Segmentation and Features Extraction of Malaysian Herbs Leaves

Suhaila Abd Halim, Normi Abdul Hadi, and Nur Syafiqah Mat Lazim
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia,
suhaila@fskm.uitm.edu.my

Abstract. In modern medical practice, herbs play a very important role as the source of biotechnology. Therefore, providing a database containing information of these herbs is significant to assist the medicinal practitioners and users. Image recognition, a process that includes image segmentation, features extraction and classification, is one of the methods to develop the database for herbs. This paper focuses on image segmentation and features extraction of 125 images with 1616×1080 pixels of Malaysian Herb Leaves including Sirih, Mexican Mint, Rerama, Belalai Gajah and Senduduk. The images were first segmented using Sobel operator to get the boundary points. Then, 7 geometrical features and 7 textural features were extracted for each image including area, perimeter, major axis length, minor axis length, roundness, smoothness and flatness. The experimental results were obtained using MATLAB R2018a. The results show that Sobel can successfully segment the images and calculates the features of the herb leaves.

1. Introduction
Traditional herbs have been a main source of medicine since ancient times. In fact, some people are still practicing this way to treat some diseases or to maintain their health. However, over the years, many species of herbs have been forgotten and today some of them are in jeopardy of becoming extinct as the result of not being well conserved [1]. This indicates the need for the development of databases and preservation areas where herb species can be documented, and information systematically gathered in order to provide valuable knowledge for the benefit of society.

Taking this as a part of the motivation for the study, image processing specifically image segmentation and features extraction are used to further support the development of databases and preservation areas for Malaysian medicinal herbs plant. Therefore, this study has developed its own database of five selected Malaysian herb leaves. The selected leaves are Sirih (Piper Betle), Mexican Mint (Plectranthus amboinicus), Rerama (Christia vespertilionis), Belalai Gajah (Clinacanthus nutans) and Senduduk (Melastoma malabathricum), which are well-known for the ability to cure various types of illness.

Sirih has the ability to help indigestion and treat diabetes [2]. Mexican Mint is believed to be very good in solving skin problem, ulceration, diarrhoea and fever [3]. Similarly, Senduduk, is often used to treat skin troubles and to accelerate wound healing [4]. Rerama and Belalai Gajah contain anticancerous and anti-inflammatory characteristics [5]. The leaf images must undergo the image
segmentation and features extraction processes in order to obtain the geometrical information of the leaves.

2. Herbs Leaves Segmentation and Features Extraction

Image segmentation is classically defined as a process that divides a digital image into different regions, wherein each region must be continuous, disconnected and have no empty subset, which is required to detect the region of interest of the image [6]. This process is significant to reduce the amount of data for later process and to preserve the important structure of the image [7].

Image segmentation is carried out by using edge detection method under gradient based method that detect the edges by finding the maximum and minimum in the first derivative of the image. Sobel operator is one of the gradient based edge operators that works by computing the gradient of image intensity at each point, giving the direction of the largest possible increase from light to dark and rate of change in that direction. It also emphasizes measuring the changes of pixel value in two directions, which are horizontal and vertical directions, and it uses 2 convolution masks to represents both directions [8].

Edge detection method, or specifically gradient based method, has been used widely in image segmentation. The methods that are commonly used for edge detection include Robert, Prewit, Sobel, Canny and wavelet. Canny operator undeniably delivers superior performance among all edge detection methods because it is robust and accurate [9,10]. However, when analysed closely, Canny has a very complex computation which makes it time consuming. On the other hand, Sobel has the advantage of being computationally inexpensive, and at the same has the ability to produce fairly good results especially when being compared to other gradient based method [7,11]. Sobel also has been used as a base for a new set of filters that aimed to lessen the presence of false edges [12].

Amlekar et al. [13] evaluated the performance of edge detection algorithm of Sobel and Canny for leaf shape analysis with certain threshold value. Canny performed better by changing parameters of threshold while Sobel produced accurate results with approximated edges for leaf shape.

Sobel has been proven a reliable edge detector for leaves such as rubber plant [14], images from leaf database such as LeafSnap [15], Folio [16] and ICL [13], and cucumber leaves [17]. Wang et al. [17] improved Sobel operator by introducing mask of 5×5 pixels of eight directions. Segmentation process without over, under or missing segmented area is important for an accurate feature detection.

