

A New Leaf Venation Detection Technique for Plant Species Classification

Abstract

This paper presents a novel approach to classify the leaf shape and to identify plant species using venation detection. The proposed approach consists of five main steps to extract the leaf venation including canny edge detection, remove leaf boundary, extract curve, and produce hue normalization image and image fusion. Moreover, to localize the edge direction efficiently, the lines that extracted from pre-processing, are further divided into smaller segments. Thirty-two leaf images of Malaysian plants are analysed and evaluated with two different datasets, Flavia and Acer. The best accuracy is obtained by 99.3% and 91.06% for Flavia and Acer datasets respectively. Experimental results show the effectiveness of the proposed approach for shape recognition with high accuracy.

Keywords: Leaf Venation; plant species; features extraction; features selection; classification.

1 Introduction

The plant is one of the most important living organisms which is categorized in kingdom Plantae. Although kingdom Plantae is not the largest kingdom it could be found everywhere on earth and was the backbone for all any others living organisms. Since stone-age, the usages of plants on earth become profoundly important in mankind activity.

Based on the plant taxonomy theory, the plant can be identifying based on their external structure such as leaf, seed, flower and fruit [1]. However, in this paper, only the characteristic of leaves are derived to identify the plant species.

Performance of leaf recognition is highly depending on the quality and trustworthy of the retrieval feature. However, it is an arduous task to acquire meaningful and unique feature from a low variation plant species leaves. It is more exhausting when implements the botanical knowledge in computer vision. Traditional plant identification use descriptions in literature or natural language to discriminate the plant species, namely terminology. However, it becomes a strenuous activity to transform them in mathematic logic for computer understanding. Based on [2], the information derived from some significant leaf parts such as leaf apex, base, margin and venation are still not well exploited despite those parts of leaves are defined as a significant morphological trait and usually used by a specialist to identify the plant species.

Human experts identify the plant by using phenetic such as dichotomous key and multi-access key. Both of this approach require auxiliary of figures, characteristic state and description. According to [3], not less than 50 units of characters are used as plant identification keys. It imparted that the number of characters used should not be less than the mentioned numbers otherwise it will be unreliable and unstable. Those characters can be described in the distinctive or characteristic state. The morphological character of plant leaves describes the multi-part of leaves such as leaf venation, shape, margin and so on. However, character states for leaf venation are parallel, palmately, pinnately and so on.

2 Related work

For venation, Amlekar et al. [4], Casanova et al. [5], Kadir et al. [6][7][8], Fern et al., [9], Mishra et al. [10], Pornpanomchai et al. [11], Rolland-Lagan et al. [12] and De Oliveira Plotze and Bruno [13] measured the area, diameter, density, length and width for the leaf venation. However, this measure may influence by many factors such as the methods used to detect the leaf venation, the scaling, and the solution of the image. Therefore, this kind of measurement is unreliable and unstable. From the botanical perspective, it should be distinguished by the leaf venation pattern. The slope of the venation is the important measurement to recognize the leaf venation. Therefore, in this proposed method, by using the theory of botanical knowledge, the pattern of leaf venation can be distinguished easily.

The popular methods that used to detect the skeleton of the leaf are Canny Edge Detection with Gradient Magnitude, Ant Colony, Advanced Shape Context, chemical solution, Independent Analysis and so on.

Amlekar et al. [4] claimed that it is the easier way to identify the plant species by using leaf venation. He first removed the noise of the input image then followed by detecting the venation by using canny edge detection. However, he did not stop after getting the leaf veins from the output of Canny Edge Detection just like others researchers had done. He continued by finding the gradients of large magnitude to discard the edges which are not connected. Figure 2 shows the research image extracted from his research paper.

The method proposed by Syahputra et al. [14] was slightly out of the ordinary. Most of the researchers only had single input image. However, the author proposed using double input image and indicated them as left image and right image. Firstly, the maximum intensity of both images was filtered. Then, algorithm SAD was proposed to measure the disparity between both images in sub-window. The size of sub-window was not mentioned in their published work.

Mouine et al. [15][16] proposed Advanced Shape Context from SC0 to SC2. They introduced voting set to find the salient point to represent the venation network. However, the salient points included margin point as well, therefore step SC2 is to distinct the margin point from salient points. SC0 represents the margin points and SC1 represents the founded salient points. Figure 3 is the venation network which retrieved from Mouine et al. [15].

Jobin et al. [17] did not mention what method they used to extract the leaf venation. However, the output image of leaf venation displayed was considered as displeased output. It is while Mishra et al. [10] applied the chemical procedure to remove the leaf pigment but remain the leaf veins. Therefore, the veinlet was apparently and clear. However, this method is time-consuming and troublesome.

Cope et al. [18] recommended Ant Colony approach to search for the leaf veins. This method ransacked the potential leaf veins pixel by pixels. Edge magnitude is the measurement to distinguish if the pixel is included in leaf veins or not. This algorithm is complicated, time-consuming and heavy computation

Li, Chi and Feng [19] used Independent Component Analysis (ICA) to extract the leaf venation. They first divided the sample image into several subcomponents. Then they used ICA to find the possible lines in each subcomponent.

3 Leaf Venation

Leaf venation is the skeleton of leaf and the examples of leaf venation are shown in Figure 4. The skeleton extends from midrib to the leaf blade. The midrib is the primary veins and the veins that extend from primary veins are called secondary veins. The fine veins that extend from secondary veins are tertiary veins. The pattern of leaf veins determines its venation states. The veins consist of vascular tissues and connect from the petiole to the leaf. The function of leaf veins is to transport the food, water, saps and soluble ions into the leaf.

The parallel venation has longitudinal veins and parallels to each other. Its veins are extended from leaf base to leaf apex. The veins of the pinnate venation are extended from the midrib to both sides of leaf margin. The reticulate venation has net-veined and its veins formed a complicated network. It is formed by midrib, secondary veins and tertiary veins. Leaf with only visible midrib is known as nervous and this kind of leaf venation is without secondary and tertiary veins. The palmate venation has several main veins which intersect at one point and extend from the leaf base to the tip of lobes.

The popular methods that used to extract the feature of leaf venation are Digital Morphology, Salient Point, Local Edge Orientation Histogram, Eagle Angle, Fractal Dimension, Bifurcation point and so on.

