



A New Leaf Venation Detection Technique for Plant Species Classification

Hoshang Kolivand¹ · Bong Mei Fern² · Tanzila Saba³ · Mohd Shafry Mohd Rahim⁴ · Amjad Rehman⁵

Received: 24 October 2017 / Accepted: 30 July 2018 / Published online: 11 August 2018
© King Fahd University of Petroleum & Minerals 2018

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 average accuracy is obtained by 98.6 and 89.83% 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 of all any others living organisms. Since the 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 characteristics of leaves are derived to identify the plant species.

Performance of leaf recognition is highly depending on the quality and trustworthiness of the retrieval feature [2,3]. However, it is an arduous task to acquire meaningful and unique feature from a low variation plant species leaves [4–6]. It is more exhausting when implementing the botanical knowledge in computer vision. Traditional plant identification uses descriptions in the 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 [7]. Based on [8], the information derived from some significant leaf parts such as leaf apex, base, margin and venation is still not well exploited although 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 [9], not less than 50 units of characters are used for 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 could be described in the distinctive or char-

✉ Tanzila Saba
tsaba@psu.edu.sa

Bong Mei Fern
bongmf@utar.edu.my

- ¹ Department of Computer Science, Liverpool John Moores University, Liverpool L3 3AF, UK
- ² Department of Computer Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Bandar Sungai Long, Cheras, Kajang, Selangor, Malaysia
- ³ College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia
- ⁴ Media and Games Innovation Centre of Excellence (MaGIC-X) UTM-IRDA Digital Media Centre, Institute of Human Centred, University Industry Research Laboratory (UIRL), Universiti Teknologi Malaysia (UTM), Skudai, Johor, Malaysia
- ⁵ College of Computer and Information Systems, Al Yamamah University, Riyadh 11512, Saudi Arabia

acteristic state. The morphological character of plant leaves describes the multi-part of leaves such as leaf venation, shape, margin. However, character states for leaf venation are parallel, palmately, pinnately, and so on.

2 Related Work

For venation, Amlekar et al. [10], Casanova et al. [11], Kadir et al. [12–14], Fern et al. [4], Mishra et al. [15], Pornpanomchai et al. [16], Rolland-Lagan et al. [17] and De Oliveira Plotze and Bruno [18] 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. [10] 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 and 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 1 shows the research image extracted from their report.

The method proposed by Syahputra et al. [19] was slightly out of the ordinary. Most of the researchers only had a single input image. However, the author proposed using double input images 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. [20,21] 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 2 shows the venation network, retrieved from Mouine et al. [20].

Jobin et al. [22] 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. [15] 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. [23] 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 whether the pixel is included in leaf veins or not. This algorithm is complicated, time-consuming and heavy computation

Li et al. [24] 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.

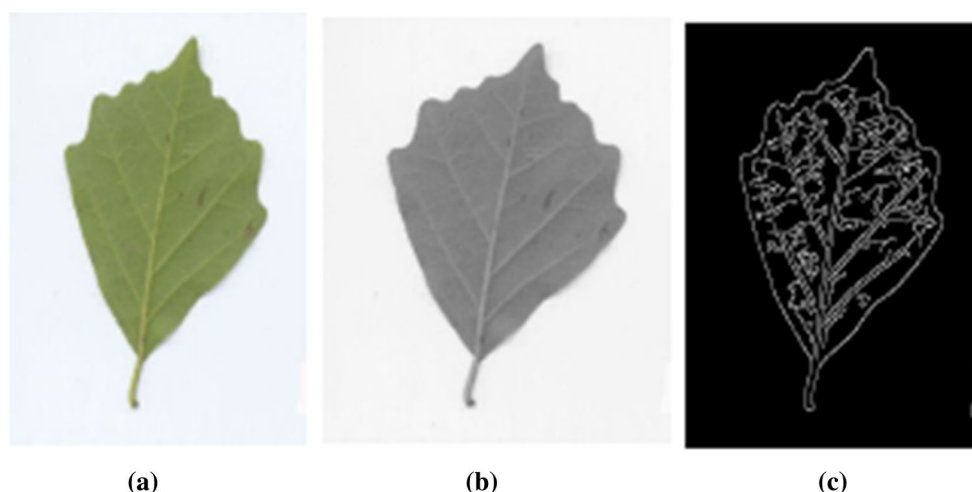


Fig. 1 Leaf image retrieved from Amlekar et al. [10]. **a** Original image, **b** grey image, **c** leaf venation

