International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Robust Pre-Processing Module for Leaf Image Analysis

Sadgir Prajakta, Ratnaparkhe Varsha

Abstract: Flora on earth is natural reserve for medicines. It has great healing power if encashed and conserved with great faith and devotion. Knowledge about flora empowers the methods for nurturing the medicines. Digital database creation, automation of plant recognition and identification of plant maturity can play a vital role in medicine extraction. The system for automation of process should be robust to handle on-site data so as to make the process less destructive. Pre-processing algorithms can equip the system with accuracy and robustness. The paper proposes the algorithms used for pre-processing the raw image which would aid the feature extraction and classification methods. Adaptive enhancement method for non-uniform illumination equalizes the underexposed and over exposed part in an image. Also methods like deblurring, orientation correction, size normalization in prescribed sequence improves the image quality for the later stages. The case study undertaken considers 38000 images and accuracy achived is of about 98%.

Keywords: Preprocessing, adaptive enhancement, non-uniform illumination, deblurring, object localization, adaptive cropping

I. INTRODUCTION

Herbal medicinal methods use different parts of plants for the extraction of medicines. Quality of medicine is determined by the quality, maturity and species of plant used. The quality and quantity of medicine will be optimal if the process is performed at proper maturity age as per literature. Analysis of age and quality of leaf manually is a tedious job and when the medicine is to be manufactured in large quantity then it becomes an arduous task. Hence an automatic system to be developed for analyzing and quantifying the leaves becomes an exigency of the day. For the analysis the image of a single leaf is beneficial. Preprocessing techniques are essential to derive a single leaf image from the raw data. The factors affecting the leaf image can be high illumination, the orientation, similarity of color of background. Also the leaf may be overlapping.

Various image processing methods like contrast adjustment, intensity adjustment, histogram equalization and morphological operation are used for the enhancement of images. Also object localization is very important which enhances the accuracy of the feature extraction.

Revised Manuscript Received on April 25, 2020.
* Correspondence Author

Sadgir Prajakta *, Department of Electronics Engineering, Government College of Engineering, Osmanpura, Station Road, Aurangabad, Maharashtra,India 431005. Email: psadgir@yahoo.in.

Ratnaparkhe Varsha, Department of Electronics Engineering, Government College of Engineering, Osmanpura, Station Road, Aurangabad, Maharashtra 431005. Email: patwadkar.varsha@gmail.com

Considering all the above criterion data acquisition and pre-processing play a vital role in the automation of the whole process. The pre-processing like orientation corrections, adaptive enhancement for non-uniform illumination, size normalization, background removal and adaptive cropping are techniques that are considered in this paper which are being used in the case study.[1,2, 4,5] Plants used in research are anti-inflammatory and analgesic according to ayurvedic literature.

II. PROPOSED METHOD

In the case study considered, pre-processing is an important factor that leads to a robust system for operation with a higher grade of accuracy which is essential in the concern to medicinal application. For the application of recognition plant species and to authenticate its maturity to respond for its age and usage ability with a robust environment of data collection following method is proposed.

The proposed system in the paper trying to address the problem that the query image will be captured from the real-time environment which will not have any specific standard setup. Block diagram in fig.1. outlines the complete flow of the system designed. The system will acquire the query image, pre-process the image and localize the leaves in the image itself. If in the image multiple leaves are located then the images will be separated for each leaf and will be recognized.

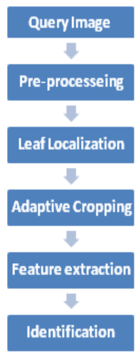

Fig.1. Basic block diagram of the system

A. Image Acquisition

Among all the parts of plants, leaves are easily available throughout the year; hence plants whose leaves are useful for medicinal purposes are selected for this project. Nine species of plants, Aegle Marmelos (Bel), Aloe Vera, Ricinus Communis (Castor), Ocimum Tenuiflorum (Tulas), Azadirachta Indica (Neem), Datura, Calotropis Gigantea (Rui), Vitex Negundo (Nirgundi) were taken in consideration for the study. Images of 10 leaves of each type of plant were selected to create the database.