Amlekar et al. [4] proposed seven morphology features to describe the leaf venation. The features used are an area, perimeter, major length, minor length, convex area and solidity. Then they classified them by using k-nearest neighbour classifier (KNN) and back propagation artificial neural network (ANN). He used 173 input image in training mode and 46 in testing mode. They got a pretty good result in ANN which was 80.43% and the KNN classifier gave a poor result with only 54.35%.

As mentioned earlier Mouine et al. [15][16] used Advanced Shape Context SC0 to SC2 to find the salient points to represents the venation networks. Then, they searched and queried the features of salient points by using Locality Sensitive Hashing (LSH). This method finds the distance and distinguish between two features and is quite a time-consuming method. The plant identification method of Mouine et al. [15] fully relied on venation features. However, in the year 2013, they included the shape features by using triangular angle area (TAR) and triangular side length (TSL). Their contribution in plant identification was then transformed into the android application (Mouine et al., [16]).

Casanova, Backes and Bruno [20] normalized the vein network to ensure the venation is independent of scale venation. Then the geometrical data of venation was determined. The maximum degree and the average degree of leaf veins were computed. Then, this measurement was then applied to the extracted leaf venation by various thresholds.

Most of the researchers retrieved the leaf texture with entire leaf, however, Mzoughi et al. [2] split the entire leaf into apex and base regions. After that, they only continued to retrieve the texture information of leaf by using Fourier histogram, edge orientation histogram and local edge orientation histogram. Local edge orientation is the revolution of edge orientation histogram, the only difference is that the local orientation accumulated the orientation within the blob.

4 Proposed Methodology

In this research, botanical features are used to detect the interest regions of leaf part and extract the features of leaf part. Total seven leaf parts are to identify the plant species. There are leaf shape, leaf lobes and sinuses, apex, base, margin, venation and texture. In these seven features, only texture's features did not embed botanical features. According to Rao (2012), texture had several branches of definition from engineer and natural sciences. In the perspective view of engineer, texture is the repetitive pattern. However, for natural sciences, texture is the tactile quality of the surface or the sense of touch. Botanists described leaf texture features as thin, waxy, soft, rough and so on. Therefore, we facing problem to describe the leaf texture based on the touching information but we able to describe the texture features based on the engineer's perspective which is the repetitive pattern. However, the other features detected and extracted based on the botanical knowledge. In this research, we have focused on venation detection which is one of the unique properties which could be used alone to detect the species.

For the features of leaf shape, it is distinguished based on the botanical knowledge. The terminology of the features is translated into the computer languages. For the features of lobes and sinuses, the numbers and the location of lobes and sinuses acted as the features of lobes and sinuses. For apex and base, the pattern of apex and base curvature is used to interpret them by using botanical features. Finally, the teeth pattern is used to represent the features of margin. Leaf venation is to find the features of the pattern of skeleton based on botanical features.

Our proposed method does not affect by the noise and the threshold which had been used by Canny Edge Detection. This is because we use Hue normalization and image fusion to filter out the noise and unrelated curved. The leaf venation is presented in a single dimensional for multiple curves which is easy to undergo analysis their characteristic. This all can be seen in Figure 1.

4.1 Leaf Venation Detection

The proposed approach consists of five main steps to extract the leaf venation as displayed in Figure 5. First of all, canny edge detection is used to detect the edge existed in the leaf samples like many other researchers did. However, the output image from canny edge detection is not that satisfactory, the detected edge not necessary is leaf veins. For example the leaf boundary is also the output of Canny Edge Detection, however, it is not leaf venation, so the second step is to remove the leaf veins. The output of the Canny edge

detection is in two dimensional, which it hard to analysis and the detected point is like scanning line by line which the lines does not form. Step 3 is to re-arrange the scatter point into curve which is 1-dimensional which is easy to do analysis later. Then step 4 hue normalization is to increase the contrast of leaf and leaf venation which is used in step 5 to distinguish which curve is leaf veins and which curve is fake curve for leaf venation. To improve the extracted venation, another 4 techniques are used which are remove leaf boundary, extract the venation in curve vector, finding hue normalization and finally fuse them.

Step 1: Canny edge detection

First of all, convert the sample image to grayscale image, then using canny edge detection with two different thresholds (lower threshold and higher threshold) to detect the existing edge in the sample image (Algorithm 1). The detected edge includes leaf venation, leaf boundary and unwanted false curve. The output image is denoted as *CannyImage*. However, many of the researchers use this to detect the leaf venation and without applying any further process to enhance the leaf venation image.

Algorithm 1: Canny Edge Detection (Step 1)

Input: Image

Output: Output image from Canny Edge Detection, *CannyImage*

Begin:

GrayImage \leftarrow *rgb2gray*(Image)

CannyImage \leftarrow *edge*(*GrayImage*, 'canny', [0.4 1])

End

Step 2: Remove leaf boundary

Binarized the sample image, found the edge of the leaf by using canny edge detection and named the output image as *LeafBoundary*. Since the *CannyImage* includes the leaf boundary which is not considered as leaf venation, therefore there is a need to exclude or remove the leaf boundary from *CannyImage* and named the desired output image as *boundaryOff_image*. To get *boundaryOff_image*, it can be obtained by finding the differences of *CannyImage* and *LeafBoundary* (equation 1 or 2).

$$\text{boundaryOff_image} = \text{CannyImage} - \text{LeafBoundary} \quad (1)$$

$$\text{boundaryOff_image} \in \text{CannyImage} \cap \text{LeafBoundary}^c \quad (2)$$

Where parameter of *boundaryOff_image* is the image where excluded the leaf boundary. In mathematics, it is called *CannyImage* complement *LeafBoundary*. This process can be summarised in Algorithm 2.

Algorithm 2: Remove leaf boundary (Step 2)

Input: Image

Output: Canny image complement leaf boundary, *boundaryOff_image*

Begin:

BinaryImage \leftarrow *im2bw(Image)*

LeafBoundary \leftarrow *edge(binary image, 'canny', [0.4 1])*

boundaryOff_image \leftarrow *CannyImage* – *LeafBoundary*

End

Step 3: Extract curve

The line and curve are a mix up together and display in the *boundaryOff_image* in two-dimensional form. In order to simplify the process of filtering the unwanted curve and line, there is a need to extract the line and curve from *boundaryOff_image*. The extracted line and curve are stored in the parameter of curve separately (Equation 3).

$$VenationCurve_i(X,Y) \in boundaryOff_image, \quad (3)$$

Where, $i = \{1, \dots, totalCurve\}$

All connected pixel in eight directions is saving in curve array separately. Each single curve stores its coordinate position in X-axis and Y-axis. Finally, the unconnected pixel are then forming $VenationCurve_i(X,Y)$. The parameter *totalCurve* is denoted as the total number of curve and line which is able to capture from *boundaryOff_image*. This process can be summarised in Algorithm 3.