Features are defined as function of one or more measurements that specify the property or characteristics of an object. A wide selection of features can be extracted from leaves, which can be comprised into two broad categories: geometric and visual [18]. However, the selection of feature extraction method must be done carefully as it can affect the final outcome of the extraction as well as the further analysis such as classification process [19].

Various geometrical features of leaves have been studied. Basic used features as suggested by Ab Jabal et al. [18] include area, perimeter, length, width and diameter. Additional features can be considered such as centroid radial distance [16], rectangularity, circularity and sphericity [15] and eccentricity [1].

3. Methodology

The general methodology is given in Figure 1. Based on the figure, the processes involved in the methodology to process the herb leaves. The process starts with data collection and image segmentation to detect the boundary of leaf from leaf dataset. Then, some features are extracted to be the input for feature vector database. Finally, the performance of the developed architecture is evaluated. The detail processes of each step are discussed in the next section.
3.1. Leaf Dataset
A dataset comprises of 125 images of randomly selected leaves which, in this study, are Sirih, Mexican Mint, Rerama, Belalai Gajah and Senduduk. The data are collected from a Kampung Sungai Balak in Selangor and captured via Sony Alpha A6000 mirrorless digital camera with 24MP. Each image consists of 1616×1080 coloured pixels.

3.2. Image Segmentation
The process of segmentation by Sobel operator starts by forming horizontal mask, \( G_x \) and vertical mask, \( G_y \) equation to form two convolution masks as in (1).

\[
\begin{align*}
G_x &= G(i+1,j) - G(i,j) \\
G_y &= G(i,j+1) - G(i,j)
\end{align*}
\]

(1)

where \((i, j)\) is the intensity value at point \((x,y)\) of the leaf. Then, the magnitude and direction of gradient are calculated using (2).

\[
\begin{align*}
\text{Magnitude, } M(x,y) &= \sqrt{G_x^2 + G_y^2} \\
\text{Direction, } \theta(x,y) &= \tan^{-1}\left(\frac{G_x}{G_y}\right)
\end{align*}
\]

(2)

The convolution masks of 3×3 pixel of Sobel edge detector are as shown in Figure 2.

\[
\begin{array}{c|c|c}
-1 & -2 & -1 \\
0 & 0 & 0 \\
1 & 2 & 1 \\
\end{array} \quad \begin{array}{c|c|c}
-1 & 0 & 1 \\
-2 & 0 & 2 \\
-1 & 0 & 1 \\
\end{array}
\]

\( G_x \) \quad \( G_y \)

Figure 2. The convolution masks of Sobel edge detector

Figure 1. Methodology of Herb Leaves Image Process
The pixel coordinates of the detected boundary are then arranged using the 8 neighboring pixels as explained in [20] before the features extraction step.

### 3.3. Features Extraction

Two types of features considered in this study are geometrical and texture features. For geometrical features, area, perimeter, major axis length, minor axis length, roundness, slimness, and smoothness [1,21] are computed. The formula of each feature is given as in (3).

- **Area**: \[ \text{Area} = \frac{1}{2} \sum_{i=1}^{n} x_i y_{i+1} - x_{i+1} y_i \]
- **Perimeter**: \[ \text{Perimeter} = \sum_{i=1}^{n} |p_i - p_{i+1}| \]
- **Major**: \[ \text{Major} = y_{\text{max}} - y_{\text{min}} \]
- **Minor**: \[ \text{Minor} = x_{\text{max}} - x_{\text{min}} \]
- **Roundness**: \[ \text{Roundness} = \frac{4\pi \text{(Area)}}{\text{Perimeter}^2} \]
- **Slimness**: \[ \text{Slimness} = \frac{\text{Major}}{\text{Minor}} \]
- **Flatness**: \[ \text{Flatness} = \frac{1}{1 + n} \]

where \((x_i, y_i)\) is the pixel coordinate of the boundary point, \(p_i\).

The texture features are contrast, correlation, energy, entropy and homogeneity depend solely to the computation of Gray Level Co-occurrence Matrix (GLCM). For both of kurtosis and skewness, the calculations are done in separate manner.

