ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2018.0272

PLANT RECOGNITION SYSTEM USING LEAF SHAPE FEATURES AND MINIMUM EUCLIDEAN DISTANCE

Farhana Haque[1] and Safwana Haque[2]

[1]*Department of Aerospace, Transport and Manufacturing, Cranfield University, United Kingdom*
[2]*Department of Computer Science and Engineering, College of Engineering and Technology, International University of Business Agriculture and Technology, Bangladesh*

Abstract

The study presents a plant recognition system that uses image and data processing techniques for recognition. A lot of research has been going on to identify plants by their leaves and one of the features that is used is the shape of the leaf but the accuracy is not high and therefore other features should also be considered to increase the accuracy. This system designed has three main steps which are image pre-processing, feature extraction and matching. Image pre-processing performs basic operations on the leaf image for segmentation which helps in making feature extraction easy. Seven (7) leaf features derived from geometric parameters of leaf shape were extracted from the pre-processed image and the simple principle of minimum Euclidean distance was used for finding the closest match to the input leaf image. The system used 10 species of leaves with a total of 50 leaf images from the flavia dataset for testing and obtained an accuracy above 90%. The algorithm is accurate and is easy to implement. However, it is slow and not tested on a large dataset. It is hoped that this proposed system will be exploited further and the speed will be improved and will also be able to give more information on the plant.

Keywords:

Euclidean Distance, Feature Extraction, Image Pre-Processing, Leaf Classification, Specie Recognition, Image Segmentation

1. INTRODUCTION

Plants are important for the sustenance of life as they provide mankind and animals with oxygen, food and medicine. Plants provide a wealth of raw materials for our industries such as wood, jute, oil, chemicals etc. Plants can also contribute in wind breaking, prevention of erosion, keeping the temperature of the earth stable and also maintaining the water, nitrogen and carbon cycles on earth. Due the numerous benefits of plants in our life, we need to study and know the features of plants and classify them so that we can prevent them from extinction, protect them from diseases, increase productivity and contribute to research areas in agriculture.

One effective way of conducting studies on plants these days is by the use of image processing. Image Processing is being used in agricultural research areas widely where images are converted into digital format and useful information are extracted from those images or characteristics associated with that information are studied. Examples of image processing usage in agriculture is the identification of plants via their leaves, flowers or seeds, measurement of leaf chlorophyll, identification of plant diseases via their leaves etc.

To conduct extensive studies on plants it is necessary to first know what a particular plant is or what specie it belongs to and one of the most common ways of doing this is by studying its leaves. Leaves are the most prominent features of plants and contains some details such as shape, colour, vein arrangement, texture etc., which can be used to distinguish plants from one another [1].

The plant recognition system developed in this study uses a leaf image for identification. The leaf image is used to extract some features derived from its geometrical parameters such as area, perimeter, length and width. These features are then used for identification of the plant. The system displays the name of the plant that has the closest match to the leaf image that was used.

2. LITERATURE REVIEW

There have been many approaches by researchers to find an effective classification of plants by image analysis and the most widely used ones were based on colour and shape of leaves. Other approaches included identifications based on leaf texture and structure, stem, flowers and seedlings [2]. However, it could be seen that identification through colour could be affected during the image acquisition process, climate changes or nutrients. It could also be seen that classification based on flowers and seedlings were difficult because of its complex 3D structure [2].

The classification of plants has used various Artificial Intelligence methods such as the Probabilistic Neural Network (PNN), Artificial Neural Network (ANN) and Back Propagation Neural Network (BPNN). It was seen that the PNN yielded faster result with the accuracy of 90.321%. The most common features that were extracted from the leaves were the leaf length, width, area, and perimeter [1].

The area of the leaf was calculated by counting the number of white pixels on the smoothed image of the leaf [2]. However, there has been an attempt to classify nine different medicinal plants based on leaf colour, area and edge features. The area of the leaf was calculated by finding the product of the number of pixels occupying the area of the binary leaf image and one-pixel count. One-pixel count was obtained from the area of a coin divided by the pixel count [3].