Algorithm 3: Extract venation curve (Step 3)

Input: *boundaryOff_image*

Output: $VenationCurve_i(X,Y)$

Begin:

//this is a function and can be found in Algorithm 4.2

Call *extractCurve(boundaryOff_image)*

End

Step 4: Produce Hue normalization image

Most of the leaf is green because of the presence of chlorophyll. This molecule absorbs most of the energy of the light spectrum except green light, therefore, in the human eye, most of the leaves are in green colour. However, in some special case, in nature, some plant have others colour of leaves.

Most of the leaves veins have very close colour to their leaves. The intensity of leaves veins and leaves are very close too. Therefore, to better separate the leaves veins from leaves there is a need to undergo the process of **hue normalization**. Intensity range of hue is 0 to 255. However, if the leaf is green the colour range used only from 42.5 to 85 intensity, if the leaf colour is purple, the range is around 170 to 212.5 [24] , but the others intensity is no use. Therefore, hue normalization able to rescales the colour intensity values between leaves vein and leaves which able to

increase the contrast between the hue range of leaf lamina and leaf venation. After rescaling, the intensity distribution of leaves veins and leaves are more apparent as shown in Figure 6.

HSL, HSV and HIS colour space (Hue-Saturation-Lightness, Hue-Saturation-Value and Hue-Saturation-Intensity) are basically more human perceptive. According to Joblove and Greenberg [21], HSL model is based on the traditional colour making procedures, for example in painting activity, they mix up the brightly coloured pigment with various proportional of black or white to lighten or darken the colour. Therefore, in this research, the RGB (Red, Green and Blue) images are encoded to HSL (Hue-Saturation-Lightness).

RGB is a powerful colour model to represent a colour feature in the machine. However, this colour model is less human perceptive and this kind of colour modal does not follow the human vision system. The other colour model can be used to represent the colour feature such as YIQ, YUV, YCbCr are device dependent and colour model like Munsell, CIE XYZ, CIE L*U*V* and CIE L*a*b* are device independent but is less human perception. Table 1 described the colour model and its classification.

Hue (H) represents the dominants wavelength in the light spectrum that is able to perceive by the human visual system. Hue is usually measured in a degree of a circle which means 360 degrees. Hue ranges can be easily classified to red, green, blue, orange, yellow, indigo and so on. This colour attribute is unaffected by shadow caused by the light source. Therefore, Hue is suitable to use to segment the leaf venation and lamina. However, since the colour of leaf lamina and leaf venation are usually very near, the process of normalization to the hue intensity is needed to find the contrast of leaves and leaves venation. The computation is displayed in Algorithm 4 and in equation 4 to 12.

Algorithm 4 : Produce hue normalization image (Step 4)

Input: Image

Output: Hue normalization image

Begin:

$R \leftarrow \text{Image}(:, :, 1)$

$G \leftarrow \text{Image}(:, :, 2)$

$B \leftarrow \text{Image}(:, :, 3)$

$RGB_{max} \leftarrow \text{maximum value in RGB}$

$RGB_{min} \leftarrow \text{minimum value in RGB}$

$C \leftarrow \text{the difference of maximum value in RGB and minimum value in RGB}$

IF maximum value in RGB is R

$H \leftarrow (G-B)/C$

ELSE IF maximum value in RGB is G

$H \leftarrow 2 + (B-R)/C$

ELSE IF maximum value in RGB is B

$H \leftarrow 4 + (R-G)/C$

End IF

$$\begin{aligned}
Hue &\leftarrow 60 * H \\
Hue_{min} &\leftarrow \text{minimum value in Hue but not included } 0 \\
Hue_{max} &\leftarrow \text{maximum value in Hue but not included } 1 \\
Hue_{normalization} &\leftarrow (Hue - Hue_{min}) ./ (Hue_{max} - Hue_{min})
\end{aligned}$$

Display the image of $Hue_{normalization}$

End

$$M = \max(R, G, B) \quad (5)$$

$$m = \min(R, G, B) \quad (6)$$

$$C = M - m \quad (7)$$

$$H' = \begin{cases} \frac{G-B}{C}, & M = R \\ 2 + \frac{B-R}{C}, & M = G \\ 4 + \frac{R-G}{C}, & M = B \end{cases} \quad (8)$$

$$HSV_{hue} = 60 * H' \quad (9)$$

$$Hue_{min} = \min(HSV_{hue}), \text{ where } RangeHue_{min} \neq 0 \quad (10)$$

$$Hue_{max} = \max(HSV_{hue}), \text{ where } RangeHue_{max} \neq 1 \quad (11)$$

$$Hue_{normalization} = \frac{HSV_{hue} - Hue_{min}}{Hue_{max} - Hue_{min}} \quad (12)$$

Step 5: Image Fusion

Figure 6 shows a zoom image for certain part of the leaf to shows the normalized Hue intensity of HSL in leaves vein and leaves. This figure had proven that leaves venation have an apparent difference in intensity after applying normalization process. In human eyes, the leaf lamina and leaf venation can be easily differentiated. However, it is difficult to teach the machine to differentiate the leaf lamina and leaf venation.

First of all, the intensity range for leaf lamina and leaf venation is computed. The intensity range for leaf venation in hue normalization image should be different. Both of them had their own intensity range. However, the intensity range of these two parts varies from plant species. Therefore, the easy way to find their intensity range for leaf lamina and leaf venation is to find the intensity of their longest straight curve from step 3. In this research, it is assumed that the longest straight curve is the primary leaf vein. Equation 13 to 14 shows the equation to find the primary vein.

$$IndexLV = \text{index}(\max(\text{length}(\text{VenationCurve}))) \quad (13)$$

$$LVein(X, Y) = \text{VenationCurve}_{IndexLV}(X, Y) \quad (14)$$

The intensity range of the longest straight curve (primary vein), $LVein(X, Y)$ can be extracted by finding the minimum and maximum intensity of the longest venation. The minimum intensity found in longest venation is the lowest intensity in

the range for leaf venation, RV_l (equation 15). In reverse, the maximum intensity found in longest venation is the highest intensity range for leaf venation, RV_h (equation 16).

$$RV_l = \min(Hue_{normalization}(Lvein(X), Lvein(Y))) \quad (15)$$

$$RV_h = \max(Hue_{normalization}(Lvein(X), Lvein(Y))) \quad (16)$$

The parameter MV_i of each curve are obtained by finding the mean of intensity for curve pixel in $hsv_{normalization}$ image. Then checks whether the MV_i for curve ($VenationCurve_i$) is in range between lowest intensity for leaf venation, RV_l and highest intensity range for leaf venation, RV_h . If they are in range, the curve is kept as valid venation, otherwise the curve is being removed.