- **Contrast**: \[ \text{Contrast} = \sum_{i,j=0}^{N-1} P_{ij}(i-j)^2 \]
- **Homogeneity**: \[ \text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i-j)^2} \]
- **Correlation**: \[ \text{Correlation} = \sum_{i,j=0}^{N-1} \frac{(i-\mu)(j-\mu)}{\sigma^2} \]
- **Kurtosis**: \[ \text{Kurtosis} = \frac{E(x-\mu)^4}{\sigma^4} \]
- **Energy**: \[ \text{Energy} = \sum_{i,j=0}^{N-1} (P_{ij})^2 \]
- **Skewness**: \[ \text{Skewness} = \frac{E(x-\mu)^3}{\sigma^3} \]
- **Entropy**: \[ \text{Entropy} = \sum_{i,j=0}^{N-1} -\ln(P_{ij})P_{ij} \]
- **Mean**: \[ \mu = \sum_{i,j=0}^{N-1} iP_{ij} \]
- **Variance**: \[ \sigma^2 = \sum_{i,j=0}^{N-1} P_{ij}(i-\mu)^2 \]

where

- \(P_{ij}\) = Element \(i,j\) of the normalized symmetrical GLCM
- \(N\) = Number of grey levels in the image
- \(\mu\) = The GLCM mean
- \(\sigma\) = The intensities variance of all references pixels in the relationship that contributed to GLCM.

### 4. Results and Discussion

This study analyses the performance of the developed method in terms of image segmentation using Sobel.
4.1. Image Segmentation
As discussed earlier, there are five types of herbs leaves that are used in this study. An example of each type with its segmented image and extracted boundary is given in Figure 3.

| Name        | Original Image | Segmented Image | Extracted Boundary |
|-------------|----------------|-----------------|--------------------|
| Sirih       | (a)            | (b)             | (c)                |
| Mexican Mint| (d)            | (e)             | (f)                |
| Rerama      | (g)            | (h)             | (i)                |
| Belalai Gajah| (j)         | (k)             | (l)                |
| Senduduk    | (m)            | (n)             | (o)                |

Figure 3. Segmented Image and Extracted Boundary of Malaysian Herb Leaves

Based on Figure 3, Sobel produced clear edge for darker green leaf, which are Sirih, Rerama and Senduduk. In contrast, in some lighter green leaves, which are Mexican Mint and Belalai Gajah, parts of the detected edge are blurred and need threshold manipulation before the boundary can be extracted. This problem can be improved if the technique to capture the image is standardized in terms of lighting, background and surrounding.

4.2. Features Extraction
Table 1 tabulates the fourteen features obtained from a sample of each category of Sirih, Mexican Mint, Rerama, Belalai Gajah and Senduduk. All the values are in pixel.
Table 1. Extracted Features

| Feature   | Sirih | Mexican Mint | Rerama | Belalai Gajah | Senduduk |
|-----------|-------|--------------|--------|---------------|----------|
| Area      | 143908| 85795        | 3468   | 3956          | 108604   |
| perimeter | 1550.46| 1367.212    | 2677.49| 3311.572      | 1497.356 |
| major     | 496.4002| 384.521150  | 647.9202| 834.2169926   | 628.9972 |
| minor     | 379.0125| 293.1575755 | 307.8841| 218.6250341   | 221.4017 |
| roundness | 0.002125| 0.002053519 | 0.00368 | 0.004376434   | 0.001337 |
| slimness  | 0.995421| 0.993979145 | 0.789738| 0.826950057   | 0.996655 |
| flatness  | 0.533854| 0.656883125 | 0.978831| 0.970352711   | 0.599003 |
| Geometrical |       |              |        |               |          |
| Contrast  | 0.034463| 0.031303283 | 0.051076| 0.082908569   | 0.033638 |
| Correlation | 0.998938| 0.998793241 | 0.99816 | 0.99781783    | 0.99932  |
| Energy    | 275.2619| 308.4713165 | 171.8734| 94.9076241    | 283.2924 |
| Entropy   | 0.752268| 0.576766622 | 0.006079| 0.004533126   | 0.608703 |
| Homogeneity | 0.763522| 0.762396491 | 0.475188| 0.262072142   | 0.351992 |
| Kurtosis  | 0.999297| 0.999212598 | 0.999559| 0.999754119   | 0.999276 |
| Skewness  | 16.56085| 17.53486004 | 13.07185| 9.69059462    | 16.80156 |
| Textural  |       |              |        |               |          |

It can be observed that, the fourteen features values provide properties that can be used to identify the image of herb category uniquely.

5. Conclusions
In this research, five selected type of Malaysian herbs images, which are Sirih Mexican Mint, Rerama, Belalai Gajah and Senduduk have been chosen for features extraction. All images have been successfully segmented by using Sobel operator and the features are extracted based on the produced segmented points. There are seven geometrical features and seven Textural features are extracted from each image. The extracted features are stored in the database which can be extended in future to Malaysian herb leaves recognition process. However, the number of samples for each herb leaf should be increased in order to develop a reliable database. The results of detected edges have proved that Sobel can achieve a good segmentation result which lead to the accuracy of the extracted features. The database acts as a source for bioproduct and eco-tourism for Malaysia.