The canny edge detection histogram was used to calculate the edge histogram. The difference in area, colour and edge histogram were calculated for test images and each database images. The average of these three values was calculated and the pair with the least difference was the identified plant. It was observed that the leaf characteristics can vary from a young age to a matured age which can affect the efficiency of a system [3].

However, leaf colour can be affected by seasons and leaf recognition based on this feature can affect the accuracy. One other factor that can affect the accuracy is when the user needs to

select the ends of a leaf for algorithms based on leaf contour extraction [4].

K. Lee and K. Hong [4] proposed a system that uses leaf veins and shapes for recognition. Firstly, the leaf contour is extracted. Secondly, the difference between grayscale image and opening operation performed on grayscale image is converted to binary image to obtain vein image. The main vein is then extracted by the projection histogram in the horizontal and the leaf direction is by measuring the histogram in vertical.

S.G. Wu et al. [1] improved the feature extraction and used Probabilistic Neural Network (PNN) for classification. To determine the best threshold for the conversion of grayscale image to binary image, the average of R, G, B of 3000 leaves were calculated. The lowest point between the two peaks i.e the pixels covering the leaf images and the other pixels covering the white background was calculated and divided by 255 which yielded a result of 0.95. Five (5) basic features were extracted; physiological length (L_p) and width (W_p), leaf area (A) and perimeter (P, measured by counting the number of pixels on the leaf contour), diameter (D). The physiological length must be stated by user by clicking the two terminals of the leaf using the mouse and as stated earlier [4] this can be a factor that affects the accuracy. Out of these five features, twelve (12) more morphological features were calculated which are: smooth factor, aspect ratio (L_p/W_p), form factor ($4\pi A/P^2$), rectangularity (L_pW_p /A), narrow factor (D/L_p), perimeter ratio of diameter(P/D), perimeter ratio of physiological length and physiological width ($P/(L_p+W_p)$) and five (5) vein features. PNN was used for classification and it has been stated that it is easier and more efficient than using ANN and BPNN. The proposed algorithm had an accuracy of approximately 91% tested on 32 kinds of leaves.

J.X. Du et al. [2] described aspect ratio as ratio of maximum length (D_{max}) to minimum length (D_{min}) of minimum bounding box, the rectangularity ratio was measured using the area of region of interest (ROI) to the area of the bounding box ($AROI/D_{max}*D_{min}$) and form factor was given by $4\pi AROI\ /P^2\ ROI$ where $PROI$ is the perimeter of the ROI.

T. Munisami et al. [5] created a dataset of 640 leaves which contained 32 different species and named it folio. Plant classification using k-Nearest Neighbour was based on shape features and colour histogram which yielded an accuracy of 87%. The algorithm stated above was also used to test the flavia dataset and it yielded a result of 91.10%. Image pre-processing included rotation if the width was greater than the height; grey scaling, thresholding, opening operation to remove small holes due to noise; inverse threshold to convert the background to black; edge extraction using Suzuki algorithm and edge filtering. Feature extractions performed were convex hull information which included area, perimeter of hull and number of vertices extracted. The length, width, area and perimeter of leaves were also calculated. Lines were drawn on the horizontal and vertical axes and the centroid was calculated too.

The information extracted above were used to create ratios for the matcher i.e. aspect ratio, white area ratio, perimeter to area, perimeter to hull, hull area ratio, distance map X, distance map Y and centroid radial distance. The colour histogram was applied to a cropped region of the leaf to avoid interference with the white background. All the ratios were normalized between 0 and 1.

The matching process stated consists of two processes. Firstly, the leaf to be recognized undergoes the same steps as mentioned above and compared to the 640 leaves in the database out of which the three leaves that produces the least Euclidean distance is returned. Secondly, for situations where the results are from different species, the colour histogram is compared and a correlation coefficient is calculated. If the result is 1, it means the plants are very similar, but if the result is -1 the plants are not close.

X. Wang et al. [6] extracted the bounding box, convex hull, inscribed circle, circumscribed circle, centroid and contour to calculate the aspect ratio, rectangularity area convexity, perimeter convexity, sphericity, circularity, eccentricity and form factor. They proposed a moving centre hyper sphere classifier and classified 20 different species. A recognition rate of 92.2% was obtained.