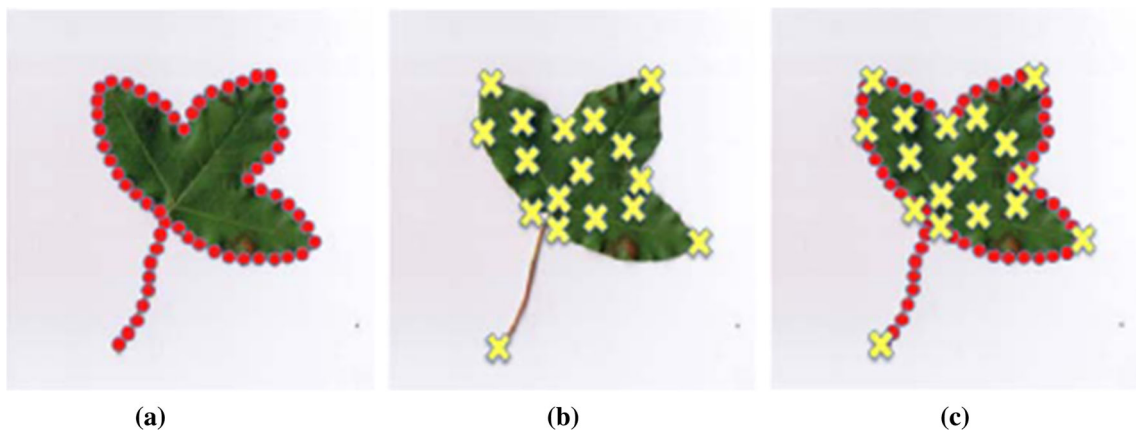


Fig. 2 Venation extracted by Mouine et al. [20,21]. a SC0, margin points, b SC1, salient points, c SC2, salient point excluded margin points

3 Leaf Venation

Leaf venation is the skeleton of leaf, and the examples of leaf venation are shown in Fig. 3. 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. [10] 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 a 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. [20,21] used advanced shape context SC0 to SC2 to find the salient points to represent the venation networks. Then, they searched and queried

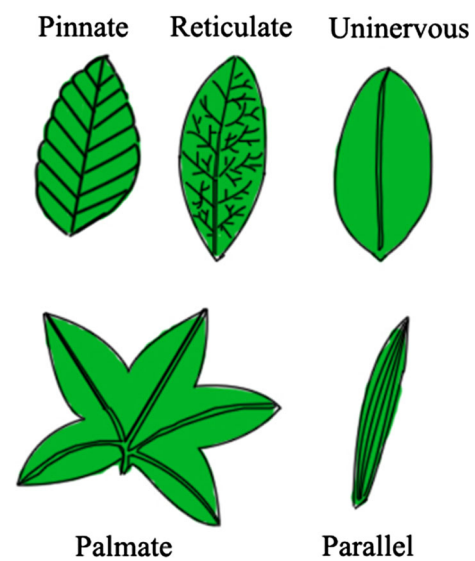


Fig. 3 Example pattern for leaf veins

the features of salient points by using locality-sensitive hashing (LSH). This method finds the distance and distinguishes between two features and is quite a time-consuming method. The plant identification method of Mouine et al. [20] 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. [21]).

Casanova, Backes and Bruno [11] normalized the vein network to ensure the venation is independent of scale venation. Then the geometrical data of venation were 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 the entire leaf; however, Mzoughi et al. [8] split the entire leaf

