

# IDENTIFICATION OF TROPICAL PLANTS LEAVES IMAGE BASE ON PRINCIPAL COMPONENT ANALYSIS

Indra Laksmiana<sup>1\*</sup>, Hendra<sup>1</sup>, Sri Aulia Novita<sup>1</sup>, Fithra Herdian<sup>1</sup>, Muhammad Riza Nurtam<sup>1</sup>, Perdana Putera<sup>1</sup>, Rosda Syelly<sup>2</sup>

<sup>1</sup>Mejor of Agricultural Machinery and Equipment, Payakumbuh Agricultural Polytechnic, Jalan Raya Negara Km 7, TanjungPati, West Sumatra, Indonesia 26271

<sup>2</sup>Computer Engineering Department, Payakumbuh School of Technology, Payakumbuh 26222, Indonesia

\*Corresponding author

E-mail: indra@politanipyk.ac.id

**Commented [A1]:** Change into: Department of Agricultural Engineering

**Abstract.** Difference and variation of leaves shape is usually used as primary identifier of the plant species. But some plants may have a similar leaf shape and thus require another more accurate identifier. This study applied principal component analysis (PCA) methods for identifying tropical plant species from the shape of the leaves. This method simplified the observed variables by reducing the dimensions of the information that is stored as much as 75%, so it did not eliminate important information and can save the data processing time. There were 100 images of leaves taken from several sides of the leaf in JPEG format with which the shape of leaves were look similar, like citrus (*Citrus aurantifolia*), durian (*Durio zibethinus*), guava (*Psidium guajava*), mango (*Mangifera indica*), jackfruit (*Artocarpus heterophyllus*), avocado (*Persea americana*), rambutan (*Nephelium lappaceum*), sapodilla (*Manilkara zapota*), red betel (*Piper crocatum*) and soursop (*Annona muricata*). Identification of those 10 kind plant leaves produced 97% accuracy rate. Measurement systems were designed using the K-fold Cross Validation with  $k = 10$ , the results of experiments shown omission error occurs on the leaves of guava, jackfruit and red betel while twice commission error were found on the leaves sapodilla and once on citrus leaves.

**Keywords:** leaves image, tropical plants identification, principal component analysis

## 1. Introduction

Leaves are the most important part of a plant because leaves supply the needs of plant survival, through the conversion of light into chemical energy. The shape or structure of leaves varies greatly, depending on the species and environmental conditions the plant lives in (Sudiana et al 2008). So that the leaves become the main characteristic in identifying plant species. In general, there are four ways to identify plants: (1) asking the identity of an unknown plant directly to the expert. (2) match the specimen or herbarium that has been identified. (3) match with the moon and pictures in flora books or monographs. (4) using an identification key or a determination key (IPB 2004).

**Commented [A2]:** Add more explanations and reference

**Commented [A3]:** Generally,

**Commented [A4]:** Mention the aim and why differentiating leaves pattern and shapes are important

There is so much knowledge or information on plant leaves, such as complete leaves, petioles, leaf blades, several possible incomplete leaves, parts of leaves, leaf layout, leaf shape and others (Tjitrosoepomo, 2009). This will complicate the classification process of plants manually, besides that information regarding plant descriptions is generally still in document form (Laksmiana et al. 2013) so that it requires a long time and special understanding. So we need a system with techniques to simplify data in identifying plants using leaf images. Data classification consists of 2 steps. The first is the training phase, where a classification algorithm

is created to analyze the training data and then represented in the form of a classification rule. The second process is classification, where the test data is used to estimate the accuracy of the classification rule (Han, 2006). One of the techniques used to simplify the data used as leaf image classifiers is Principal Component Analysis (PCA).

PCA can be used to reduce the dimensions of a data without significantly reducing the characteristics of the data (Smith, 2002). According to Santosa (2007), Principal Component Analysis is a reliable technique for extracting structure from a data set with quite a lot of dimensions. PCA projects images into its eigenspace by finding the eigenvectors of each image and projecting them into the eigenvectors obtained. Research conducted by Ehsanirad (2010) in his journal "Plant classification based on leaf recognition" using the PCA method can classify plants with an accuracy rate of more than 90%. Dunlop et al (2000) used PCA to identify typical compounds from various species of eucalyptus plants and using PCA obtained groupings of compounds that clearly described the specifications of each species.

This study attempts to apply Principal Component Analysis to identify plant species based on leaf images. Leaf images can be simplified for variables by shrinking (reducing) the image matrix into a new matrix, the transformed matrix which has smaller dimensions but still retains 75% of the original matrix information, and finds the average of each class (each leaf) from the transformation results matrix, then find the smallest distance between the testing leaf images and the average of each class. This is to get information compatibility between the testing image and the plant species. So that the accuracy of the system can be known and it is hoped that it can help identify plant species that are around quickly and precisely

## 2. Methods

The stages in the research process consist of taking image data or data acquisition, converting image data into vectors, sharing data (training and testing), processing Principal Component Analysis, then analyzing accuracy.