Retrieval Number: D9099049420/2020©BEIESP
DOI: 10.35940/ijeat.D9099.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Fig.2. show the setup developed for database acquisition with customized illumination and rotating platform, with this rotating platform angular position can be changed concerning X and Y-axis. Fig. 3. exhibit an outer view of the setup. Leaf images from both dorsal as well as ventral sides are considered for work. Thus database of each leaf of each plant was obtained with 15 different conditions of illumination, different angular positions of the leaf and from dorsal and ventral sides of a leaf. In all 4200 images of each species are collected. In total 40000 images are considered for the research.

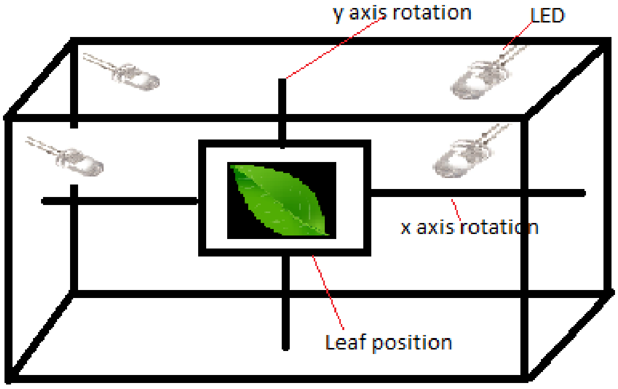

Fig.2. Inner view of database creation setup

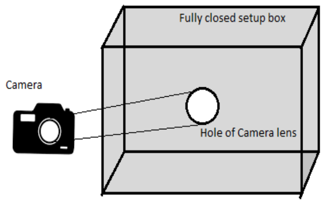

Fig.3. Outer view of the database creation setup

B. Pre- Processing

The pre-processing step in the system proposes algorithms in the order; adaptive enhancement of non-uniform illumination images, de-blurring, size normalization, orientation correction and background noise cancellation; to get an optimum image to be provided to next steps for further processing.

i) **Adaptive enhancement for non-uniform illumination images:**

In real-life situations the placement of the light source is unpredictable and the numerous objects can block the light towards the object. Hence the illumination is not uniform in the image acquired and also shadows persist. A color image acquired by the camera captures the inherent c characteristics of the scene also the lights irradiating on it. So the images have nonuniform illumination, i.e. some part is underexposed and some are overexposed.

This paper uses the technique in [1] to enhance images with nonuniform illumination with a technique that considers the local illumination patterns. In the local adaptation approach first image illumination is estimated using just a noticeable difference (JND) based low pass filter used on the Y channel of YCbCr space. The calculations for the Y channel and JND low pass filter are as follows:

$$Y(i,j) = 0.2989R(i,j) + 0.578G(i,j) + 0.114B(i,j) - (1)$$

Where R, G, and B are the intensities of RGB channels, respectively, and (i, j) represents pixel location in the two-dimensional spatial space of an image.

$$P_{JND}(i,j) = T^L(i,j) + T^t(i,j) - C^{Lt}(i,j) * min\{T^L(i,j), T^t(i,j)\} - (2)$$

where $T^L(i,j)$and $T^t(i,j)$ are the thresholds for luminance adaptation and texture masking, respectively; $C^{LT}(i,j)$ represents the overlapping effect, and $0 < C^{LT}(i,j) \leq 1$.

Next calculation is conducted to separate the over and under exposed image area. Here the two separated local areas are demarcated in two classes. The proposed demarcation changes as local luminance varies thus adapting to complicated luminance and helping to control the enhancement degree for various regions in an image. The eq.3 is used for the demarcation T (i, j) as follows:

$$T(i,j) = \frac{1 - Y_{median}}{1 + exp\{10[Y_{JND}(i,j) - 0.7]\}} - (3)$$

Where Y*median* is the median intensity value in the input luminance image Y, and is used to represent global luminance.The notation Y_{JND} is the estimated luminance by JNDbased filter.

Next calculation is conducted to separate the over and under exposed image area. Here the two separated local areas are demarcated in two classes. The proposed demarcation changes as local luminance varies thus adapting to complicated luminance and helping to control the enhancement degree for various regions in an image. The eq.3 is used for the demarcation T (i, j) as follows:

$$T(i,j) = \frac{1 - Y_{median}}{1 + exp\{10[Y_{JND}(i,j) - 0.7]\}} - (3)$$

Where Y*median* is the median intensity value in the input luminance image Y, and is used to represent global luminance.The notation Y_{JND} is the estimated luminance by JNDbased filter.