Acknowledgement
This research is supported by FRGS-RACER (600-IRMI/FRGS-RACER 5/3 (053/2019)), Ministry of Higher Education and Universiti Teknologi MARA Malaysia.

References
[1] Chaki J 2017 Template based shape feature selection for plant leaf classification Int. J. Eng. Technol.
[2] Swathi Handoo 2019 Can Betel Leaf Be Good For Your Health? STYLECRAZE
[3] Ong S, Paneerechelvan S, Lai H and Koteswara Rao N 2014 IN VITRO LIPASE INHIBITORY EFFECT OF THIRTY TWO SELECTED PLANTS IN MALAYSIA Asian J. Pharm. Clin. Res. 7
[4] Napishah H, Azmahani A, Zubaidi A L, Intan A and Nazifah A 2011 Short Communication A Preliminary Study on the Antimicrobial Properties of Several Plants Collected from
Terengganu, Malaysia J. Agrobiotech 2 99–106

[5] Azman Zakaria 2015 UPM Runs Stage Two Of Anti-cancerous Red Butterfly Wing Research | Universiti Putra Malaysia Univ. Putra Malaysia

[6] Shahidan M and Suhaila A H 2018 Image Enhancement Process on Digital Radiographic Image with Weld Discontinuities J. Mech. Eng. 5 275–92

[7] Jain A, Namdev A and Chawla M 2016 Parallel edge detection by SOBEL algorithm using CUDA C 2016 IEEE Students’ Conference on Electrical, Electronics and Computer Science, SCEECS 2016 (IEEE) pp 1–6

[8] Shrivakshan G T and Chandrasekar C 2012 A comparison of various edge detection techniques used in image processing Int. J. Comput. Sci. Issues 9 269

[9] Muthukrishnan R and Radha M 2011 Edge detection techniques for image segmentation Int. J. Comput. Sci. Inf. Technol. 3 259

[10] Zeng M, Li Y, Meng Q, Yang T and Liu J 2012 Improving histogram-based image contrast enhancement using gray-level information histogram with application to X-ray images Optik (Stuttg). 123 511–20

[11] Aggrawal H and Maini R 2009 Study and Comparison of Various Image Edge Detection Technique Int. J. Image Process. 3

[12] Chaudhary A and Gulati T 2013 Segmenting digital images using edge detection Int. J. Emerg. Technol. Adv. Eng. 3

[13] Amlekar M, Gaikwad A, Yannawar P, Manza R, Yannawae P and Manza R 2015 Comparison of edge detection methods in leaf shape analysis for plant classification Gate to Comput. Vis. Pattern Recognit. 1 19–22

[14] Suyanto and J Munte 2017 Implementation of sobel method to detect the seed rubber plant leaves Related content J. Phys. Conf. Ser. 983

[15] Kala J R and Viriri S 2018 Plant species classification using sinuosity coefficients of leaves Image Anal. Stereol. 37 119–26

[16] Munisami T, Ramsum R, Kishnah S and Pudaruth S 2015 Plant leaf recognition using shape features and colour histogram with K-nearest neighbour classifiers Procedia Comput. Sci. 58 740–7

[17] Wang Z, Wang K, Yang F, Pan S and Han Y 2017 Image segmentation of overlapping leaves based on Chan-Vese model and Sobel operator Inf. Process. Agric. 5 1–10

[18] Ab Jabal M F, Hamid S, Shuib S, Ahmad I, Faizal M, Jabal A, Hamid S, Shuib S, Ahmad I, Ab Jabal M F, Hamid S, Shuib S and Ahmad I 2013 Leaf features extraction and recognition approaches to classify plant J. Comput. Sci. 9 1295

[19] Medjahed S A 2015 A comparative study of feature extraction methods in images classification Int. J. Image, Graph. signal Process. 7 16

[20] Halim S A, Halim M S A and Hadi N A 2018 Surface reconstruction from computed tomography (CT) image of human head with the effect of noise AIP Conf. Proc. 2013

[21] Halim S A, Hadi N A, Ibrahim A and Manurung Y H 2011 The geometrical feature of weld defect in assessing digital radiographic image 2011 IEEE Int. Conf. Imaging Syst. Tech. IST 2011 - Proc. 189–93