Algorithm, 5: Image fusion (step 5)

Input: Venation Curve

Output: Enhanced leaf venation

Begin:

IndexLV \leftarrow find which venation curve is longest straight curve and get its index

Lvein(X, Y) \leftarrow finding the X and Y axis coordinate in longest Venation Curve

//finding the lowest intensity range for leaf venation

$RV_l \leftarrow \min(Hue_{normalization}(Lvein(X), Lvein(Y)))$

//finding the highest intensity range for leaf venation

$RV_h \leftarrow \max(Hue_{normalization}(Lvein(X), Lvein(Y)))$

For each venation curve

$MV_i \leftarrow$ finding the average intensity for VenationCurve

IF the average MV_i is within RV_l and RV_h

This curve is valid venation and store it

Else

This curve is invalid and eliminate

End IF

End For

Display image of the deformation of valid leaf venation

End

4.2 Leaf Venation Feature

To localize the edge direction efficiently, the lines that extracted from pre-processing, are further divided into smaller segmented lines called L . Every line segment is bounded by two distinct endpoints with ten continuous pixels linked between starting point, P_{start} and ending point, P_{end} . L is a subset of B , $L \in B$ and L is a line segment which parameterized as Equation 17. The parameters t should be natural number which is a whole number that is positive number but excluded zero. However, in this situation, the t value is 10.

$$L = \{u + tu \mid t > 0, t \in N\}, \text{ where } N = \{1, 2, 3, \dots\} \quad (17)$$

The parameter u represents the starting point, P_{start} of line segment and $u + tu$ represents the ending point P_{end} of line segment. Above statement is presented as Equation 18 and 19.

$$P_{start} = u \quad (18)$$

$$P_{end} = u + tu \quad (19)$$

In this case, line segment is assumed as the hypotenuse of right triangle as presented in Figure 7. The line \overrightarrow{AB} represents the line segment, point A represents the start point of line segment, P_{start} and B denoted the end point of the line segment, P_{end} . However, the dash line \overrightarrow{AC} denotes the horizontal distance and the dash line \overrightarrow{CB} represents the vertically distance. The angle of the line segment ($angleL$) can be obtained by using Equation 20.

The directional line segment encodes in eight directional which is $Angle_i, i = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$ or equivalent to $Radian_i, i = \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi, \frac{5\pi}{4}, \frac{3\pi}{2}, \frac{7\pi}{4}\}$. However, for the line segment which is bended, arbitrary and deviated from being straight line is denoted as curve or strands and it is non-directional line segment. In total, the line segment is grouped into nine directional unit which are $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ$ and *curve*. The traditional edge histogram is excluded from the existence of curve. However, in leaf venation especially the tertiary venation it presence in strand condition.

The line segment which lay along starting point P_{start} and ending point P_{end} called directional line segment. Conversely, the non-directional lines do not lie straight but diverged and deviated. Non-directional lines are then grouped in curve unit. Equation 20 is used to determine whether the line segments are directional lines or non-directional lines.

$$lineType(D) = \begin{cases} \text{directional line, } D < Threshold \\ \text{non-directional line, } D \geq Threshold \end{cases} \quad (20)$$

Where,

$$D = \sum_{i=0}^N |Y_i - [slope * (X_i - P_{start}(X)) + P_{start}(Y)]|$$

$$lineType(D) = \begin{cases} \text{directional line, } D < Threshold \\ \text{non-directional line, } D \geq Threshold \end{cases}$$

The frequency and the number of occurrence of each directional unit are computed. Then convert them into a binary pattern. If the frequency of each directional unit is greater than a predefined threshold it is 'true' or else it is 'false' as presents in Equation 21.

$$H_B(x) = \begin{cases} 0, & x < thres \\ 1, & x \geq thres \end{cases} \quad (21)$$

Figure 8 compares the angle in a circle with the matrix in angle form and angle code. The example of binary edge matrix is presented in Figure 9 (a) and (b) is the binary pattern filter which is used to convert the binary edge matrix to the binary directional pattern.

Equation 22 shows the computation to convert the binary edge matrix to binary direction pattern.

$$\text{BiDirP} = \sum_{j=1}^{D_c} \sum_{i=1}^{D_r} (MA_{ij} \times MB_{ij}) \quad (22)$$

Where,

$$i = \{1, 2, 3, \dots, D_r\}, j = \{1, 2, 3, \dots, D_c\}$$

The result of BiDirP is the summation of multiplication of each in Matrix MA and matrix MB . This result is used to determine the leaves venation state. Table 2 shows the venation type with the reasonable index of BiDirP. The calculation of index of BiDirP is based on the angle of leaf veins. For example, the parallel veins, the veins is 0 degree so the angle code for the parallel veins is 2^2 . For the pinnately veins, the veins had one primary veins which is 0 degree (2^2) and two secondary which their angle codes are 2^1 and 2^3 , therefore the index for BiDirP is the summation of 2^1 , 2^2 and 2^3 .

5 Result and Discussion

This section presents the comparison between venation detection by using Canny Edge Detection and proposed a method. The proposed method outperformed the traditional Canny Edge Detection. We have employed the presented method on both Flavia and Acer datasets with 32 samples. Some of the characteristic of leaf venation in Flavia dataset are described in Figure 10. Figure 11 presents the output image of some venation of Acer dataset.

As can be observed in both Figure 10 and 11, the accuracy of venation detection is remarkable free from comparing to previous ones. To reveal the accuracy Canny method is shown on the left side of each sample. Our results are shown in the right-hand side of each pair images. As an example in Figure 10, top row in the middle which is labelled with *Kalopanax septemlobus* (Thunb. ex A.Murr.) Koidz, our method shows a perfect curacy free from any edge and additional required data.

