

# LEAF FEATURES EXTRACTION AND RECOGNITION APPROACHES TO CLASSIFY PLANT

Mohamad Faizal Ab Jabal, Suhardi Hamid, Salehuddin Shuib and Illiasaak Ahmad

Department of Computer Science,  
Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Malaysia

Received 2013-06-02; Revised 2013-08-18; Accepted 2013-08-23

## ABSTRACT

Plant classification based on leaf identification is becoming a popular trend. Each leaf carries substantial information that can be used to identify and classify the origin or the type of plant. In medical perspective, images have been used by doctors to diagnose diseases and this method has been proven reliable for years. Using the same method as doctors, researchers try to simulate the same principle to recognise a plant using high quality leaf images and complex mathematical formulae for computers to decide the origin and type of plants. The experiments have yielded many success stories in the lab, but some approaches have failed miserably when tested in the real world. This happens because researchers may have ignored the facts that the real world sampling may not have the luxury and complacency as what they may have in the lab. What this study intends to deliver is the ideal case approach in plant classification and recognition that not only applicable in the real world, but also acceptable in the lab. The consequence from this study is to introducing more external factors for consideration when experimenting real world sampling for leaf recognition and classification does this.

**Keywords:** Leaf Features, Extraction, Recognition, Plant Classification

## 1. INTRODUCTION

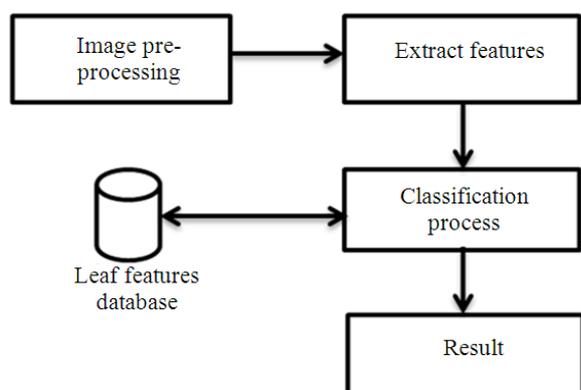
Each leaf has its own features and carries significant information that can help people to recognise and classify the plant by looking at it. Leaf shape is a prominent feature that most people use to recognise and classify a plant (Hossain and Amin, 2010). Wu *et al.* (2007) in had stated that diameter, physiological length, physiological width, leaf area and perimeter are basic geometry information can be extract from the leaf shape (Hossain and Amin, 2010). In addition, leaf colour, textures and vein are also considered as features (Kadir *et al.*, 2011a). All these features are useful for recognition and classification of leaf image. **Figure 1** illustrates the fundamental of recognition and classification process by computer using a leaf image in order to recognise and classify a plant.

Previously, most of the proposed approaches are focused on recognising and classifying method. Recognition process normally happens during pre-processing, followed by the extraction process as shown in

**Fig. 1.** After that, classification process will look up into a database to comparing the leaf features. Kadir *et al.* (2011a) had mentioned there are two categories of recognising method which are (i) contour-based and (ii) region-based approaches. However, contour-based approach has a difficulty in finding a correct curvature point compared to region-based approach (Kadir *et al.*, 2011a). Another recognition method is moment invariants, which was proposed by (Zulkifli *et al.*, 2011) where he worked with 10 kinds of leaves.

Although many approaches have been proposed and tested with almost the entire leaf features successfully extracted and recognised, still those approaches have their own limitations. It is clear that some approaches are found to be inaccurate, primarily because the input image contains noise. Besides that, different understanding on consideration to the extracted features also influence the finding because a different definition to the features or different dataset has been used for testing. The summary of all findings from the comparative study is presented in the last section of this study.

**Corresponding Author:** Mohamad Faizal Ab Jabal, Department of Computer Science, Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Malaysia



**Fig. 1.** Fundamental of plant recognition and classification

### 1.1. Leaf Features

There are 5 basic features that can be used to recognise a leaf as stated by Wu *et al.* (2007) and those features are basic geometric features that are diameter, physiological length, physiological width, leaf area and leaf perimeter. On top of that, there are 12 digital morphological features that have been defined by Wu *et al.* (2007) based on the basic geometric features (Hossain and Amin, 2010; Kadir *et al.*, 2011a; Zulkifli *et al.*, 2011).

Mentioned by Kadir *et al.* (2011b) and Wu *et al.* (2006) was divided the leaf features into 2 categories, which are the general visual features and domain-related visual features. The general visual features are consists of colour, texture and shape. It was defined as common features on images and no relation with specific type and content of images. Domain-related visual features combined with morphology characteristics of a leaf are shape, dent and vein. In addition, domain-related visual features are compulsory for extraction process (Kadir *et al.*, 2011a).