### Image Data Acquisition

The data used in this study were 100 images, which consists of 10 types of plants of which 10 comparisons were taken. Capturing images of leaves under relatively the same lighting conditions and distances and for each type of plant consists of two parts: the upper and lower parts, because most leaves have body features such as color, protrusions of leaf veins or something else.

Ten types of leaves have a relatively close general resemblance, such as lime leaves (*Citrus aurantifolia*), durian leaves (*Durio zibethinus*), guava leaves (*Psidium guajava*), mango (*Mangifera indica.*), jackfruit (*Artocarpus heterophyllus*), avocado (*Persea americana*), rambutan (*Nephelium lappaceum*), sapodilla (*Manilkara zapota*), red betel (*Piper crocatum*) and soursop (*Annona muricata*), the leaf images used can be seen in Figure 2. All leaf images have JPG extension and RGB mode with image dimensions of 100×75 pixels.

This image is stored sequentially from the first type of leaf to the tenth type of leaf with names 10.jpg, 11.jpg, 12.jpg up to 109.jpg. The first type of leaf, *Citrus aurantifolia*, is named 10.jpg, 11.jpg, 12.jpg to 19.jpg. Furthermore, the second type of leaf *Durio zibethinus* is named 20.jpg, 21.jpg, 22.jpg up to 29.jpg and so on until the tenth leaf type of *Annona muricata* is named 100.jpg, 101.jpg up to 109.jpg.

**Commented [A5]:** Revise the paragraph, use appropriate transition markers, and pay attention to the cohesiveness of ideas.

**Commented [A6]:** Should provide what technic or method that have been used to differentiate leaves. What are their advantages and disadvantages

**Commented [A7]:** Why do chose this method instead of another method

**Commented [A8]:** Make the sentences parallel

**Commented [A9]:** Begin the concluding paragraph by summarizing the research mentioned

**Commented [A10]:** Flowchart will be more helpful to describe the process

**Commented [A11]:** Move this part to the appropriate order

**Commented [A12]:** Make it plural and pay attention to the tense



Figure 2. Leaf image.

### Image reading and data sharing

The first process is to read an image into a matrix. Image data ~~that was previously~~ in RGB mode is converted to greyscale mode. The matrix of an image is stored as a line vector measuring  $1 \times 7500$ . The process is repeated for all images and the image data ~~is~~ combined to obtain a combined matrix, the matrix X and the matrix O. The matrix X is a collection of vector lines for testing images, while the matrix O is a collection of row vectors for testing images. Image division uses k-fold cross validation with a k value of 10. Thus, it is called 10-fold cross validation. In 10-fold cross validation, the data is divided into two parts: a training set of 9 images and a testing set of 1 image. Images as training data are taken for each leaf, one image will be used as a testing image. So that there are  $9 \times 10 = 90$  images for training data. Whereas as a testing image is one image taken from each type of leaf, so there are as many as  $1 \times 10 = 10$  images. The trial was carried out 10 times based on k-fold cross validation with a value of  $k = 10$ . The first experiment is to use the first image as the testing image, the second experiment is to use the second image as the testing image and so on until the tenth experiment is to use the tenth image as the testing image.

In one experiment, the process of reading the training image was repeated 9 times for 10 leaves, resulting in an X composite matrix measuring  $90 \times 7500$ . For image testing in an experiment, it is an exception or other than data from the X matrix, because there is one image from each class (leaf) which will become a training image or which is not processed when reading the X matrix, then there will be 10 images which will become row vectors. for matrix O, so matrix O is a combined matrix of testing images measuring  $10 \times 7500$ .

### The Process of Principal Component Analysis (PCA)

**Commented [A13]:** Figure 1

**Commented [A14]:** The process was not so clear to reader some technical terms are introduced without sufficient explanation, assuming prior knowledge on the part of the reader. It would be beneficial to provide a glossary or define key terms within the context of the article itself.

**Commented [A15]:** Pay attention to the use of present and past tenses

**Commented [A16]:** Pay attention to the use of present and past tenses

**Commented [A17]:** Provide research methodology scheme

The PCA process is to eliminate the correlation between independent variables by transforming the initial independent variables into new independent variables that have no correlation at all. This new variable contains the principal component (PC) values of that variable. The PCA procedure basically aims to simplify the observed variables by reducing their dimensions (Abdi, 2010). So that PCA can be used for pattern or image recognition processes and also for image reconstruction. Digital image consists of the smallest elements is usually called pixels. Pixels store information in the form of image color intensity at these coordinates. Images can be translated as matrices and pixels can be translated as elements of the matrix (Rahmad A. 2012).