Figure 12 summarized these steps and shows how the proposed method achieved the goal. Canny Edge Detection detects the edge of the image by the intensity changes. However, most of the intensity changes does not mean that, there are leaf venation. First of all, leaf boundary is not the leaf venation, so step 2 is to remove the leaf boundary. The output of Canny edge detection is 2 dimensional, which is difficult to do analysis. Therefore, we extract the curve and make it in one dimensional. Most cases, leaf are constructed by only small range of intensity especially the changes of Hue. So we used Hue normalization to make the leaf and the venation high contrast. This step makes it possible to differentiate the leaf and its venation easily. Image fusion is used to determine whether the curve is leaf veins or unwanted curves, then eliminate the unwanted curves.

The proposed method outperformed the other previous works in term of the number of the dataset used and the accuracy as more external leaf structure features are applied. Table 3 shows that the more features that are integrated into leaf identification, the higher the

accuracy are. However, the selection of features plays the key role. Besides that, 10-fold cross validation is applied to reduce the overfitting and improve the accuracy. There is 99.3 percent of accuracy is achieved using the proposed method. The accuracy equation is as shown as below,

$$Accuracy = \frac{\text{total number of query}}{\text{relevant number of images}} \quad (23)$$

The obtained results are outperforming than the previous works as the previous works are lacking botanical knowledge. The obtained features are unreliable and not worthy. For example for leaf shape, the length, diameter, width and so on of the leaf should not be used as features to recognize the leaf. The age of the leaf may influence the result. The other features such as the ratio of the venation pixel versus leaf area pixel are considering as unreliable too, as the detected venation using different methods and a different magnitude of thresholds provided different answers. By using botanical features, the actual methods to recognize the leaf provided accurate features.

Because the dataset used is a collection of the Acer genus samples leaves from many others dataset, therefore, there are no previous works found on them. However, previous methods are applied on the dataset to compare with the proposed method. The comparison is shown in Table 4. The proposed method still outperformed other previous works as the accuracy achieved is at 91.6 percent.

Previous works may give good results to the Flavia dataset which are form from higher taxa's plant, however, when it is from the same genus, the results are not promising. Acer dataset form from the plant species of the same genus, therefore, their similarity is very high. The previous work unable to distinguish them. By embedded the botanical features, the achievement in identifying the plant species is promising.

Conclusion

To preserve of ecosystems, plant conservation is an urgent necessity in the current area. This research proposed a methodology to identify the plant species using leaf venation detection based on botanical knowledge. Edge detection, leaf boundary removal, extracting curve, producing hue normalization image and image fusion are five steps which completed our approach to identify the species. Our experimental results in two different datasets demonstrated the effectiveness of the presented method for shape recognition with high accuracy. An accurate comparison between presented method and existing ones shows the effectiveness of our method. We believe that this idea will seriously affect on plant conservation and consequently preserve the ecosystems.

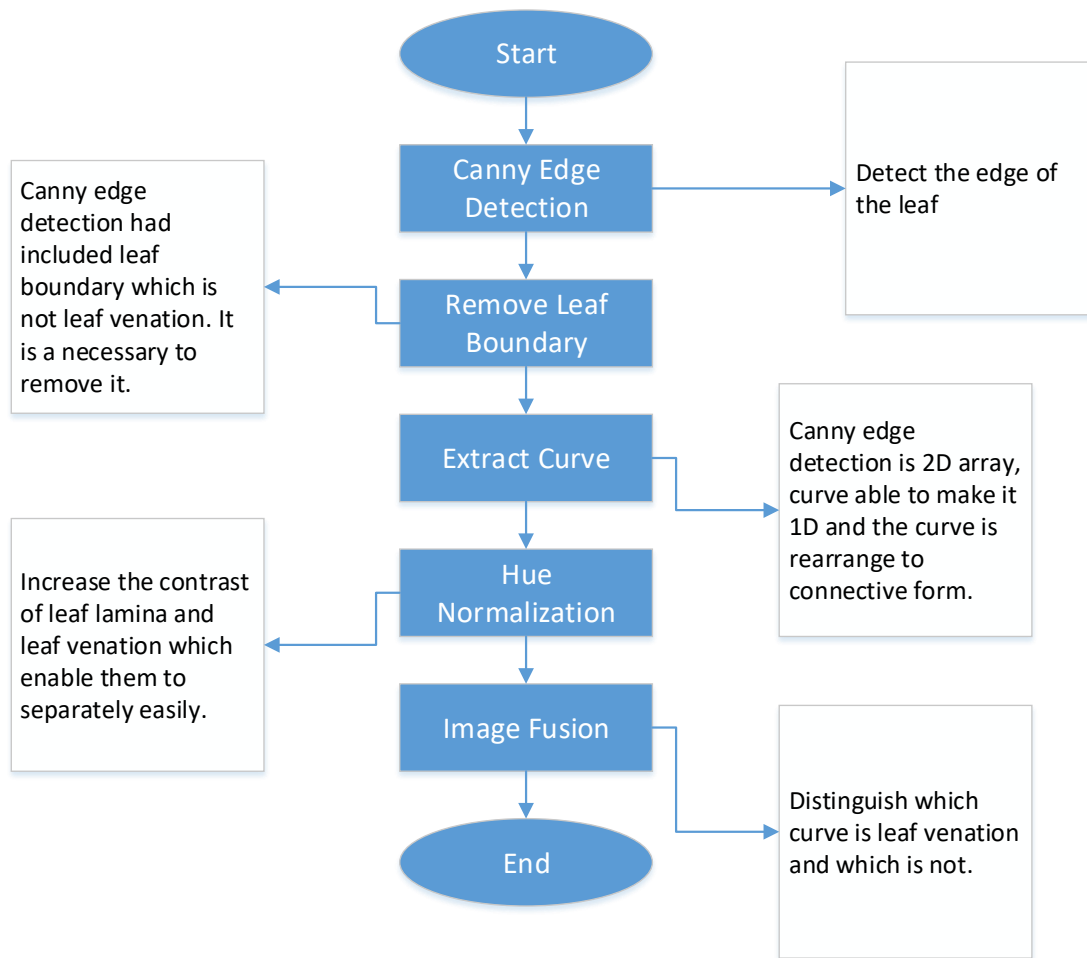


Figure 1: Block diagram for the proposed method

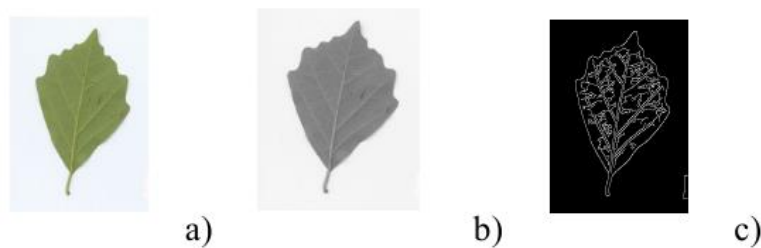


Figure 2: Leaf image retrieved from Amlekar et al. [4]. (a) Original image (b) grey image (c) leaf venation



Figure 3: Venation extracted by Mouine et al. [15][16] (a) SC0, margin points (b) SC1, salient points (c) SC2, salient point excluded margin points.