Global features and local descriptors are 2 categories, which are broken down from common features as defined by Shabanzade *et al.* (2011). The global features are properties that define a leaf shape in general, such as length, width and leaf area. Local descriptors describe leaf details such as texture, contrast, correlation and homogeneity (Shabanzade *et al.*, 2011). The shape, colour and skeleton are basic features for plant classification according to (Jing *et al.*, 2009). Ehsanirad (2010) had considered in his research that leaf shape or leaf texture or combination of both properties as extracted and recognised leaf features. In

addition, the different features are chosen to describe different properties of a leaf. Kadir *et al.* (2011a) used 2 geometric features for recognition, which are slimness and roundness. Then, he used additional leaf features in his research, which are colour, vein and texture (Kadir *et al.*, 2011b).

These facts were further confirmed by (Fotopoulou *et al.*, 2011; Valliammal and Geethalakshmi, 2012) who stated in their publications that leaf image could be categorised based on colour, texture, shape or combination of these properties. Later, Zhang and Zhang (2008) was enhanced that the properties for these features such as surface area, surface perimeter and the disfigurement are inherited from the shape features, variance of red, green and blue channels are belonging to the colour features and texture energy, texture entropy and texture contrast are fitting to the texture features. Hossain and Amin (2010) were done a research on leaf shape in order to improve the previous shape feature extraction method. Subsequently, they had defined several morphological features, which are properties from the shape features such as eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area and extent. Meanwhile, Li *et al.* (2005) and Fu and Chi (2003) have used a leaf vein features for plant recognition (Hossain and Amin, 2010; Prasad *et al.*, 2011). The essential properties have been considered from the leaf vein features are vein pixels and width of the vein. According to Najjar and Zagrouba (2012) and Lee and Chen (2003) had used region-based features for the proposed method in order to classifying the leaf (Najjar and Zagrouba, 2012). Moreover, the features have been defined consists of aspect ratio, compactness, centroid and horizontal or vertical projections. In order to classify weeds from sugar beet Jafari *et al.* (2006) was focused on leaf colour features in his research (Swain *et al.*, 2011). The purpose is weed colour is different from main plant and soil. Therefore, it is easy to differentiate between weeds and sugar beet. Intent to identify citrus disease, Pydipati *et al.* (2006) had mentioned that colour texture features is a key features (Cubero *et al.*, 2011).

On top of that, there are various definitions to the leaf features accordingly to the researchers and their research objectives. **Table 1** presents the appropriate view to the definitions that was defined by previous researchers according to certain criteria that may useful to our research.

**Table 1.** Leaf features definition

Authors	Leaf features definition	Sub-Definition
Wu <i>et al.</i> (2007); Hossain and Amin (2010); Kadir <i>et al.</i> (2011a) and Zulkifli <i>et al.</i> (2011)	5 basic features (Basic Geometric Features): <ul style="list-style-type: none"> <li>• Diameter</li> <li>• Physiological length</li> <li>• Physiological width</li> <li>• Leaf area</li> <li>• Leaf perimeter</li> </ul>	From Basic Geometric Features can another 12 Digital Morphological Features: <ul style="list-style-type: none"> <li>• Smooth factor</li> <li>• Aspect ratio</li> <li>• Form factor</li> <li>• Rectangularity</li> <li>• Narrow factor</li> <li>• Perimeter ratio of diameter</li> <li>• Perimeter ratio of physiological length</li> <li>• Perimeter ratio of physiological width</li> <li>• Vein features</li> </ul>
Fotopoulou <i>et al.</i> (2011) and Valliammal and Geethalakshmi (2012)	<ul style="list-style-type: none"> <li>• Colour</li> <li>• Texture</li> <li>• Shape or</li> <li>• Combination of these properties</li> </ul>	Shape features: <ul style="list-style-type: none"> <li>• Surface area</li> <li>• Surface perimeter</li> <li>• Disfigurement</li> </ul> Colour Features: <p>Variance of:</p> <ul style="list-style-type: none"> <li>• Red</li> <li>• Green</li> <li>• Blue</li> </ul> Texture Features: <ul style="list-style-type: none"> <li>• Texture energy</li> <li>• Texture entropy</li> <li>• Texture contrast</li> </ul>
Hossain and Amin (2010)	<ul style="list-style-type: none"> <li>• Shape</li> </ul>	Shape properties: <ul style="list-style-type: none"> <li>• Eccentricity</li> <li>• Area</li> <li>• Perimeter</li> <li>• Major axis</li> <li>• Minor axis</li> <li>• Equivalent diameter</li> <li>• Convex area</li> <li>• Extent</li> </ul>
Li <i>et al.</i> (2005); Fu and Chi (2003); Hossain and Amin (2010) and Prasad <i>et al.</i> (2011)	<ul style="list-style-type: none"> <li>• Vein</li> </ul>	Vein properties: <ul style="list-style-type: none"> <li>• Vein pixels</li> <li>• Width</li> </ul>
Shabanzade <i>et al.</i> (2011)	Global features: <ul style="list-style-type: none"> <li>• Leaf shape</li> </ul> Local descriptors (Describing the leaf details): <ul style="list-style-type: none"> <li>• Example- Texture</li> </ul>	Properties of leaf shape: <ul style="list-style-type: none"> <li>• Length</li> <li>• Width</li> <li>• Leaf area</li> </ul> Texture details: <ul style="list-style-type: none"> <li>• Contrast</li> <li>• Correlation</li> <li>• Homogeneity</li> </ul>
Jing <i>et al.</i> (2009)	Basic leaf features: <ul style="list-style-type: none"> <li>• Shape</li> <li>• Colour</li> <li>• Skeleton</li> </ul>	
Ehsanirad (2010)	<ul style="list-style-type: none"> <li>• Shape</li> <li>• Texture or</li> <li>• Combination of both properties</li> </ul>	