It can be seen that the leaf shape has been a common feature that was used in most plant identification approaches and the basic features that were calculated were leaf length, width, area and perimeter which were further used to extract some values such as the aspect ratio, form factor, rectangularity, eccentricity and convex hull information [7]. However, to increase the accuracy, other features such as the colour and texture of leaves were considered [8].

Complex techniques such as PNN, BPN, k-nearest neighbour (kNN) or hypersphere classification can be used for identification of plants which have a large dataset, but for small datasets, simple algorithms such as average of parameters can be utilised. However, these algorithms can affect the accuracy if the dataset is large.

3. MATERIALS AND METHODS

3.1 METHODOLOGY

Experimental design has been applied for this research and it is the process of planning an experiment in advance. It is considered the most accurate form of experimental research as it tries to prove some hypothesis mathematically with statistical analysis.

3.2 RESEARCH PLAN

The research is aimed towards designing and implementing a plant recognition system. It is important to recognize plants as it can be used for medicinal, educational or herbal purposes. The research plan is an important part as it helps in gathering and answering important questions that are related to the research. The steps taken are as described below.

3.2.1 Problem Definition:

As it has been seen from earlier researches carried out, that plant recognition has applied a lot of techniques and shape is the most common feature. However, shape features might need to adapt other features such as colour, texture or vein arrangement for more accuracy.

3.2.2 Review of the Plant Recognition Method:

As seen from the literature review, a lot of features have been considered in plant recognition systems such as shape, colour,

texture and vein arrangement. These have been used for enhancing accuracies. The general shape features include perimeter, area, length and width. It was also stated that ANN or hypersphere classifiers give higher accuracies but for small data, a simpler technique can be used such as the averaging.

3.2.3 Dataset Preparation:

The leaf images were obtained from flavia dataset. However, this study makes use of 10 different species with five image samples from each specie shown in Fig.1. The leaf images had white background which makes it easy to perform image pre-processing. However, other backgrounds can be used but must have a high contrast with the leaf image so that it would be easy to differentiate the object from the background [9]. 50 images were used for testing the system and these images were RGB images in '.jpeg' format (Fig.2). These images were segmented and stored earlier for matching purpose in '.bmp' format as seen in Fig.3.

3.3 RESEARCH FRAMEWORK

Research framework in Fig.4 shows the overall flow that was followed during the development of the system. The Fig.1 shows the flow of the plant recognition system.

Fig.1. 10 different species that were used

3.3.1 Input of Leaf Image:

The user needs to launch the software MATLAB R2013a and make sure that the segmented images folder is added to the correct path. The user needs to run the program and a simple GUI as shown in Fig.5 appears, which is designed for locating and selecting an image from anywhere in the PC. The GUI has a button for selecting the image and once it is clicked, a popup window comes up to locate the place where the image to be tested has been stored.

The selected image for recognition should be an RGB image in '.jpeg' and the leaf image should be captured on a white background, so that there would be less noises in the image. Segmentation process can be carried out with convenience on images with white background and this leads to better chances of accuracy.

3.3.2 Application of Image Pre-Processing:

The framework shown in Fig.6 describes the steps taken for pre-processing the input leaf image. The following is the description of each step used in the framework:

- *Resize the image*: The RGB leaf image is resized to 800×600 resolution. The image is processed faster if it is not very large.
- *Convert from RGB to grayscale*: The RGB image is converted to a grayscale with a threshold frequency of 0.95 as it gives very good results. The converted image is shown in Fig.7 below.
- *Convert from grayscale to binary*: The image is further converted to a binary image as seen in Fig.8, so that morphological operations can be carried out with ease.
- *Swap between the white pixels and black pixels*: The binary image has the leaf region pixels in 0 and the background at 1, but it is swapped to make the leaf image in 1 and the background in 0, so that the ROI can be easily distinguished.
- *Perform morphological opening*: Morphological opening performs two operations on the image i.e. erosion followed by dilation. It removes the smaller objects from the object and preserves the shape and size of bigger objects.
- *Use average filter*: The average filter is applied to remove noise from the image and gives a smoothness to the leaf image.
- *Extract the Region of Interest (ROI)*: It is necessary to specify the ROI of the image to extract some details. The ROI takes the white pixels of the image into consideration as seen in Fig.10.
- *Obtain the properties of ROI*: Some properties of the ROI are calculated such as area, perimeter, eccentricity, etc. for the features extraction process.
- *Convert the image into convex hull image*: In mathematics, the convex hull of a set of points, Y, in the Euclidean space is the smallest convex set that contains Y. This is done to be able to get properties of a convex image as shown in Fig.11.
- *Obtain the properties of convex hull image*: Some properties of the convex hull image are calculated such as area, perimeter etc. for the features extraction process.
- *Obtain the length and width of bounding box*: The bounding box is the smallest rectangle that contains all the white pixels of the leaf image shown in Fig.12. It is done to get the length and width of the image.

Fig.2. 50 RGB images that were used for testing the system

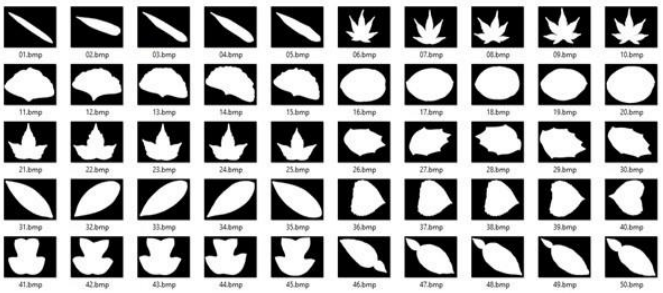

Fig.3. Fifty segmented images used for matching with five from each specie

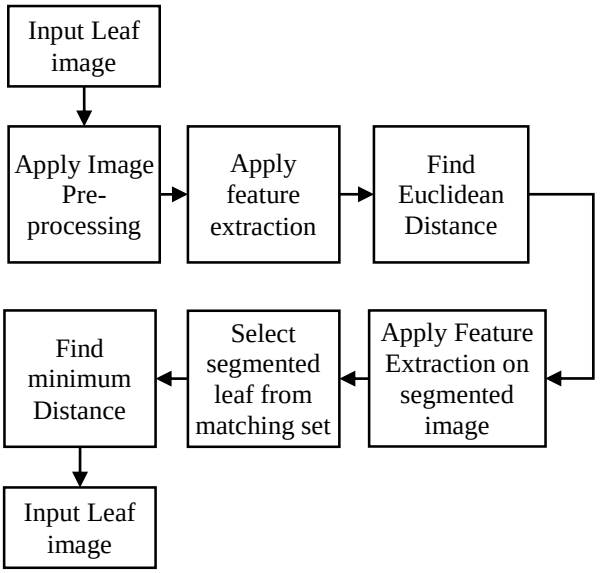

Fig.4. Framework of the plant recognition system

3.3.3 Application of Feature Extraction:

The area is calculated by counting the number of white pixels in the region and the perimeter is calculated by counting the distance between the adjacent pixels in the border of the ROI. The features below are derived from the geometric parameters.

The following ratios are extracted and computed from the leaf image [6].