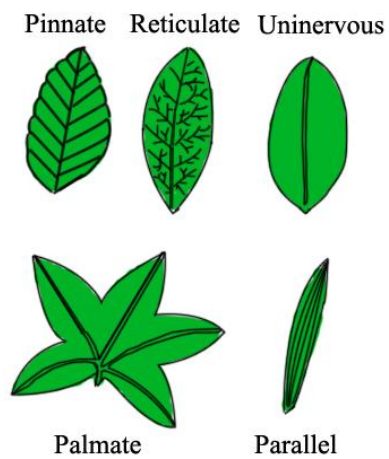


Figure 4: Example pattern for leaf veins

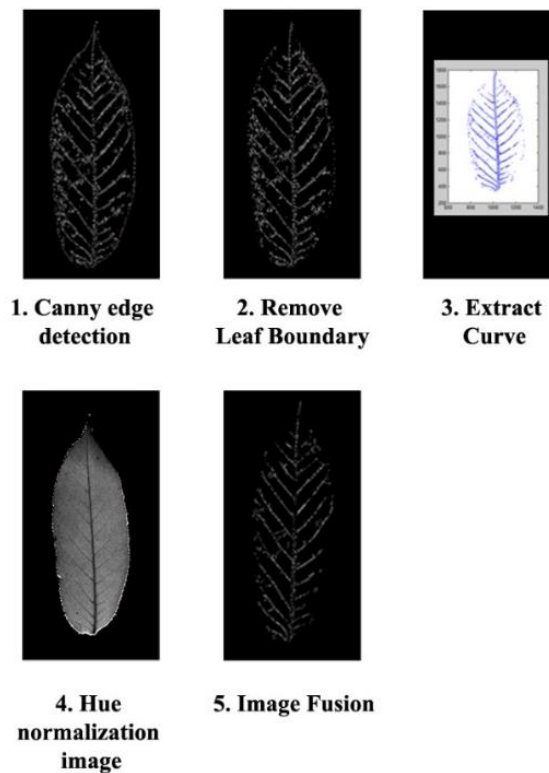


Figure 5: Leaf venation extraction steps

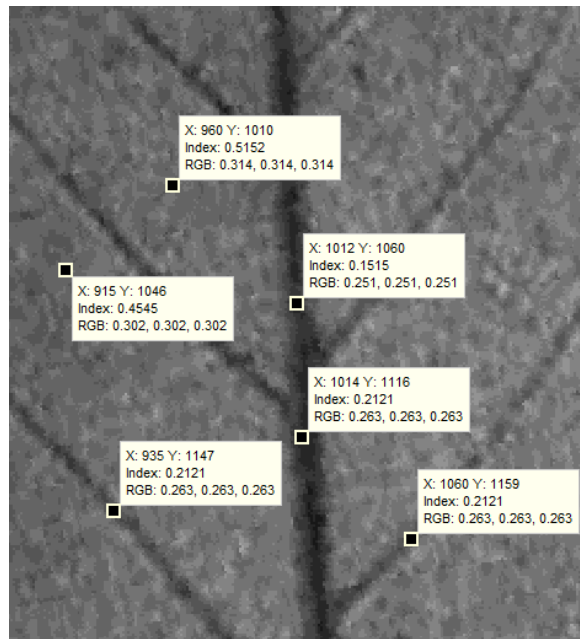


Figure 6: Hue intensity of leaves vein and leaves after normalization process

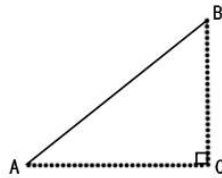


Figure 7: Line segment act as hypotenuse of right triangle

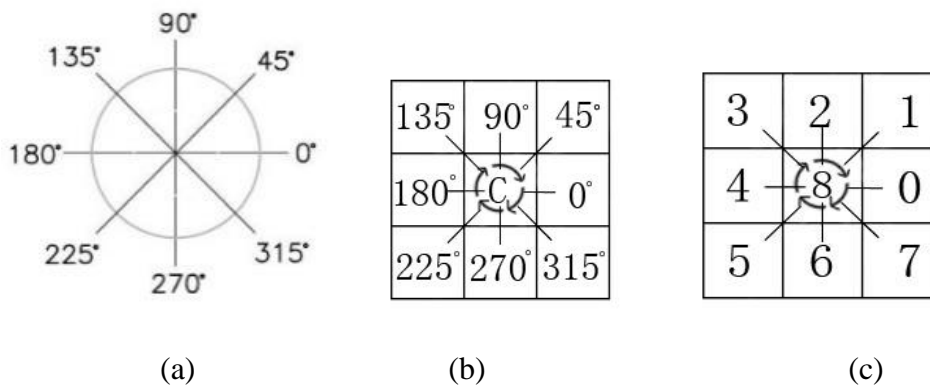


Figure 8: (a) Angle presented in circle (b) Angle presented in matrix form (c) Angle code presented in matrix form according to its angle or directional