**Table 1.** Continue

Kadir <i>et al.</i> (2011a; 2011b)	Geometric features: <ul style="list-style-type: none"> <li>• Slimness</li> <li>• Roundness</li> </ul> Additional features: <ul style="list-style-type: none"> <li>• Colour</li> <li>• Vein</li> </ul>
Wu <i>et al.</i> (2006) and Kadir <i>et al.</i> (2011a)	General Visual Features: <ul style="list-style-type: none"> <li>• Colour</li> <li>• Texture</li> <li>• Shape</li> </ul> Domain-Related Visual Features (Combination of morphology characteristic of leaf): <ul style="list-style-type: none"> <li>• Shape</li> <li>• Dent</li> </ul>
Lee and Chen (2003) and Najjar and Zagrouba (2012)	Region-Based features: <ul style="list-style-type: none"> <li>• Aspect ratio</li> <li>• Compactness</li> <li>• Centroid</li> <li>• Horizontal projections</li> <li>• Vertical projections</li> </ul>

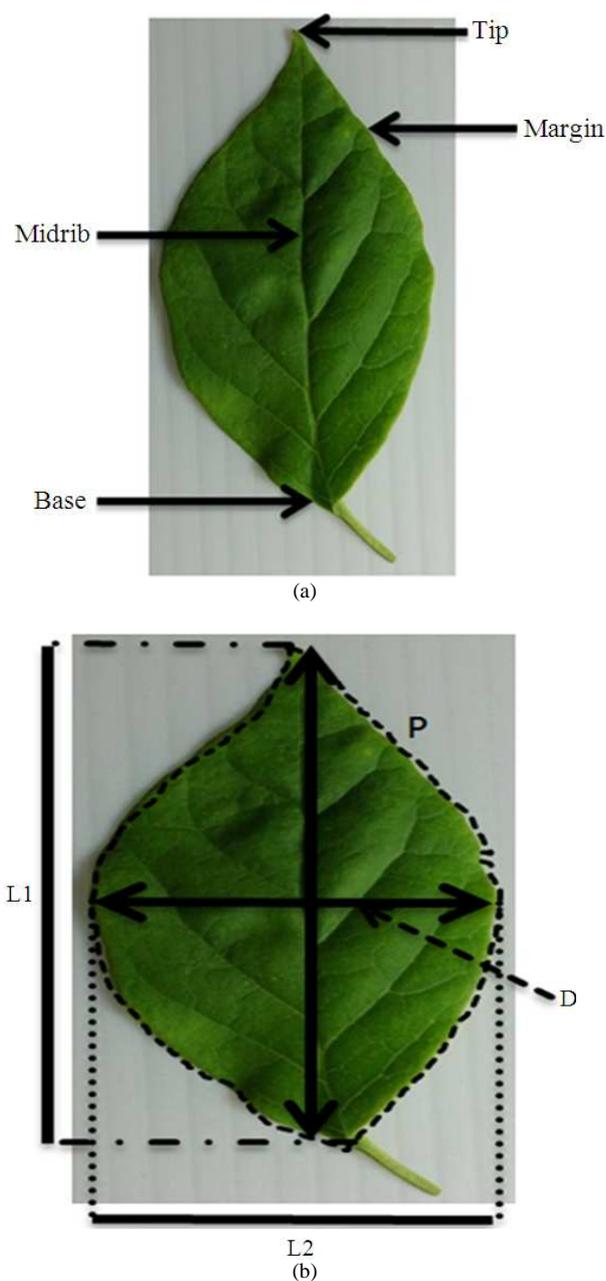
Based on the finding of the previous works, as summarised in **Table 1**, the leaf features have been redefined. There are 2 categories of features, which are geometric and visual feature. Geometric features generally defined as features that can be physically touched by humans and manually measure using common measurement tool such as a ruler. The feature is only the shape of the leaf. But, the shape of the leaf consists of several attributes or properties, which are diameter, perimeter, margin, slimness, roundness, shape of the tip and midrib. **Figure 2** depicts the leaf structure that shows the shape of the leaf, or the geometric feature. Visual feature can be measured by using a special method and a computer and is not tangible. These features are colour, vein and texture of the leaf. Colour feature is a variance of red, green and blue. The texture features are the contrast, correlation and homogeneity.