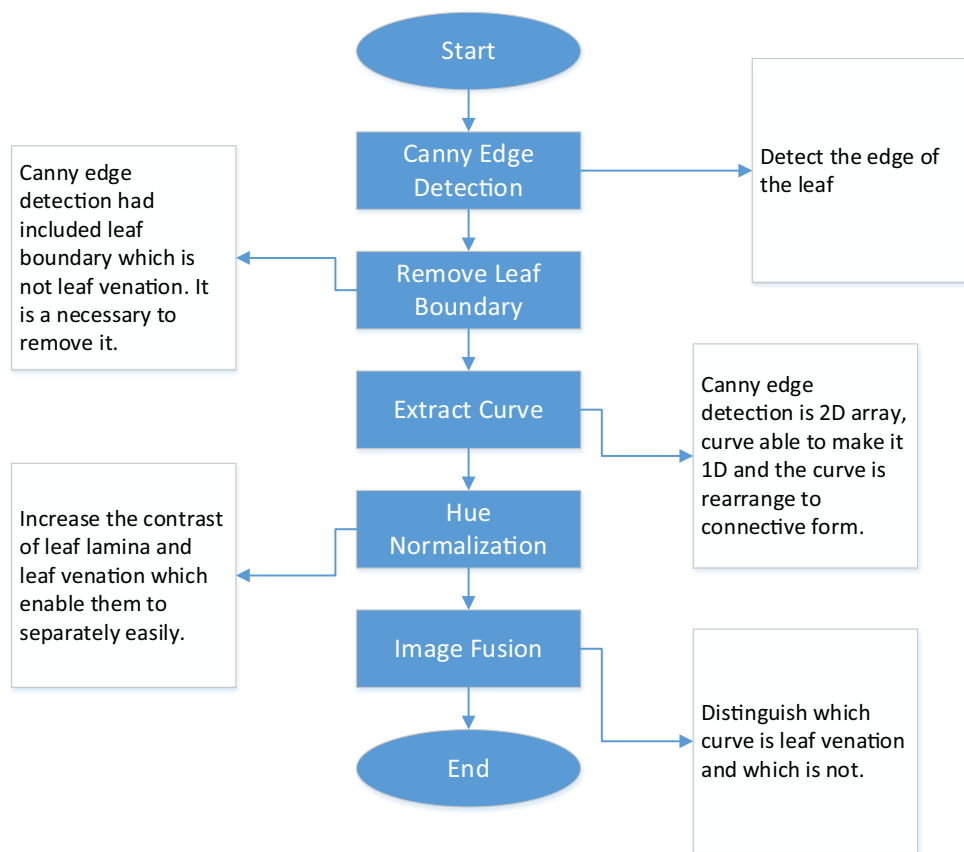


Fig. 4 Block diagram for the proposed method

into the 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 [25], the texture had several branches of definition from engineer and natural sciences. In the perspective view of an engineer, the texture is the repetitive pattern. However, for natural sciences, the 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 face

problems to describe the leaf texture based on the touching information but we are 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. The focus of research is leaf venation extraction, and research in [5] deals on leaf venation feature analysis. However, in [6], we use another feature such as leaf margin, leaf apex, base, lobe and sinus to identify the plant 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.

The 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 dimension for multiple curves which are easy to undergo analysis of their characteristic. This all can be seen in Fig. 4.

4.1 Leaf Venation Detection

The proposed approach consists of five main steps to extract the leaf venation as displayed in Fig. 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, and 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 dimensions, 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 rearrange the scatter point into a curve which is one dimensional which is easy to do the 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 a fake curve for leaf venation. To improve the extracted venation, another four techniques are used to 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 greyscale 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 the leaf boundary

Binarize the sample image, find the edge of the leaf by using canny edge detection and name the output image as *LeafBoundary*. Since the *CannyImage* includes the leaf boundary which is not considered as leaf venation, there

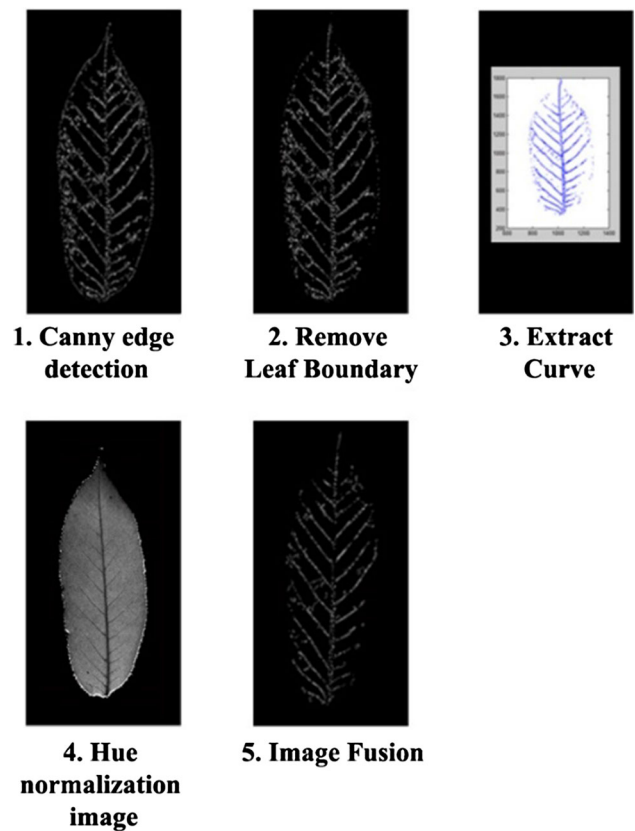


Fig. 5 Leaf venation extraction steps

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 between *CannyImage* and *LeafBoundary* (Eqs. 1 or 2).

$$boundaryOff_image = CannyImage - LeafBoundary \quad (1)$$

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

where the parameter of *boundaryOff_image* is the image where excluded the leaf boundary. In mathematics, it is called *CannyImage* complement *LeafBoundary*. This process can be summarized 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 the curve separately (Eq. 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 is 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 summarized 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 is 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 0–255. 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 170–212.5 [26], but the intensity of the other is no use. Therefore, hue normalization is able to rescale the colour intensity values between leaves vein and leaves which is 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 is more apparent as shown in Fig. 6.

HSL, HSV and HIS colour space (hue-saturation-lightness, hue-saturation-value and hue-saturation-intensity) are basically more human perceptible. According to Joblove and Greenberg [27], HSL model is based on the traditional colour making procedures; for example, in painting activity, they mix up the brightly coloured pigment with various propor-