After the demarcation the local luminance calculated will be modified by SNRF to suppress bright areas and light up darker areas as in eq.4 and eq.5.

$$H_{Low}(i,j) = Y_{JND}(i,j) + 0.5Y_{mlow} - (4)$$

$$H_{High}(i,j) = 2Y_{JND}(i,j) * 0.5Y_{mhigh} - (5)$$

Where H_{low} and H_{high}, control the degree of adaptation produced by SNRF Y_{JND} represents local luminance. The notation Y_{mlow} denotes the mean value of pixels that are categorized to be under-exposure by the pixel-wise demarcation T (i, j). Y_{mhigh} represents the mean value of pixels that are categorized to be overexposure by the pixel-wise demarcation T (i, j).

Then image reconstruction based on the enhanced luminance and original image is completed using following equations.Eq. 6 denotes red plane. Eq. 7. and 8 formulates green plane and blue plane accordingly.

$$R'(i,j) = R(i,j)\left[\frac{Y'(i,j)}{Y(i,j)}\right]^{1-sqrt[R(i,j)]} - (6)$$

$$G'(i,j) = G(i,j)\left[\frac{Y'(i,j)}{Y(i,j)}\right]^{1-sqrt[G(i,j)]} - (7)$$

$$B'(i,j) = B(i,j)\left[\frac{Y'(i,j)}{Y(i,j)}\right]^{1-sqrt[B(i,j)]} - (8)$$

Retrieval Number: D9099049420/2020©BEIESP
DOI: 10.35940/ijeat.D9099.049420
Journal Website: www.ijeat.org

Where Y' denotes the modified luminance and Y is the original Luminance. Notations R, G, and B represent the RGB channels of the original color image and R', G' and B' are reconstructed RGB channel. [3].The results obtained using the following methods are as shown in fig. 4 below.

(a) (b)

Fig.4. Adaptive enhancement for non-uniform illumination images (a) Original Image (b)Image after adaptive enhancement

ii) De-blurring

Deblurring function is restoration of the original image from the blurred image . The deblurring function is designed from the knowledge of the optical system for image acquisition i.e. point speed function. The blurring of an image can have many causes:

1) Movement of the camera
2) exposure time is long
3) Out of focus object,
4) Wide angle lens
5) Atmospheric turbulence
6) Short exposure time

The blurred image can be approximately described by the eq.9.

$$g=H*f+n \quad -(9)$$

Where g is blurred image, H is point spread function, f is true image, and n is additive noise. From the equation, it can be concluded that when the PSF is convolved with the image blurred image is achived. Based on the above equation the model for de-blurring can be de-convolving the blurred image with PSF. If the PSF describes the distortion exactly then the original image can be restored exactly [4].

Also the point to be considered is de-blurring algorithm is periodic frequency pattern. Edges in the image create high frequency drop off which in-turn creates ringing effect. Algorithms are implemented to eliminate ringing by blurring the entire image. Then replacing the center pixels of the blurred image in the original image. In this way the edges of the image tapper off to a lower frequency.

In the paper the algorithm used is the blind de-convolution algorithm. The algorithm tries to improve the likelihood, when convolved with the resulting PSF. The blind de-convolution algorithm is effective when no information is available about the distortion. Using iterative process the function restores the image and PSF simultaneously. When likelihood is maximized then noise amplification tend to occur. To control the effect use of damping parameter is preferred. This parameter specifies the threshold level for the values which can deviate image from original one. Iterations are suppressed for the pixels that deviate from original values.

iii) Normalization:

As the images are from the real world can be acquired from various sources. The size of the query images can be variety of sizes and hence can size normalization is essential for further processing to optimize the time required for data processing. In the case study the image size is standardized as 300*300 pixels.

iv) Object localization

For the object localization i.e locating all the different leaves in single image. Using this step the software designed locates all the leaves of the classes selected for the case study in query image and put the bounding box around those leaves. For this the template of each class of leaves is provided to the system and its calculates SURF features of each template. Also the SURF features of the query image are calculated. Then it locates the matching templates in the image and puts bounding box around it. Here the U-SURF algorithm is utilized. In the algorithm the image is spilted in smaller 4*4 square sub regions. Simple features spaced regularly at 5*5 are computed. Then horizontal as well as the vertical direction HAAR wavelets are calculated. the responses are first weighted with a Gaussian centered at the interest point to increase the robustness towards geometric deformation and localization errors. The descriptor of length 64 is formed from all 4*4 sub regions. The wavelet responses are added from each sub region which are part of feature vectors. Absolute values of responses gives information about polarity of intensity changes.