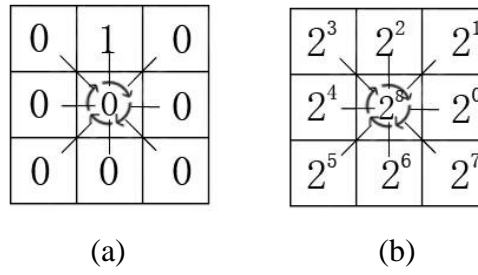


Figure 9: (a) Binary Edge Matrix, MA (b) Binary Pattern Filter, MB

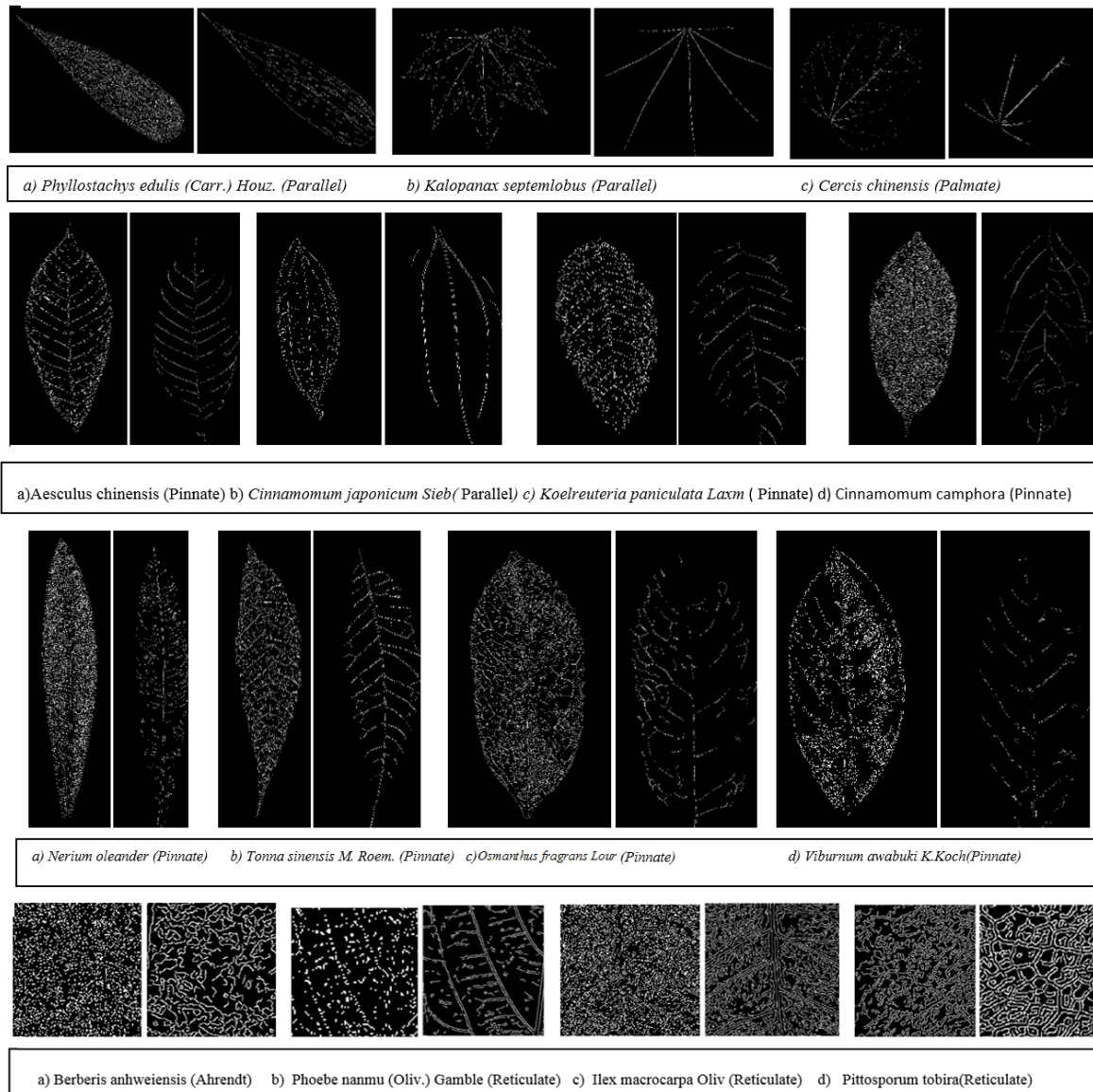


Figure 10: Leaf venation detection in Flavia dataset

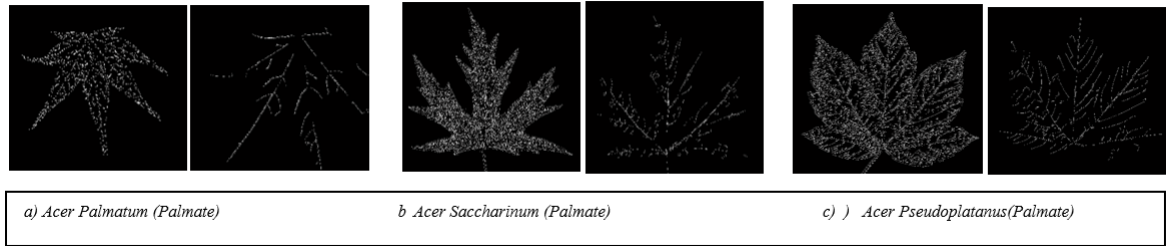


Figure 11: Leaf venation detection in Acer dataset

Table 1: Color models classifications, Source from Ibrahim, Selim and Zayed [23].

Color Model	Classifications
Munsell	Device independent
RGB, CMY(K)	Device dependent
YIQ, YUV, YCbCr	Device dependent
HIS, HSV, HSL	User oriented-Device dependent
CIE XYZ, CIE L*u*v*, CIE L*a*b*	Device independent, color metric

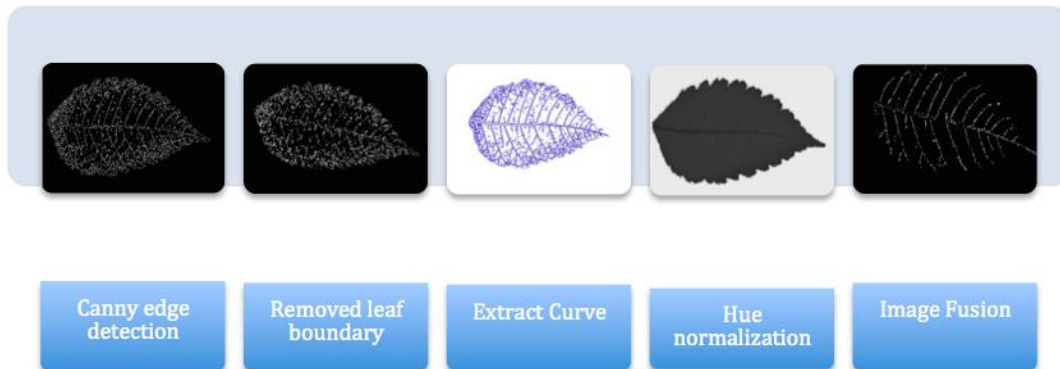


Figure 12: Output at each and every step.