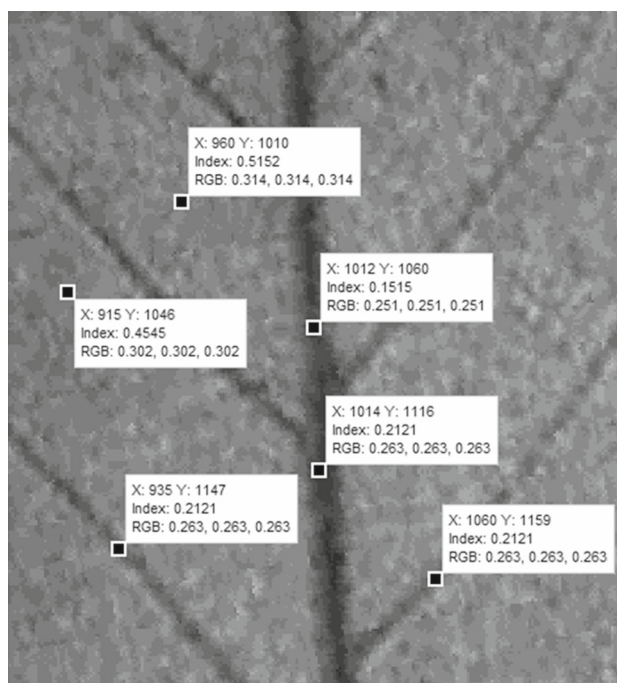


Fig. 6 Hue intensity of leaves vein and leaves after the normalization process

Table 1 Colour model classifications. Source from Ibrahim et al. [28]

Colour 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, colour metric

tional 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 perceptible and this kind of colour modal does not follow the human vision system. The other colour model can be used to represent the colour features such as YIQ, YUV, YCbCr that are device dependent and colour models like Munsell, CIE XYZ, CIE L*U*V* and CIE L*a*b* are device independent but is less human perception. Table 1 describes 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°. Hue ranges can be easily classified as 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 is 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 Eqs. 4–11.

Step 5: Image Fusion

Figure 6 shows a zoom image for a 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 the normalization process. In human eyes, the leaf lamina and leaf

Algorithm 4 : Produce hue normalization image (Step 4)

Input: Image

Output: Hue normalization image

Begin:

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

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

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

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

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

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

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

$\text{Hue} \leftarrow 60 * H$

$\text{Hue}_{min} \leftarrow \text{minimum value in Hue but not included 0}$

$\text{Hue}_{max} \leftarrow \text{maximum value in Hue but not included 1}$

$\text{Hue}_{normalization} \leftarrow (\text{Hue} - \text{Hue}_{min}) ./ (\text{Hue}_{max} - \text{Hue}_{min})$

Display the image of $\text{Hue}_{normalization}$

End

$$M = \max (R, G, B) \tag{4}$$

$$m = \min (R, G, B) \tag{5}$$

$$C = M - m \tag{6}$$

$$H' = \begin{cases} \frac{G-B}{C}, & M = R \\ 2 + \frac{B-R}{C}, & M = G \\ 4 + \frac{R-G}{C}, & M = B \end{cases} \tag{7}$$

$$HSV_{hue} = 60 * H' \tag{8}$$

$$\text{Hue}_{min} = \min (HSV_{hue}), \text{ where } \text{RangeHue}_{min} \neq 0 \tag{9}$$

$$\begin{aligned} \text{Hue}_{max} &= \max (HSV_{hue}), \text{ where } \text{RangeHue}_{max} \neq 1 \end{aligned} \tag{10}$$

$$\text{Hue}_{normalization} = \frac{HSV_{hue} - \text{Hue}_{min}}{\text{Hue}_{max} - \text{Hue}_{min}} \tag{11}$$

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. Equations 12–13 show the equation to find the primary vein.

$$\begin{aligned} \text{IndexLV} &= \text{index} (\max (\text{length} (\text{VenationCurve}))) \end{aligned} \tag{12}$$

$$LVein(X, Y) = VenationCurve_{IndexLV}(X, Y) \quad (13)$$

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 (Eq. 14). In reverse, the maximum intensity found in longest venation is the highest intensity range for leaf venation, RV_h (Eq. 15).

$$RV_l = \min(Hue_{normalization}(Lvein(X), Lvein(Y))) \quad (14)$$

$$RV_h = \max(Hue_{normalization}(Lvein(X), Lvein(Y))) \quad (15)$$

The parameter MV_i of each curve is obtained by finding the mean of intensity for curve pixel in $hsv_{normalization}$ image. Then it 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 a 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 Eq. 16. The parameters t should be natural number which is a natural positive number excluded zero. However, in this situation, the t value is 10.

$$L = \{u + tu \mid t > 0, t \in N\}, \text{ where } N = \{1, 2, 3, \dots\} \quad (16)$$

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 Eqs. 17 and 18.

$$P_{start} = u \quad (17)$$

$$P_{end} = u + tu \quad (18)$$

In this case, line segment is assumed as the hypotenuse of right triangle as presented in Fig. 7. The line \overrightarrow{AB} represents the line segment, point A represents the start point of line seg-



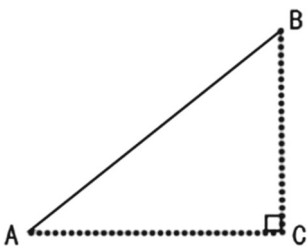


Fig. 7 Line segment acts as the hypotenuse of a right triangle

ment, 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 Eq. 19.