v) Adaptive Cropping:

After number of objects are localized in the image then based on the values generated for plotting the bounding box the same values are used to crop the image. If there are 2 leaves in the image then 2 different images are generated for further processing. The number of images generated depends on the objects localized in the query image.

vi**) Orientation**:

The orientation of object in the image plays vital role in the feature extraction step. Here in the case study reference orientation of the leaf is kept as 90° with reference to mid row of the image. It means that the tip of the leaf in the image should be 90° with respect to the mid row of the image. For the same the image is threshold and segmented from the background. Then using region properties function in MATLAB the orientation of the leaf is calculated. It based on the earlier steps of pre-processing the leaf is not completely one object in the case then average of the orientation is calculated and the current leaf orientation is predicted. Then if the reference orientation and current orientation does not match then the object in the image is rotated by the difference degrees. This results in orientation correction. [5]

C) Feature Extraction

Classification of plants can be based on features such as colour, texture and shape of the leaf. The proposed system uses texture and shape among these features.

Retrieval Number: D9099049420/2020©BEIESP
DOI: 10.35940/ijeat.D9099.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

i) Texture feature

Texture feature is prominent for the application. A ridge filter is used to accentuate the ridge patterns of the leaf surface. The image is binarized with average threshold and the total proportions of white pixels are used to discriminate between single midrib, no midrib and multiple midrib leaves as represented in fig.5.

ii) Shape feature extraction:

These are very important features in the leaf segregation. This stage considered three types of shape measures such as symmetry test, geometrical & morphological features and width plot based parameters [7]

Images indicate that some leaves like *Neem* and *Datura* are not symmetrical along the midrib. This feature is used to classify the plants. The distance between the edges and the midrib is calculated according to eq.10.

$$ST = \sum_{k=i}^{n} (Side\ 1(k) - Side\ 2(k)) \quad (10)$$

Where k is point selected for test in image, *Side 1* is distance between k^{th} points on midrib to right side edge, *Side 2* is distance between k^{th} points on midrib to left side edge. Equality of both side distances with respect to midrib indicates symmetry of the leaves as shown in fig.6.

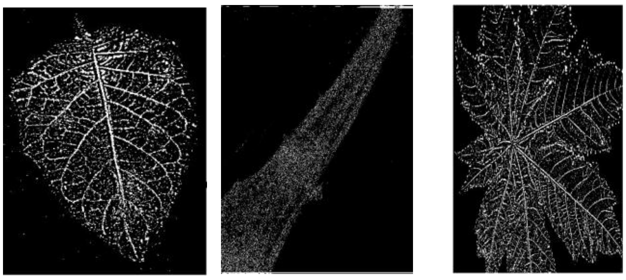

Figure.5. (a)Single midrib (b) No midrib (c) Multiple midrib leaves from database

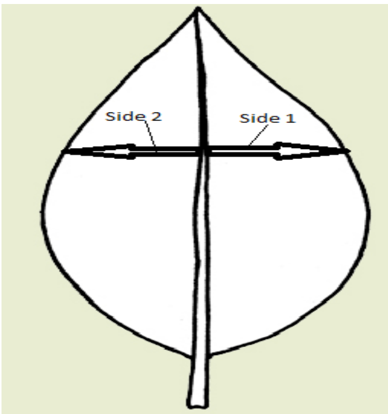

Fig.6. Image showing distances for symmetry test.

The Geometrical and morphological features of the plant leaf is calculated and are concatenated to form the feature vector. Classification based on Major axis length, Minor axis length, area, convex area, Eccentricity, perimeter, solidity, extent and equidiameter features are carried out. Complete width profile of leaf specifies the shape and hence the type of leaf. Width contour signal is considered as the digital signature of the leaf under consideration. Two level decomposition using Wavelet transform of this digital signature gives four feature values. Fig.7. shows the width plot of the Parijat Leaf.