Most of the proposed approaches started the process with image pre-processing. Many techniques have been implemented during pre-processing; for instance, to convert the RGB image to gray scale, transform the gray scale image to binary image and the use image enhancement. The purpose of implementing these processes is to minimize the noise in the image that can disturb the extraction and classification process. Almost all approaches are implemented in the same manner during image pre-processing. Therefore, this study will

not elaborate much on the image pre-processing. This section will focus on the proposed leaf feature extraction and classification approach by previous researchers.

## 1.2. Features Extraction

Hossain and Amin (2010) was extracted a geometric features by using feature extraction method in which the features presented the shape of the leaf as shown in **Table 1**. Moreover they defined all the extracted features that consist of diameter as the longest distance between any two points on the margin of the leaf as shown in **Fig. 2(b)** and denoted as D; physiological length is the distance between two terminals with human interfered to mark the terminals by using mouse click as shown in **Fig. 2(b)** and denoted as L1, physiological width is the longest distance between points of those intersection pairs and denoted as L2, leaf area is a number of pixels of binary value 1 and leaf perimeter is denoted as P; in **Fig. 2(b)**. On the other hand, (Valliammal and Geethalakshmi, 2012) used image segmentation for leaf feature extraction in order to locate object shape. Instead of segmentation, (Fotopoulou *et al.*, 2011) was implemented Centroid Contour Distance (CCD) and Angle Code (CD) measurement for extracting the leaf edges. Although there are different approaches for extraction have been used, but they share a common goal, that is to extract the leaf features, as summarised in **Table 1**.



**Fig. 2.** Geometric Feature (a) Basic structure of the leaf (b) Example of leaf shape properties

Extracting vein feature is a key for modelling plant organs and living plant recognition according to Li *et al.* (2005); Fu and Chi (2003); Hossain and Amin (2010) and Prasad *et al.* (2011). On the other hand, Li *et al.* (2005) was proposed a new vein extraction method by integrating snake technique with Cellular Neural

Network (CNN) (Najjar and Zagrouba, 2012). While Fu and Chi (2003) had stated that edge operator methods such as Sobel, Prewitt and Laplacian are more suitable method for extracting vein features. Therefore by proposing two-stage approach, he has successfully extracted the vein features (Prasad *et al.*, 2011). Furthermore, the first stage process is preliminary segmentation based on the intensity histogram of the leaf image. Second stage process is fine checking using Artificial Neural Network (ANN) classifier.

Shabanzade *et al.* (2011) used statistical moments and histogram-based features method in order to extract a leaf texture feature. The reason method being used because to avoid lose some significant information regarding pixel, pixel position and information of the texture. All extracted features are classified into local descriptors category. Besides, thresholding and segmentation method has been used where the image has converted into binary image for separating the background and the object. Finally the extracted information has been classified into global features category.

Locally Linear Embedding (LLE) is to form a vector value from the image in order to extract leaf features (Jing *et al.*, 2009). The shape feature is extracted using the method that constitutes a local coordinate of the leaf edges and maps them to the global one. Meanwhile, Ehsanirad (2010) was implementing Gray-Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA) algorithms to extract leaf texture features in order to measure classification accuracy of both algorithms to further enhance the results. Even with both algorithms used for extraction, but the extraction process will influence the classification result. Hence, any mistake that occurs during extraction that will inadvertently affect the classification accuracy.

Kadir *et al.* (2011a) was recorded the experimental results of several methods with initial assumption that the methods have potentials to be used for plant extraction and classification. At the end of the experiment, he concluded that Polar Fourier Transform (PFT) method is highly potential to achieve his objective, where the objective has been set before the experiment started which was to measure the performance of accuracy for every method in order to extract and identify plants. Later, he used feature extractor in his plant identification system for extracting the leaf features as stated in **Table 1** (Kadir *et al.*, 2011b). According to Kadir *et al.* (2011a) and Wu *et al.* (2006) was grouped the features into two categories as shown in **Table 1**. Furthermore, to extract the general visual features he used moment invariants to describe the shape and shape properties features. Subsequently he used Artificial Neural

Network (ANN) for extracting vein features, which classified in domain-related visual features category. Another recent work done by Lee and Chen (2003) that was defined region-based features to be extracted for plant identification and implementing Region of Interest (ROI) to the leaf in order to extract the features by using single thresholding method (Najjar and Zagrouba, 2012).