The directional line segment encodes in eight directions which are $Angle_i, i = \{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$ 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 it is denoted as curve or strands and it is non-directional line segment. In total, the line segment is grouped into nine directional units 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 tertiary veins its always present in the reticulate or scalariform pattern which is unable to determine using the traditional histogram.

The line segment which lays along starting point P_{start} and ending point P_{end} is 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 19 is used to determine whether the line segments are directional lines or non-directional lines.

$$lineType(D) = \begin{cases} directional\ line, & D < Threshold \\ non-directional\ line, & D \geq Threshold \end{cases} \quad (19)$$

where,

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

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

The frequency and the number of occurrence of each directional unit are computed and then converted 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 presented in Eq. 20.

$$H_B(x) = \begin{cases} 0, & x < thres \\ 1, & x \geq thres \end{cases} \quad (20)$$

Figure 8 compares the angle in a circle with the matrix in angle form and angle code and we called it as MA. The example of binary edge matrix presented in Fig. 9a, b is the binary pattern filter which is used to convert the binary edge matrix to the binary directional pattern and we call it as MB. Equation 21 shows the computation to convert the binary edge matrix to binary direction pattern.

$$BiDirP = \sum_{j=1}^{D_c} \sum_{i=1}^{D_r} (MA_{ij} \times MB_{ij}) \quad (21)$$

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 the index of BiDirP is based on the angle of leaf veins. For example, for the parallel veins, the veins are 0° so the angle code for the parallel veins is 2^2 . For the pinnate veins, the veins had one primary vein which is 0° (2^2) and two secondaries 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 . For more information, you can refer our paper published in [5].

5 Results and Discussion

This section presents the comparison between venation detection 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 characteristics of leaf venation in Flavia dataset are described in Fig. 10. Figure 11 presents the output image of some venation of Acer dataset.

As can be observed in both Figs. 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 of images. As an example in Fig. 10, the 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.

Fig. 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

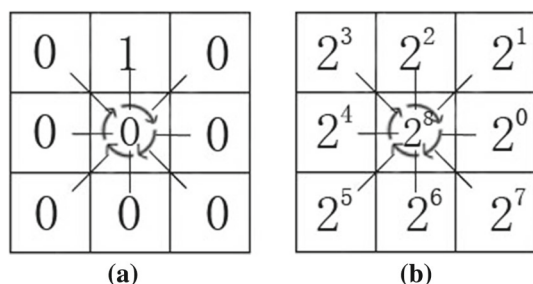
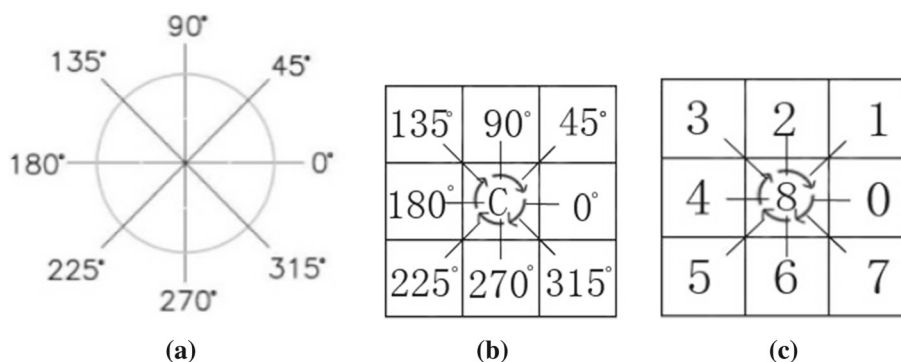


Fig. 9 **a** Binary edge matrix, MA , **b** binary pattern filter, MB

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

Figure 12 summarizes 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 is 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 two dimensional, which is difficult to do the analysis. Therefore, we extract the curve and make it in one dimension. In most cases, leaf is constructed by an 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 between the leaf and its venation easily. Image fusion is used to determine whether the curve is leaf veins or unwanted curves, and then eliminate the unwanted curves.

The proposed method outperformed the other previous works in terms of the number of the dataset used and the accuracy as more external leaf structure features is 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, ten-fold cross-validation is applied to reduce the over-fitting and improve the accuracy of the artificial neural network used for the leaf venation recognition. Since there are 1907 sets of data in the Flavia dataset, each fold would have 190 sets of data with the last one contains 197 data. After the training of the algorithm, the average accuracy obtained is at 98.60%. The average accuracy equation is as shown below,

$$\text{Average accuracy} = \left[\sum_{k=1}^{10} \left(\frac{\text{Total number of correctly recognized leaf}}{\text{Total number of leaf data}} \right) \right] / 10 \tag{22}$$

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, there are no previous works found on them. However, previous methods are applied to 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 average accuracy achieved is at 89.83%.