Aspect Ratio = length of Bounding Box / Width of Bounding Box

Rectangularity = area of ROI / area of bounding box

Area convexity = area of ROI / area of convex hull

Fig.5. GUI to select the image that needs to be tested

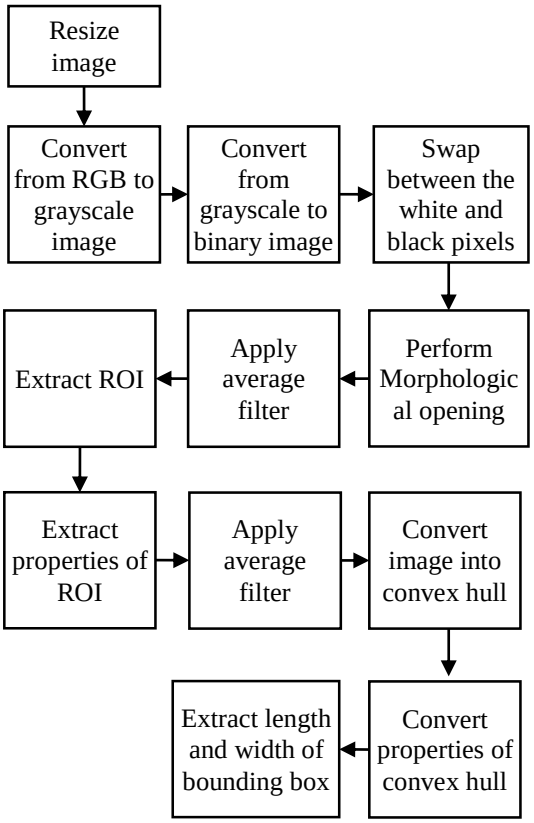

Fig.6. Framework of Image Pre-processing

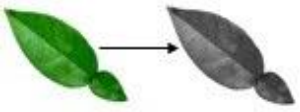

Fig.7. RGB image to Grayscale Image

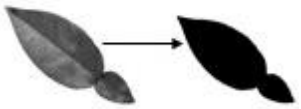

Fig.8. Grayscale Image to Binary image

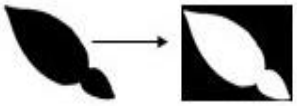

Fig.9. Swap between the pixel values

Fig.10. Image showing the selected ROI

Fig.11. Image showing the selected convex hull

Fig.12. Image showing the bounding box

Perimeter convexity = Perimeter of ROI/perimeter of convex hull

Circularity = 4**PI** area of ROI/(perimeter of convex hull)^2

Eccentricity = foci of the ellipse/major axis length

Form Factor = 4* *PI** area of ROI/(*perimeter of ROI*)^2

3.3.4 Leaf Selection, Feature Extraction on Segmented Image and Euclidean Distance Calculation:

The system selects one segmented image that was stored earlier and extracts the ratios that are stated above from the binary image. It then computes the Euclidean distance from the ratios of the input image and the segmented image. Afterwards, it calculates the Euclidean distance between the input leaf and the next segmented images and so on till all the distances have been found. The Euclidean distance, D, is calculated using the formula below:

$$D = ((x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2)^{0.5}$$

3.3.5 Minimum Distance Calculation:

The smallest distance between the input image and the segmented images after finding the Euclidean distance is computed.

The leaf image that has the least Euclidean distance is displayed with its name. It is considered to be the closest match of the input image.

3.3.6 Display of the Closest Match:

The leaf image that has the least Euclidean distance is displayed with its name as shown in Fig.13. It is considered to be the closest match of the input image.

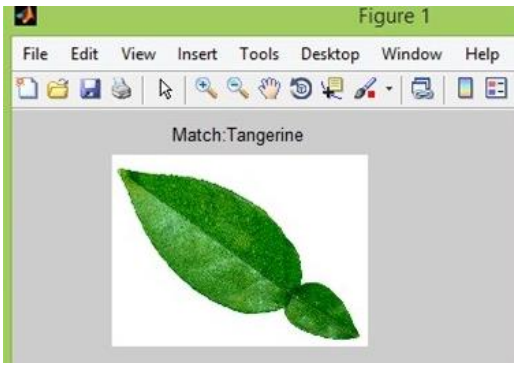

Fig.13. Image with the closest match is displayed

3.4 BUILDING THE STAGES

The software and hardware used for developing the system are as discussed below:

3.4.1 Software:

The image processing software used is MATLAB R2013a for a 64-bit system. The software has been used for the entire process of development. MATLAB is a high performance language used for computation and it integrates visualization, computation and programming in a user friendly environment. It can handle two dimensional and three dimensional visualizations and can be used to build graphical user interfaces (GUI).