In short, there are many methods that have been used and proposed previously in order to extract leaf features. Based on the papers reviewed, most of the feature extraction methods concentrate around extracting the leaf shape. Feature extraction is a key process that will influence final results. Any mistakes performed during extraction can generate incorrect output to the system. However, concentrating to extract leaf shape features is irrelevant if the condition of the leaf shape is cropped by the human or eaten by the insect. This limitation can cause incorrect calculation in the system and finally produce incorrect output. In other word, studies on the acceptable testing data are needed. The purpose is to identify the data either it can be processed by the system or not. In addition, the process shall be embedded in the pre-processing method. Finally, improvement must be made in the future to overcome the limitations.

### 1.3. Plant Classification

Classification process is the final phase in the plant identification system. Almost all methods that have been proposed in this phase are to retrieve the processing input in a vector value format from the extraction process. The examples of the vector values are diameter, perimeter, aspect ratio, colour variances and extent value. All values will be trained in the classification methods or algorithms in order to recognise the plants.

Wu *et al.* (2007) used Probabilistic Neural Network (PNN) to train the extracted values of 1800 leaves that was used and classified into 32 species of plant (Hossain and Amin, 2010; Kadir *et al.*, 2011a; Zulkifli *et al.*, 2011). The result is on average 90.312% accuracy. The testing to the proposed approach was conducted also with other general-purpose classification algorithms and it was found out that the algorithms only focused on leaf shape information. In other words, the proposed approach has an advantage because the approach is not only concentrating on leaf shape information in order to classify the plants. Meanwhile, Kadir *et al.* (2011b) implemented the same method in Wu *et al.* (2007) proposed approach. According him, PNN method consists of several layers and the input layer will retrieve the vector values from the extraction process for training the method. However, colour and texture features become additional input to train the

method, which was not previously applied by Wu *et al.* (2007) and Kadir *et al.* (2011b). Consequently he has shown that there is an improvement of 3.44% in accuracy of the plant classification compared to 90.312% previously done by Wu *et al.* (2007) and Kadir *et al.* (2011b).

In different research, (Jing *et al.*, 2009) proposed moving center hypersphere classifier method to identify the plant. The method consists of four processing steps which implement k-Nearest Neighbours (kNN) algorithm refers (Jing *et al.*, 2009) for process details. At the same time, other approaches also have been tested such as most close neighbour classifier, 4-close neighbour classifier and BP nerve network algorithm. All algorithms have been trained with a same dataset, which include 20 species of plants and for each species, 20 samples of leaf are collected. Furthermore, neighbours k was set from 4 to 15 for every algorithm. Finally, the training result to the proposed approach was 92.4% rate of the average for plant identification. Beside, Liao *et al.* (2010) had applied Euclidean distances to calculate the features vector to recognise the plant species. The successfulness of recognition process was 92% average of accuracy. However, the testing that has been conducted was set up with a different objective from the start, where main goal is to increase the recognition processing time. By the way she also recorded the recognition result for reference. On the other hand, Shabanzade *et al.* (2011) used Linear Discriminant Analysis (LDA) technique to discriminate between two or more categories based on a series of variables. Then, he used nearest neighbour classifier algorithm to identify the plant. 60 species of plants and 20 samples for every species have been used for training. Consequently the recognition rate is 94.3% accuracy. In addition, he mentioned the proposed features and method has an advantage because it is able to tolerate with the expecting problem that occur on the leaf.

One of the neural network methods are known as feed-forward back-propagation neural network was executed by Wu *et al.* (2007) as a recognition method in his proposed approach (Kadir *et al.*, 2011a). The number of nodes of input layer is the same as the number of extracted features and similarly with the output layer is same, as the number of plant categories, become the main reason why the method has been used for recognition purpose. Furthermore, the method that has been structured consists of three layers, which are 16 nodes of input layer, 32 nodes of hidden layer and 6 nodes of the output layer. The method has been trained based on 1200 samples which consist of 6 species of

plant and 30 leaf images from each species. The result of the training was recorded based on the species.

Still neural network algorithm has been chosen for recognition purpose. This time Xiao *et al.* (2005) was used Nearest Neighbour classifier (1-NN), k-Nearest Neighbour classifier (k-NN) and Radial Basis Probabilistic Neural Network (RBPNN) methods to train the samples (Beghin *et al.*, 2010; Prasad *et al.*, 2011; Cope *et al.*, 2012). The vector values were retrieved from the previous segmentation process, where the algorithm was proposed to integrate with Wavelet Transform (WT) and Gaussian interpolation methods. As a result, it was 93.17% for (1-NN), 85.47% for (k-NN where  $k = 5$ ) and 91.18% for (RBPNN). The finding is increasing the value of  $k$  will improve the stability of the proposed method in order to recognise the plant. Du *et al.* (2007) was proposed Move Median Centers (MMC) hypersphere as a classifier to recognise and classify the plant. 20 species of plant have been used to be sampled for testing and greater than 75% of success rate for recognition process (Beghin *et al.*, 2010; Cope *et al.*, 2012). General Regression Neural Network (GRNN) was used by Zulkifli *et al.* (2011), for recognition purposes. Similarly with other methods, the vector values from the extraction process are input into the classifier to be trained. 10 species and 10 samples from each species have been used for training. The result from the testing is 100% accuracy rate of plant recognition and classification. Not only that any changes in the spread parameters of the GRNN will not affect the process of leaf recognition.