Previous works may give good results to the Flavia dataset which are form from higher taxa's plant; however, when it

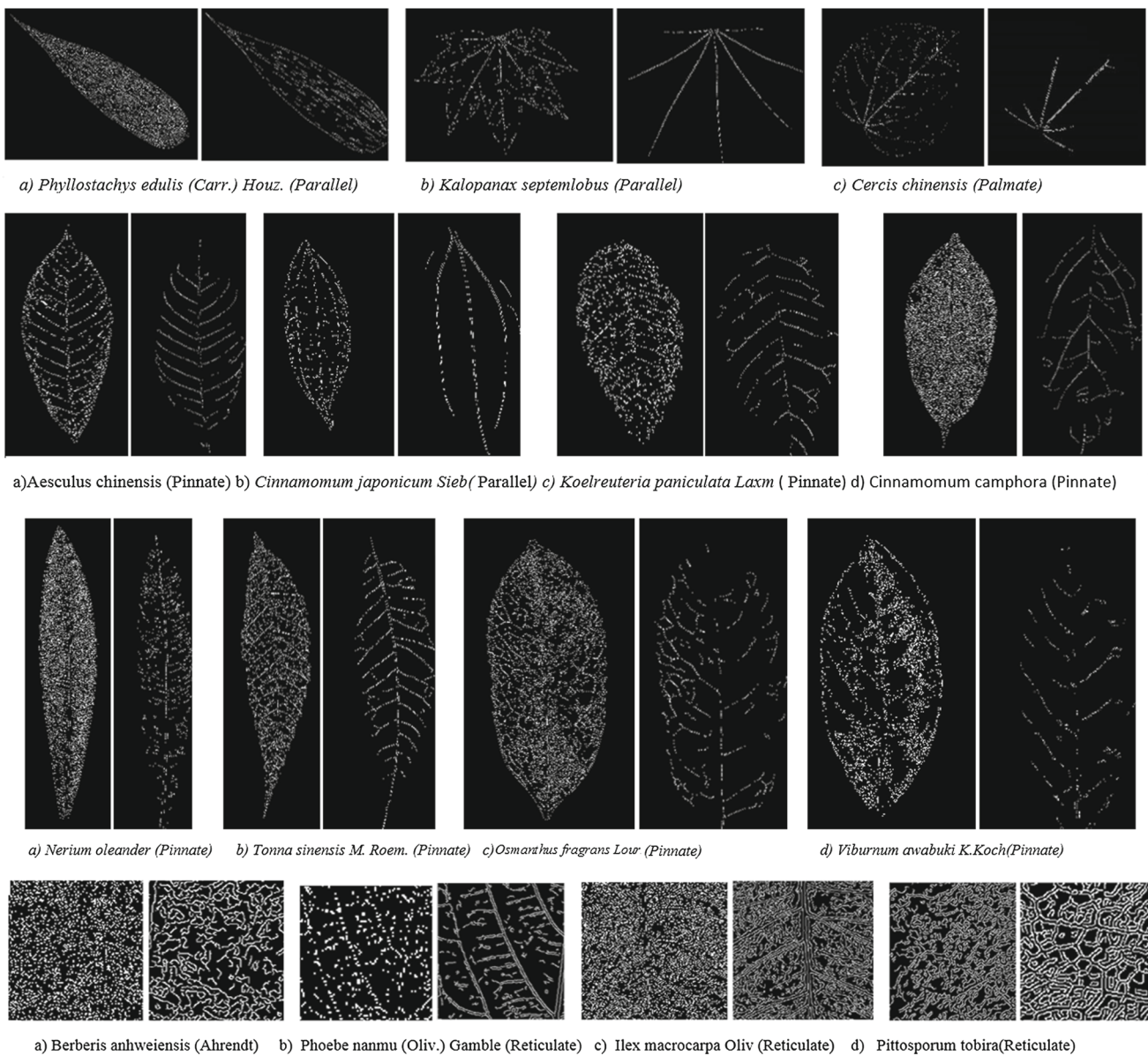


Fig. 10 Leaf venation detection in Flavia dataset



Fig. 11 Leaf venation detection in Acer dataset

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 is unable

to distinguish them. By embedded the botanical features, the achievement in identifying the plant species is promising.