3.4.2 Hardware:

The hardware used in the development of the system is Hewlett Packard (HP) Envy 15k019nr Notebook PC. It uses 4th generation Intel Core i7-4510U processor with a frequency of 2.0GHz and maximum turbo at 2.6GHz. It also contains NVIDIA GeForce Graphics card and has a capacity of 2GB. It has a hard disk of ITB capacity and 8GB DDR3 SDRAM. It contains a Windows 8.1 64-bit Operating system. The lithium-ion battery lasts approximately. 3.5 hours and this work was completed on time and successfully without any technical faults.

3.4.3 Testing and Evaluation:

The experiment was carried out on 50 RGB leaf images as shown in Fig.3. The minimum Euclidean distance was used for recognition because it is an easy method to implement when a small number of images are used for matching. The leaves are labelled and the percentage of match for each specie is shown in Table.1 below.

Table.1. Results from the experiment

Leaf Image Number Label	Leaf Name	Match (%)
1 – 5	Pubescent Bamboo	100
6 -10	Japanese Maple	100
11 -15	Ginkgo	60
16 -20	Crape Myrtle	100
21 – 25	Trident maple	100
26 – 30	Beale's barberry	80
31 – 35	Southern Magnolia	100
36 40	Canadian Poplar	100
41 – 45	Chinese tulip tree	100
45 – 50	Tangerine	100

The line graph in Fig.14 shows the leaf image numbers that failed to give the correct result. As it can be seen, leaf images labelled 12, 14 and 30 gave a wrong match to the leaf image that was entered for testing. If a correct match is identified the value corresponds to 1 and an incorrect match corresponds to 0. The overall percentage match for the leaf species are shown in Fig.15.

4. RESULTS AND DISCUSSIONS

Fifty leaf images were used to test the system and three images failed to return the correct results. The accuracy was measured by the following formula:

$$Accuracy = IC/\ IT \times 100$$

where *IC* is the total number of correct image matches and *IT* is the total number of leaf images used for testing. The result is multiplied by 100 to give the percentage value.

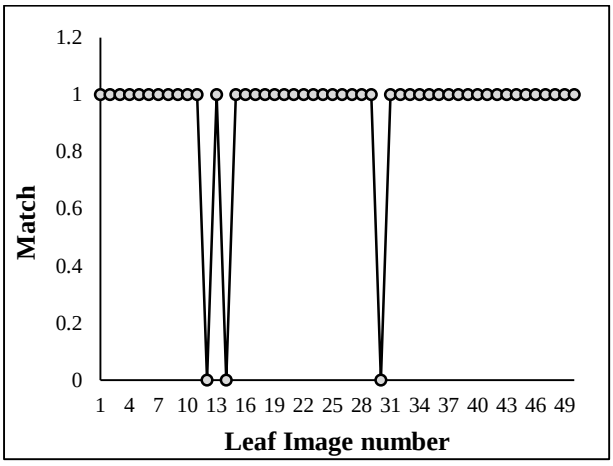

Fig.14. Line graph showing correct and incorrect matches of sampled leaf images

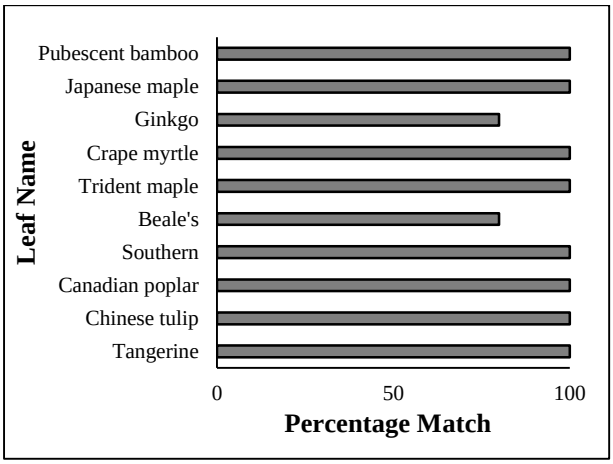

Fig.15. Graph showing the percentage match for each leaf specie

It was computed from the formula above that the accuracy of the system was 94%. The pie chart in Fig.16 shows the accuracy of the system.

The objectives of the system were achieved using simple algorithms. The image processing algorithm used on the input image removed the noise with convenience and gave good results to perform feature extraction with ease and the algorithm steps were accurate for this research.