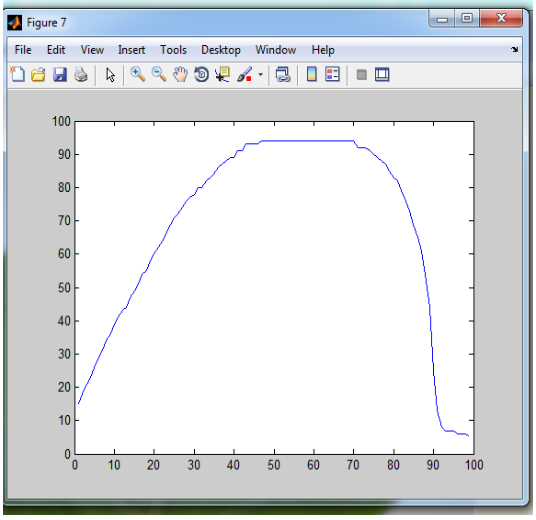

Fig.7. Width plot of Parijat Leaf

D) Classification

Tree classifier is implemented to separate gives easy and accurate way to differentiate the plants with reduced feature computation complexity and reduced time. The stepwise classification starts with the ridge based classifier. The plants are categorized into three classes i.e. no ridge, multiple ridges, single ridge. Among database plants only Castor plant in multiple midrib and only aloe-vera with no ridge. In the remaining plant database now, all are single ridge plants. By symmetry test *Neem* and Datura are segregated as they are asymmetric and single ridge. So after symmetry test 2 classes will be formed symmetrical and asymmetrical. Within asymmetrical leaves category plants, further classification is done based on geometrical and morphological features. The sub classification of symmetrical leaves plants is proposed based on width plot analysis. Wavelet decomposition based features of width contour of leaves gives classification for five types of plants *viz. Tulas, Nirgudi, Parijat, Rui, and Bel*. Fig.8. depicts the tree classification flow chart.

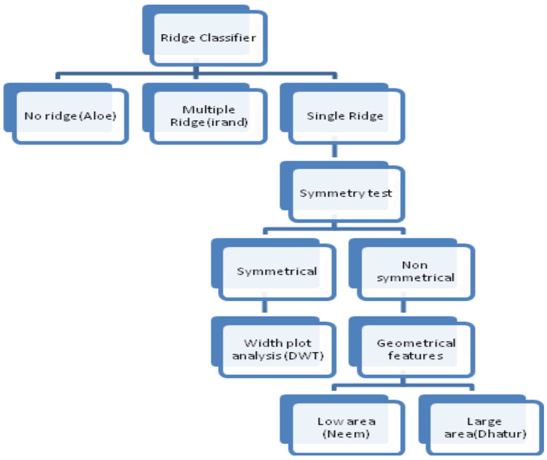

Fig.8. Tree Classification method flowchart

III. RESULT

In the addressed case study, the total database of 38000 images handled for implementation and validation of system. Among the database images of each species 70% of images i.e. 26600 images are used for training of system.

Retrieval Number: D9099049420/2020©BEIESP
DOI: 10.35940/ijeat.D9099.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

While rest case study species images as well as random plant species images are used for validation of algorithm. The results of algorithm with 3 different feature extraction methods implemented as mentioned in table I.

Table I. Accuracy of different feature extraction algorithms.

Sr.no.	Features	Accuracy	Accuracy with pre-processing
1.	Geometrical and morphological features	70%	80%
2.	Ridge filter based geometrical features	80%	85%
3.	Width plot analysis based Features	92%	96%

In order to tune the accuracy image pre-processing methods are implemented. The accuracy of three methods improves after inclusion of pre-processing algorithm as shown in the table I.

When the decision tree classification is implemented the accuracy improves. The accuracy is 94% with higher speed. If tree classification is done on pre-processed images, the accuracy is increased to 98%.

IV. CONCLUSION

Computer based identification of plants enables to precisely identify plants on large scale with less domain knowledge. The pre-processing of image leads to improve the plant images to enhance the features which will assist in better classification. Al sthWidth plot analysis a novel approach proposed in the paper which provides accuracy up to 98% when tree classification is implemented and it also provides increase in speed. Tree classification reduces load on classifiers and facilitates effective use of features. As the study consists of medicinal plants the basic level of maturity prediction of plant leaf is proposed on the discrimination and can be improved in further research. As all plants selected are green hence for maturity decision using hue discriminator is useful. All the plants considered in the studies are used for medicinal purpose during its mature i.e. mid age.