Table 2: Matching table for venation type with its BiDirP index

Venation type	Index of BiDirP
Parallel	4
Pinnate	14
Palmate	31
Reticulate	>256
Uninervous	0

Table 3: Comparison between our method and previous methods using Flavia dataset

Author	Features	Accuracy	Dataset	Training	Testing	Species
Satti, Satya and Sharma (2013)	Shape, colour	93.3%	1907	1742	165	33
Chaki, Parekh and Bhattacharya (2015)	Shape, texture	87.1%	930	620	310	31
Arun, Emmanuel and Durairaj (2013)	Texture	94.7%	250	175	75	5
Wu et al. (2007)	Shape, veins	90.0%	1800	1800	320	32
Kadir et al. (2013b)	Shape, colour, vein, texture	93.4%	1600	1280	320	32
Our proposed method	venation	99.06%	1907	1280	627	32

Table 4: Comparison proposed method with previous methods using Acer dataset

Author	Features	Accuracy	Dataset	Training	Testing	Species
Wu et al. (2007)	Shape, veins	37.3%	600	450	150	32
Kadir et al. (2013b)	Shape, colour, vein, texture	62.0%	600	450	150	32
Arun, Emmanuel and Durairaj (2013)	Texture	43.3%	600	450	150	32
Our proposed method	venation	91.6%	600	450	150	32

References

- [1] Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Yahiaoui, I., Selmi, S., Carré, J., Barthélémy, D., Boujemaa, N. and others (2013) "PI@ ntnet mobile app," in Proceedings of the 21st ACM international conference on Multimedia, pp. 423–424.
- [2] Mzoughi, O., Yahiaoui, I., Boujemaa, N. and Zagrouba, E. (2013) "Advanced tree species identification using multiple leaf parts image queries," in Image Processing (ICIP), 2013 20th IEEE International Conference on, pp. 3967–3971.
- [3] Plant Taxonomy 2E (2009). McGraw-Hill Education (India) Pvt Limited. https://books.google.com.my/books?id=_px_WAWHiZIC.
- [4] Amlekar, M., Manza, R. R., Yannawar, P. and Gaikwad, A. T. (2014) "Leaf features based plant classification using artificial neural network," IBMRD's Journal of Management & Research, 3(1), pp. 224–232.

- [5] Casanova, D., Backes, A. R. and Bruno, O. M. (2013) "Pattern recognition tool based on complex network-based approach," in Journal of Physics: Conference Series, p. 012048.
- [6] Kadir, A., Nugroho, L. E., Susanto, A. and Santosa, P. I. (2013b) "Leaf classification using shape, color, and texture features," arXiv preprint arXiv:1401.4447.
- [7] Kadir, A., Nugroho, L. E., Susanto, A. and Santosa, P. I. (2013c) "Neural network application on foliage plant identification," arXiv preprint arXiv:1311.5829.
- [8] Kadir, A., Nugroho, L. E., Susanto, A. and Santosa, P. I. (2011b) "Foliage plant retrieval using polar Fourier transform, color moments and vein features," arXiv preprint arXiv:1110.1513.
- [9] Fern, B.M., Rahim, M.S.M., Saba, T., Almazyad, A.S., Rehman, A. (2017) Stratified classification of plant species based on venation state Biomedical Research, vol. 28(13), pp.5660-5663.
- [10] Mishra, M. K., Dandamudi, P., Nayani, S. P., Munikoti, S. S., Chelukunda, S. S. and others (2011) "Variability in stomatal features and leaf venation pattern in Indian coffee (*Coffea arabica* L.) cultivars and their functional significance," Botanica Serbica (Serbia).
- [11] Pornpanomchai, C., Supapattranon, C. K. and Siriwisesokul, N. (2011) "Leaf and flower recognition system (e-Botanist)," International Journal of Engineering and Technology. IACSIT Press, 3(4), p. 347.
- [12] Rolland-Lagan, A.-G., Amin, M. and Pakulska, M. (2009) "Quantifying leaf venation patterns: two-dimensional maps," The Plant journal : for cell and molecular biology, 57(1), pp. 195–205. doi: 10.1111/j.1365-313X.2008.03678.x.
- [13] De Oliveira Plotze, R. and Bruno, O. M. (2009) "Automatic leaf structure biometry: computer vision techniques and their applications in plant taxonomy," International Journal of Pattern Recognition and Artificial Intelligence. World Scientific, 23(02), pp. 247–262.
- [14] Syahputra, H., Harjoko, A., Wardoyo, R. and Pulungan, R. (2014) "Plant recognition using stereo leaf image using gray-level co-occurrence matrix," Journal of Computer Science. Science Publications, 10(4), p. 697.

[15] Mouine, S., Yahiaoui, I. and Verroust-Blondet, A. (2012) "Advanced shape context for plant species identification using leaf image retrieval," in Proceedings of the 2nd ACM international conference on multimedia retrieval, p. 49.

[16] Mouine, S., Yahiaoui, I. and Verroust-Blondet, A. (2013b) "Combining leaf salient points and leaf contour descriptions for plant species recognition," in Image Analysis and Recognition, pp. 205–214.

[17] Jobin, A., Nair, M. S. and Tatavarti, R. (2012) "Plant Identification Based on Fractal Refinement Technique (FRT)," Procedia Technology. vol. 6, pp. 171–179.

[18] Cope, J. S., Remagnino, P., Barman, S. and Wilkin, P. (2010b) "The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms," in Advanced Concepts for Intelligent Vision Systems, pp. 135–144.

[19] Li, Y., Chi, Z. and Feng, D. D. (2006) "Leaf vein extraction using independent component analysis," in Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on, pp. 3890–3894.

[20] Casanova, D., Backes, A. R. and Bruno, O. M. (2013) "Pattern recognition tool based on complex network-based approach," in Journal of Physics: Conference Series, p. 012048.

[21] Joblove, G. H., & Greenberg, D. (1978, August). Color spaces for computer graphics. In ACM siggraph computer graphics (Vol. 12, No. 3, pp. 20-25). ACM.

[23] Ibrahim, N. B., Selim, M. M. and Zayed, H. H. (2012) "A dynamic skin detector based on face skin tone color," in Informatics and Systems (INFOS), 2012 8th International Conference on, pp. MM–1.

[24] WorkWithColor.com. "Cyan." Accessed December 27, 2017. <http://www.workwithcolor.com/cyan-color-hue-range-01.htm>.