Even most of the proposed classifiers were successful at achieving their objectives, but the situation can only be true with certain assumption created at early of the project. For instance, the data samples must be in a very good condition or in other words the image must not contain substantial noise that can affect the process. That is the purpose of this study which to search and select the optimum classifier that can be applied in our project and be useful with our data samples. **Table 2** presents the appropriate view of the proposed classifiers and the rate of success for each to recognise the plant.

Most of researchers implemented nearest neighbour or neural network as the classifiers in order to recognise the plant. However, even with the same classifier used for recognition, they still yield different rate of accuracy at the end of the testing as shown in **Table 2**. There are several situations that can be assumed as factors influencing the recognition and classification process. The factors can be categorised into two groups, which

are physical and technical factors. Detail explanation will be mentioned in a next section.

#### 1.4. Physical and Technical Factors

Recently, most of the researchers are focusing on increasing the rate of accuracy and processing time for extracting, recognising and classifying processes. However, there is less discussion on the factors that may influence those processes. As mentioned in a previous section, physical factor and technical factor are categories of factors that influence those processes. Physical factor can be defined as a factor on the data side that can be seen by using naked eyes while technical factor is a factor that occurs on the approach side. **Table 3** shows the factors in detail.

As shown in **Table 3** both factors may affect the classification rate. The reason for highlighting the factors is because usually the dataset will be tested is in a very good condition of the leaf image. In other words, the leaf image will be photographed in a plain background, good lighting condition, just a piece of the leaf in an image and a complete shape of the leaf (*situation a*). However, the real world situation of dataset which the image is photographed directly from the tree together with other leaves on branch and the sun as a source of the lighting, disregarding any condition of the shape of the leaf and without the use any plain background (*situation b*). Most of the approaches are proposed based on the *situation a*. However, in certain situations, some plants are restricted from plucking because the plants are under supervision of the responsible institution to save from extinction. Therefore, the *situation b* will occur, but the previous proposed approaches are no longer applicable. Both situations are the physical factor and both situations will affect the classification rate.

Technical factor is another category shown in **Table 3**. The factor has been highlighted because most approaches have been developed based on the *situation a* and just focusing on improvement and proving the theory or knowledge matters. However, in terms of implementation in the real world situation, the result of the testing will not be as expected. The researchers may have different objectives or focus that they want to achieve or they may have set the initial assumption such that the image must be clear of any noise before the development. That is a reason for most proposed approaches to have a limitation when they are tested in the real world condition.

**Table 2.** List of classifier and the successfulness rate of classifying and recognizing

Classifier name	Accuracy rate of plant classifying and recognizing (%)
Probabilistic Neural Network (PNN) Wu <i>et al.</i> (2007); Hossain and Amin (2010); Kadir <i>et al.</i> (2011a) and Zulkifli <i>et al.</i> (2011)	90.312%
Probabilistic Neural Network (PNN) + Color and Texture (Kadir <i>et al.</i> , 2011b)	93.752%
Moving Center Hypersphere Classifier Method+ k-Nearest Neighbours (kNN); where k = 4 to 15, (Jing <i>et al.</i> , 2009)	92.400%
Euclidean Distances (Liao <i>et al.</i> , 2010)	92.000%
Linear Discriminant Analysis (LDA) + Nearest Neighbour (1-NN) (Shabanzade <i>et al.</i> , 2011)	94.300%
Integration of Wavelet Transform (WT) & Gaussian Interpolation Methods +	
i. Nearest Neighbour (1-NN)	93.170%
ii. k -Nearest Neighbour (k-NN)	85.470%
iii. Radial Basis Probabilistic Neural Network (RBPNN)	91.180%
Xiao <i>et al.</i> (2005); Beghin <i>et al.</i> (2010); Prasad <i>et al.</i> (2011) and Cope <i>et al.</i> (2012)	
Move Median Centers (MMC) Hypersphere Classifier Du <i>et al.</i> (2007); Beghin <i>et al.</i> (2010) and Cope <i>et al.</i> (2012)	75.000%
General Regression Neural Network (GRNN) (Zulkifli <i>et al.</i> , 2011)	100.000%

**Table 3.** Categories of factor may affect the plant classification rate

	Physical Factor	Technical Factor
Definition	Occur on the data side and can be seen by using naked eyes. More focus on the test objects.	Occur on the method side and can be seen during or after the testing. More focus on the algorithms, techniques or methods use for extraction, recognition and classification process.
Examples	<ul style="list-style-type: none"> <li>i. Incomplete shape of the objects</li> <li>ii. Blurring, distortion, brightness, contrast on the images</li> <li>iii. Weather condition that may affect the images during photographing</li> </ul>	<ul style="list-style-type: none"> <li>i. Project or research objectives want to be achieved</li> <li>ii. Project or research focus</li> <li>iii. Initial assumption or scope to the dataset of the project or research</li> </ul>

In addition, the extraction process and the extracted features also affect the classification rate. The reason is a different extraction methods will extract different features. Besides, more features are extracted and considered for recognition process, hence, more accurate the classification output. In fact, the same classifier has been used for recognition and classification, but because of different extraction methods were used earlier, therefore the accuracy rates are not the same as what has been produced.