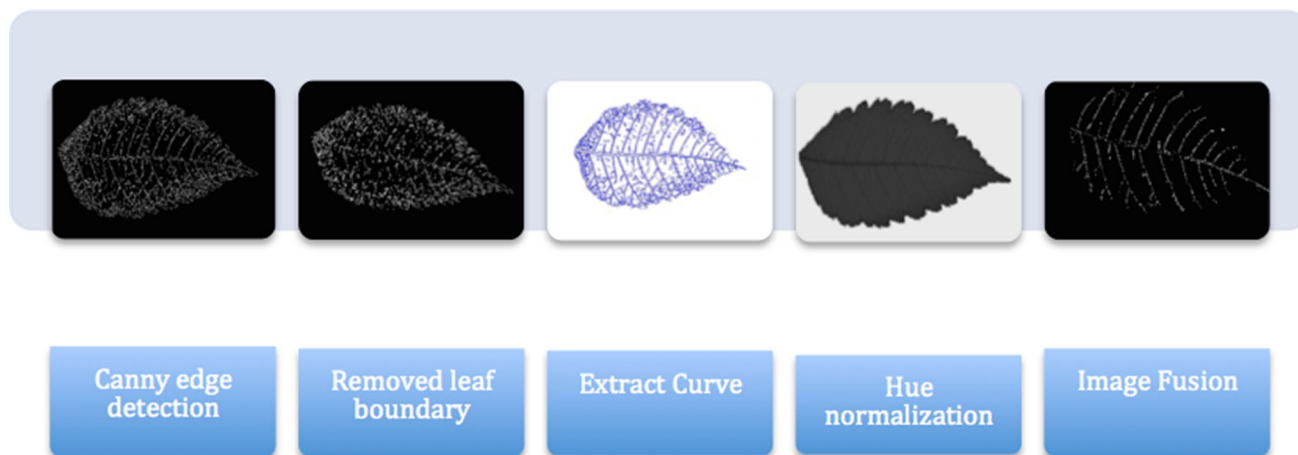


Fig. 12 Output at each and every step

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

Author	Features	Accuracy (%)	Dataset	Training	Testing	Species
Satti et al. [29]	Shape, colour	93.3	1907	1742	165	33
Chaki et al. [30]	Shape, texture	87.1	930	620	310	31
Arun et al. [31]	Texture	94.7	250	175	75	5
Wu et al. [32]	Shape, veins	90.0	1800	1800	320	32
Kadir et al. [12]	Shape, colour, vein, texture	93.4	1600	1280	320	32
Our proposed method	Venation	98.6	1907	1717	190	32

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

Author	Features	Accuracy (%)	Dataset	Training	Testing	Species
Wu et al. [32]	Shape, veins	37.3	600	450	150	32
Kadir et al. [12]	Shape, colour, vein, texture	62.0	600	450	150	32
Arun et al. [31]	Texture	43.3	600	450	150	32
Proposed method	Venation	89.83	600	540	60	32

Conclusion

To preserve 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 the presented method and existing ones shows the effectiveness of our method. We believe that this idea will seriously affect plant conservation and consequently preserve the ecosystems.

Acknowledgements This work is output of the collaboration of Department of Computer Science, Liverpool John Moores University, Liverpool, UK, University Industry Research Laboratory (UIRL), Universiti Teknologi Malaysia UTM, Skudai, Johor, Malaysia, and Machine Learning Research Group [RG-CCIS-2017-06-02] Prince Sultan University Riyadh Saudi Arabia.

References

- Goëau, H.; Bonnet, P.; Joly, A.; Baki'c, V.; Barbe, J.; Yahiaoui, I.; Selmi, S.; Carré, J.; Barthélémy, D.; Boujemaa, N.; et al.: PI@ntnet mobile app. In: Proceedings of the 21st ACM International Conference on Multimedia, pp. 423–424 (2013)
- Mehmood, Z.; Abbas, F.; Mahmood, T.; et al.: Content-based image retrieval based on visual words fusion versus features fusion of local and global features. Arab. J. Sci. Eng. (2018). <https://doi.org/10.1007/s13369-018-3062-0>