REFERENCES:

1. Joao B. F., Odemir M. B., Davi R. R., Rosana M., Maria C. and Gabriel L. 2016, identifying plant species using architectural features in leaf microscopy images, Botany. 94, No.1, 15-21.
2. Hiew B.Y., Teoh A. B.J., Nago 2006, Automatic Digital Camera Based Fingerprint Image Pre-processing, *International Conference on Computer Graphics, Imaging and Visualisation*, 182-189.
3. Yanfang Wang, Qian Huang, Jing Hu, "Adaptive enhancement for non-uniform illumination images via nonlinear mapping," J. Electron. Imaging 26(5), 053012 (2017),
4. P. S. Seem V., 2019, Comparative Analysis of Segmentation techniques for Progressive Evaluation and Risk Identification of Diabetic Foot Ulcers, *4th MEC International Conference on Big Data and Smart City.* ii-iii.
5. Petrou M., Pedro G. S. 1999 *Image Processing the fundamentals, Wiley Publication*, ISBN 0-471-99883-4.
6. Gonzalez, Woods and Eddins ,2002, *Digital Image Processing Using Matlab*, Prentice-Hall, Inc. ISBN 0-201-18075-8.
7. Biva Shrestha 2010, Thesis on Classification of Plants Using Images of Their Leaves, Appalachian State University, Master of Science.9-20.
8. Daniel S. 2018 Local Normalization: Filter to reduce the effect on a non-uniform illumination, *Biomedical Image Group, EPFL*, Switz, Retrieved from http://bigwww.epfl.ch/sage/soft/localnormalization/.
9. Jyotismita C, Ranjan P. and Samar B. 2018, Plant leaf classification using multiple descriptors: A hierarchical approach, *Journal of King Saud University – Computer and Information Sciences*, 1-15
10. Nursuriati J., Nuril A. C., Sharifalillah N., Khalil A., 2015 Automatic Plant Identification: Is Shape the Key Feature?, *IEEE International Symposium on Robotics and Intelligent Sensors, Procedia Computer Science* 436 – 442.
11. Dengsheng Z., Guojun Lu. 2004 Review of shape representation and description Techniques, Pattern Recognition ,Pattern Recognition Society. Elsevier Ltd.1-19.
12. Jyotismita Chaki, Ranjan Parekh,2011 Plant Leaf Recognition Using Shape Based Features And Neural Network Classifiers, *International Journal of Advanced Computer Science and Applications*, 2, 41-47
13. James S. C., David C., Jonathan Y. Clark, Paolo R., Paul W., 2012, Plant Species Identification using Digital Morphometrics: A Review, *Expert Systems with Applications, Elsevier*, 7563-7577.
14. Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang, 2007, Leaf shape based plant species recognition, *Applied Mathematics and Computation*, 185, 883–893
15. David Dagan Feng, Wan-Chi Siu, Hong-Jiang Zhang (Eds.), 2003 Multimedia information Retrieval and Management Technological Fundamentals and Applications, *Springer-Verlag Berlin Heidelberg, Softcover reprint* ,1 st edition. 432-457

AUTHORS PROFILE

Prajakta Sadgir received her B.E. degree in Electronics and Telecommunication engineering from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India in 2012, M.E. (Electronics) degree from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India, M.S., India in 2014 and at present research scholar pursuing Ph.D. degree in Electronics from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India. Currently working as Assistant Professor in College of military Engineering, Pune, M.S., India. Her area of interest includes Image Processing, Pattern Recognition, Machine learning.

V.R. Ratnaparkhe received her B.E. degree in Electronics and Telecommunication engineering from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, M.S., India in 1989, M.Tech.(EDT) degree from C. E. D. T. Aurangabad, M.S., India in 1996 and the Ph.D. degree in Electronics from Swami Ramanand Teerth, Marathwada University, Nanded, M.S., India in 2008. Currently she is Associate Professor in Government Engineering College, Aurangabad, M.S., India.

She has published many research papers in Journals (12) and Conferences (8). Her area of interest includes Image Processing, Pattern Recognition.

Retrieval Number: D9099049420/2020©BEIESP
DOI: 10.35940/ijeat.D9099.049420
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering & Sciences Publication