In short, this section highlights and categorises the factors that affect plant classification process that was previously ignored by researchers. Generally, both factors are needs to be considered in the future development.

## 2. CONCLUSION

The findings of this study are the types of leaf features that should be extracted, external factors that

must be considered before the extraction process, types of extraction and classification methods that can be used for plant recognition and classification. In other words, the results of this study can be used as a specification of leaf features that must be considered for plant recognition and classification purposes as shown in **Table 4**.

Finally, three classifiers have been selected for testing and future development. The selection will be based on type of leaf features that can be extracted and recognised and ability of the pre-processing method to handle the noise or other external factors in the image. The selected classifiers are Probabilistic Neural Network (PNN) + Colour and Texture as proposed by (Kadir *et al.*, 2011b), Linear Discriminant Analysis (LDA) + Nearest Neighbour (1-NN) proposed by (Shabanzade *et al.*, 2011) and General Regression Neural Network (GRNN) proposed by (Zulkifli *et al.*, 2011).

**Table 4.** Specification of the leaf features

Main features	Sub-features
Shape	<ul style="list-style-type: none"> <li>• Diameter</li> <li>• Perimeter</li> <li>• Degree of the leaf tip</li> <li>• Roundness</li> <li>• Slimness</li> </ul>
Color	<ul style="list-style-type: none"> <li>• Variance of Red, Green and Blue (RGB)</li> </ul>
Texture	<ul style="list-style-type: none"> <li>• Contrast</li> <li>• Correlation</li> <li>• Homogeneity</li> </ul>

### 2.1. Future Work

Next step in the future work, three selected classifiers will be tested based on our dataset and the results will be recorded. Only the better classifier will be used in our research. However, we may have to consider images that contain many leaves in order to test the ability of the classifiers.

### 3. ACKNOWLEDGEMENT

We would like to express our gratitude and thank you to the Division of Research and Industrial Linkages and, Faculty of Computer Science and Mathematical, University Technology MARA for the funding and those who are involved directly or indirectly in this research.

### 4. REFERENCES

- Beghin, T., J.S. Cope, P. Remagnino and S. Barman, 2010. Shape and texture based plant leaf classification. *Adv. Concepts Intell. Vision Syst.*, 6475: 345-353. DOI: 10.1007/978-3-642-17691-3\_32
- Cope, J.S., D. Corney, J.Y. Clark, P. Remagnino and P. Wilkin, 2012. Plant species identification using digital morphometrics: A review. *Expert Syst. Applic.*, 39: 7562-7573. DOI: 10.1016/j.eswa.2012.01.073
- Cubero, S., N. Aleixos, E. Molto, J. Gomez-Sanchis and J. Blasco, 2011. Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. *Food Bioprocess Technol.*, 4: 487-504. DOI: 10.1007/s11947-010-0411-8
- Du, J.X., X.F. Wang and G.J. Zhang, 2007. Leaf shape based plant species recognition. *Applied Mathem. Comput.*, 185: 883-893. DOI: 10.1016/j.amc.2006.07.072
- Ehsanirad, A., 2010. Plant classification based on leaf recognition. *Int. J. Comput. Sci. Inform. Security*, 8: 78-81.
- Fotopoulou, F., N. Laskaris, G. Economou and S. Fotopoulos, 2011. Advanced leaf image retrieval via Multidimensional Embedding Sequence Similarity (MESS) method. *Patt. Anal. Appli.* DOI: 10.1007/s10044-011-0254-6
- Fu, H. and Z. Chi, 2003. A two-stage approach for leaf vein extraction. *Proceedings of the International Conference on Neural Networks and Signal Processing*, Dec.14-17, IEEE Xplore Press, Nanjing, pp: 208-211. DOI: 10.1109/ICNNSP.2003.1279248
- Hossain, J. and M.A. Amin, 2010. Leaf shape identification based plant biometrics. *Proceedings of 13th International Conference on Computer and Information Technology*, Dec. 23-25, IEEE Xplore Press, Dhaka, Bangladesh, pp: 458-463. DOI: 10.1109/ICCITECHN.2010.5723901
- Jafari, A., H.E. Jahromi, S.S. Mohtasebi and M. Omid, 2006. Color segmentation scheme for classifying weeds from sugar beet using machine vision. *Int. J. Inform. Sci. Manage.*, 4: 1-12.
- Jing, L., Z. Shanwen and L. Jiandu, 2009. A method of plant leaf recognition based on locally linear embedding and moving center hypersphere classifier. *Proceedings of the Intelligent Computing 5th International Conference on Emerging Intelligent Computing Technology and Applications*, Springer-Verlag, (ASV '09), Ulsan, South Korea, pp: 645-651. DOI: 10.1007/978-3-642-04020-7\_69
- Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa, 2011a. A comparative experiment of several shape methods in recognizing plants. *Int. J. Comput. Sci. Inform. Technol.*, 3: 256-263. DOI: 10.5121/ijcsit.2011.3318
- Kadir, A., L.E. Nugroho, A. Susanto and P.I. Santosa, 2011b. Leaf classification using shape, color and texture features. *Int. J. Comput. Trends Technol. Seventh Sense Res. Group*, 1: 225-230.
- Lee, C.L. and S.Y. Chen, 2003. Classification for leaf images. *Proceedings of the 16th IPPR Conference on Computer Vision Graphics and Image Processing*, (CVGIP '03), pp: 355-363.
- Li, Y., Q. Zhu, Y. Cao and C. Wang, 2005. A leaf vein extraction method based on snakes technique. *Proceedings of the International Conference on Neural Networks and Brain*, Oct. 13-15, IEEE Xplore Press, pp: 885-888. DOI: 10.1109/ICNNB.2005.1614763