3. Jabeen, S.; Mehmood, Z.; Mahmood, T.; Saba, T.; Rehman, A.; Mahmood, M.T.: An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. *PloS one* **13**(4), e0194526 (2018)
4. Fern, B.M.; Rahim, M.S.M.; Saba, T.; Almazyad, A.S.; Rehman, A.: Stratified classification of plant species based on venation state. *Biomed. Res.* **28**(13), 5660–5663 (2017)
5. Bong, M.F.; Sulong, G.; Kumoi, R.; Rahim, M.S.M.: Classify the plant species based on lobes, sinuses and margin. *Jurnal Teknologi.* **75**(2), 27–31 (2015)
6. Kolivand, H.; Fern, B.M.; Rahim, M.S.M.; Sulong, G.; Baker, T.; Tully, D.: An expert botanical feature extraction technique based on phenetic features for identifying plant species. *PloS one* **13**(2), e0191447 (2018)
7. Neamah, K.; Mohamad, D.; Saba, T.; Rehman, A.: Discriminative features mining for offline handwritten signature verification. *3D Res.* (2014). <https://doi.org/10.1007/s13319-013-0002-3>
8. Mzoughi, O.; Yahiaoui, I.; Boujemaa, N.; Zagrouba, E.: Advanced tree species identification using multiple leaf parts image queries. In: 20th IEEE International Conference on Image Processing (ICIP), 2013, pp. 3967–3971 (2013)
9. Plant Taxonomy 2E: McGraw-Hill Education (India) Pvt Limited. https://books.google.com.my/books?id=_px_WAwHiZIC (2009)
10. Amlekar, M.; Manza, R.R.; Yannawar, P.; Gaikwad, A.T.: Leaf features based plant classification using artificial neural network. *IBMRD's J. Manag. Res.* **3**(1), 224–232 (2014)
11. Casanova, D.; Backes, A.R.; Bruno, O.M.: Pattern recognition tool based on complex network-based approach. In: *Journal of Physics: Conference Series*, p. 012048 (2013)
12. Kadir, A.; Nugroho, L.E.; Susanto, A.; Santosa, P.I.: Leaf classification using shape, color, and texture features. *arXiv preprint arXiv:1401.4447* (2013b)
13. Kadir, A.; Nugroho, L.E.; Susanto, A.; Santosa, P.I.: Neural network application on foliage plant identification. *arXiv preprint arXiv:1311.5829* (2013c)
14. Kadir, A.; Nugroho, L.E.; Susanto, A.; Santosa, P.I.: Foliage plant retrieval using polar Fourier transform, color moments and vein features. *arXiv preprint arXiv:1110.1513* (2011b)
15. Mishra, M.K.; Dandamudi, P.; Nayani, S.P.; Munikoti, S.S.; Chelukunda, S.S.; et al.: Variability in stomatal features and leaf venation pattern in Indian coffee (*Coffea arabica* L.) cultivars and their functional significance. *Bot. Serbica (Serbia)* (2011)
16. Pornpanomchai, C.; Supapatranon, C.K.; Siriwisetkukul, N.: Leaf and flower recognition system (e-Botanist). *Int. J. Eng. Technol.* **3**(4), 347 (2011)
17. Rolland-Lagan, A.-G.; Amin, M.; Pakulska, M.: Quantifying leaf venation patterns: two-dimensional maps. *Plant J. Cell Mol. Biol.* **57**(1), 195–205 (2009). <https://doi.org/10.1111/j.1365-313X.2008.03678.x>
18. De Oliveira Plotze, R.; Bruno, O.M.: Automatic leaf structure biometry: computer vision techniques and their applications in plant taxonomy. *Int. J. Pattern Recognit. Artif. Intell.* **23**(02), 247–262 (2009)
19. Syahputra, H.; Harjoko, A.; Wardoyo, R.; Pulungan, R.: Plant recognition using stereo leaf image using gray-level co-occurrence matrix. *J. Comput. Sci.* **10**(4), 697 (2014)
20. Mouine, S.; Yahiaoui, I.; Verroust-Blondet, A.: Advanced shape context for plant species identification using leaf image retrieval. In: *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, p. 49 (2012)
21. Mouine, S.; Yahiaoui, I.; Verroust-Blondet, A.: Combining leaf salient points and leaf contour descriptions for plant species recognition. In: *Image Analysis and Recognition*, pp. 205–214 (2013b)
22. Jobin, A.; Nair, M.S.; Tatavarti, R.: Plant identification based on fractal refinement technique (FRT). *Procedia Technol.* **6**, 171–179 (2012)
23. Cope, J.S.; Remagnino, P.; Barman, S.; Wilkin, P.: 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 (2010b)
24. Li, Y.; Chi, Z.; Feng, D.D.: Leaf vein extraction using independent component analysis. In: *IEEE International Conference on Systems, Man and Cybernetics, 2006, SMC'06*, pp. 3890–3894 (2006)
25. Alkawaz, M.H.; Sulong, G.; Saba, T.; Almazyad, A.S.; Rehman, A.: Concise analysis of current text automation and watermarking approaches. *Secur. Commun. Netw.* **9**(18), 6365–6378 (2017)
26. WorkWithColor.com.: Cyan. <http://www.workwithcolor.com/cyan-color-hue-range-01.htm>. Accessed 27 Dec 2017
27. Joblove, G.H.; Greenberg, D.: Color spaces for computer graphics. In: *ACM Siggraph Computer Graphics*, vol. 12, no. 3, pp. 20–25. ACM (1978)
28. Ibrahim, N.B.; Selim, M.M.; Zayed, H.H.: A dynamic skin detector based on face skin tone color. In: *8th International Conference on Informatics and Systems (INFOS)*, 2012, pp. MM-1 (2012)
29. Satti, V.; Satya, A.; Sharma, S.: An automatic leaf recognition system for plant identification using machine vision technology. *Int. J. Eng. Sci. Technol.* **5**(4), 874 (2013)
30. Chaki, J.; Parekh, R.; Bhattacharya, S.: Plant leaf recognition using texture and shape features with neural classifiers. *Pattern Recogn. Lett.* **58**, 61–68 (2015)
31. Arun, C.; Emmanuel, W.S.; Durairaj, D.C.: Texture feature extraction for identification of medicinal plants and comparison of different classifiers. *Int. J. Comput. Appl.* **62**(12), 1–9 (2013)
32. Wu, S.G.; Bao, F.S.; Xu, E.Y.; Wang, Y.X.; Chang, Y.F.; Xiang, Q.L.: A leaf recognition algorithm for plant classification using probabilistic neural network. *IEEE Int. Symp. Sig. Process. Info. Technol.* 11–16 (2007)