It can also be seen that the features derived from the geometric parameters of the leaf shape can be used for recognition instead on using contour based features. The simple minimum Euclidean distance technique can be used for matching and has proved to work efficiently with a small dataset.

However, if a large dataset of images is used with various sizes of leaves, the system's accuracy might reduce and other features with a better classification technique might need to be adapted to improve the recognition rate. The system is easy to use and understand, but it is a bit slow as it needs to do a lot of mathematical calculations. The system only displays the name and image of the plant that has the closest match to the input leaf and this need not be necessarily correct all the time as the system matches based on values extracted from the leaf images.

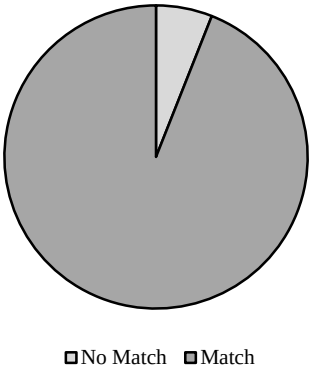

Fig.16. Graph showing the total percentage for correct and incorrect match

5. CONCLUSION

The plant recognition system was developed based on shape features and minimum Euclidean distance. There were basically three steps that were used which were image pre-processing, feature extraction and matching using minimum Euclidean distance. The algorithms used proved to be accurate for the dataset used.

The system is efficient for use as it has given a high accuracy of 94%, but the number of leaf species in the segmented images needs to be increased so that it could be used for a wide range of leaves which will be more beneficial for plant recognition.

The system performs the computations correctly, but it takes time to process the images and find the closest match because it has a lot of computations that need to be carried out. A better matching technique needs to be adapted that would be faster in processing and displaying the result.

Future works may be carried out to improve the speed and also to expand the plant recognition system by adding more dataset and also more information about the plant. It would be beneficial to the society to give more information about the plant as plant recognition systems are used in a wide range of domains.

REFERENCES

[1] S.G. Wu, F.S. Bao, E.Y. Xu, Y. X. Wang, Y.F. Chang and Q.L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pp. 15-18, 2007.

[2] J.X. Du, X.F. Wang and G.J. Zhang, "Leaf Shape based Plant Species Recognition", *Applied Mathematics and Computation*, Vol. 185, No. 2, pp. 883-893, 2007.

[3] S. Kumar, "Leaf Color, Area and Edge Features based Approach for Identification of Indian Medicinal Plants", *Indian Journal of Computer Science and Engineering*, Vol. 3, No. 3, pp. 436-442, 2012.

[4] K. Lee and K. Hong, "An Implementation of Leaf Recognition System using Leaf Vein and Shape", *International Journal of Bio- Science and Bio- Technology*, Vol. 5, No. 2, pp. 57-66, 2013.

[5] T. Munisami, M. Ramsurn, S. Kishnah and S. Pudaruth, "Plant Leaf Recognition using Shape Features and Colour Histogram with K-Nearest Neighbour Classifiers", *Procedia Computer Science*, Vol. 58, pp. 740-747, 2015.

[6] X. Wang, J. Du and G. Zhang, "Recognition of Leaf Images based on Shape Features using a Hypersphere Classifier", *Proceedings of International Conference on Intelligent Computing*, pp. 87-96, 2005.

[7] J.S. Cope, D. Corney, J.Y. Clark, P. Remagnino and P. Wilkin, "Plant Species Identification using Digital Morphometrics: A Review", *Expert Systems with Applications*, Vol. 39, No. 8, pp. 7562-7573, 2012.

[8] J. Chaki and P. Ranjan, "Plant Leaf Recognition using Shape based Features and Neural Network Classifiers", *The Journal of Advanced Computer Science and Applications*, Vol. 2, No. 10, pp. 41-47, 2011.

[9] S.B. Patil and S.S. Patil, "Measurement of Sugarcane Leaf Chlorophyll", *International Journal of Application or Innovation in Engineering and Technology*, Vol. 1, No. 3, pp. 97-102, 2014.