- Liao, Y.P., H.G. Zhou and G.R. Fan, 2010. Accelerating recognition system of leaves on nios ii embedded platform. Proceedings of the International Symposium on Computer Communication Control and Automation, May 5-7, IEEE Xplore Press, Tainan, pp: 334-337. DOI: 10.1109/3CA.2010.5533816
- Najjar, A. and E. Zagrouba, 2012. Flower image segmentation based on color analysis and a supervised evaluation. Proceedings of the International Conference on Communications and Information Technology (ICCIT), Jun. 26-28, IEEE Xplore Digital Library, pp: 397-401. DOI: 10.1109/ICCITechnol.2012.6285834
- Prasad, S., K.M. Kudiri and R.C. Tripathi, 2011. Relative sub-image based features for leaf recognition using support vector machine. Proceedings of the International Conference on Communication, Computing and Security, (ICCCS' 11), ACM Press, pp: 343-346. DOI: 10.1145/1947940.1948012
- Pydipati, R., T.F. Burks and W.S. Lee, 2006. Identification of citrus disease using color texture features and discriminant analysis. Comput. Elect. Agric., 52: 49-59. DOI: 10.1016/j.compag.2006.01.004
- Shabanzade, M., M. Zahedi and S.A. Aghvami, 2011. Combination of local descriptors and global features for leaf recognition, signal and image processing. Int. J., 2: 23-31. DOI: 10.5121/sipij.2011.2303
- Swain, K.C., M. Nørremark, R.N. Jørgensen, H.S. Midtby and O. Green, 2011. Weed identification using an Automated Active Shape Matching (AASM) technique. Biosyst. Eng., 110: 450-457. DOI: 10.1016/j.biosystemseng.2011.09.011
- Valliammal, N. and S.N. Geethalakshmi, 2012. A novel approach for plant leaf image segmentation using fuzzy clustering. Int. J. Comput. Applic., Foundat. Comput. Sci., 44: 10-20. DOI: 10.5120/6322-8669
- Wu, Q.W., C. Zhou and C. Wang, 2006. Feature extraction and automatic recognition of plant leaf using artificial neural network. Adv. Ciencias Comput.
- Wu, S.G., F.S. Bao, E.Y. Xu, Y.X. Wang and Y.F. Chang *et al.*, 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. Proceedings of the IEEE 7th International Symposium on Signal Processing and Information Technology, Dec. 15-18, IEEE Xplore Press, Giza, pp: 11-16. DOI: 10.1109/ISSPIT.2007.4458016
- Xiao, G., D. Ji-Xiang and W. Xiao-Feng, 2005. Leaf recognition based on the combination of wavelet transform and gaussian interpolation. Proceedings of the International Conference on Advances in Intelligent Computing, Aug. 23-26, Springer Berlin Heidelberg, Hefei, China, pp: 253-262. DOI: 10.1007/11538059\_27
- Zhang, X. and F. Zhang, 2008. Images features extraction of tobacco leaves. Proceeding of the Congress on Image and Signal Processing, May 27-30, IEEE Xplor Press, Sanya, China, pp: 773-776. DOI: 10.1109/CISP.2008.88
- Zulkifli, Z., S. Puteh and I.A. Mohtar, 2011. Plant leaf identification using moment invariants and general regression neural network. Proceedings of the 11th International Conference on Hybrid Intelligent Systems, Dec. 5-8, Melacca, pp: 430-435. DOI: 10.1109/HIS.2011.6122144