In this study the detection of leaf images was carried out from several stages such as the process of reading the image, determining the covariance matrix, determining the feature vector and feature roots corresponding to the covariance matrix, selecting the largest root characteristic. So that it contributes the desired information and determines the transformation matrix using the vector characteristic that corresponds to the characteristic of root, deriving the matrix from transformation of leaves images, determines the average vector of the transformation results for each class (plant type), determines the Euclidean distance between the testing image and each class average vector, then recognizes the image based on the minimum Euclidean distance. If the minimum distance of the image is obtained from the average class vector of the same class, then the image is recognized.

After obtaining the X matrix, the covariance calculation process is carried out, so that a sigma matrix is obtained. This covariance calculation uses the equation of variance ( $y_i$ ) =  $\sigma_{y_i}^2 = \mathbf{a}_i^T \Sigma \mathbf{a}_i$ . To perform covariance calculations, the data matrix X type or unit was converted into double so that the range of values of matrix elements change from 0-255 to 0.0-1.

After obtaining the covariance matrix sigma, then determine the vector and root characteristics of the covariance matrix. The resulting output is matrices V and D which respectively represent vector and root characteristics. Besides that, it also determines the diagonal of the matrix d. The output of the diagonal matrix d is lambda that is root characteristic values of the covariance matrix sigma. Furthermore, after obtaining the values of lambda roots characteristic value and vectors characteristics V, the extraction image characteristic and feature process was carried out. Roots characteristics were large enough to represent the desired information. In this study, the information retrieved was 75% of the initial information.

In the tenth experiment which brings out the tenth image of each class/leaf, the number of feature vectors taken to obtain information up to 75% is 14 feature vectors. The feature vectors corresponding to the feature roots are arranged as column vectors in matrix A. Matrix A is the transformation matrix that will be used to reduce matrix X. Since there are 14 feature vectors characteristics taken as column vectors for matrix A, then the matrix A shrink to 7500 x 14. The percentage value of p information for the tenth experiment is equal to 75.333%.

By using the transformation matrix A, the matrix Y was obtained as result of multiplication the matrix X with the transformation matrix A. Reduction of the matrix X into matrix Y will store the information that reached 75% of the information held by the matrix X. Matrix Y size was 90 x 14, obtained from:

$$Y_{90 \times 14} = X_{90 \times 7500} \cdot A_{7500 \times 14}$$

**Commented [A18]:** The writing style of the article is also quite dry and lacks a cohesive flow. It would greatly benefit from clearer transitions between concepts and the use of subheadings to organize the content more effectively. Furthermore, the article could benefit from incorporating visuals such as before-and-after images to illustrate the impact of the discussed techniques on different types of images.

**Commented [A19]:** conducted through

**Commented [A20]:** was

**Commented [A21]:** the next step was to

Beside the reduction of testing image, the matrix X, reduction also applied to training image, the matrix O. result of transformation matrix O to matrix A is the matrix U which has size 10 x 14, obtained from:

$$U_{10 \times 14} = O_{10 \times 7500} \cdot A_{7500 \times 14}$$

The matrix U will also provide information up to 75% of the information owned by the O matrix.

After obtaining the Y-transformed matrix, the average for each class will be calculated. The first class mean (first leaf) is obtained from calculating the first to ninth row average of matrix Y. Meanwhile, the second class average (second leaf) is obtained from calculating the tenth to eighteenth row average of matrix Y, and so on until the tenth class (tenth leaf). The result of the first to tenth class averaging is in the form of a vector stored in matrix C as a row vector. So, matrix C will be 10×14 in size. To perform the calculation of the average of each class, the initial value used is 1, which indicates that the calculation starts from the first row of the matrix Y. The addition of steps in each iteration is given a value of 9 to indicate that there are 9 rows of the matrix Y to be processed. To calculate the average value in each class. This step stops until the value is 90 or the 90th row of matrix Y. So, what we do is find the average value of every 9 rows in matrix Y. Hence, each row of matrix Y is the image data for 1 leaf.

The Euclidean distance is used to measure the level of similarity of the leaf images that you want to detect, by calculating the Euclidean distance between the average vectors of each class and the testing image. The mean vector of each class is expressed as the row vector of matrix C and the testing image in an experiment is expressed as the row vector of the matrix U. The Euclidean distance of one testing image and the average vector of a class is a scalar. So that the results of calculating the Euclidean distance of one testing image with the mean vector of each class, 10 scalar values are obtained which are stored as column vectors of matrix D. So, matrix D has dimensions of 10 × 10.

Image recognition is done by finding the minimum Euclidean distance from matrix D. Because each column vector of matrix D is the Euclidean distance between the testing images for a class and the average vector for all classes. This is to find out which testing image is closer (more similar) to which class as the identification of a testing image. The calculation results for the minimum Euclidean distance from matrix D are stored in matrices M and I. The matrix M contains the minimum Euclidean distance for each class in an experiment, so the matrix M is 1×10 in size. While matrix I contains information about which class the minimum distance is in. From this matrix I can identify which testing image is detected in which class or which leaf it is identified as.

### Accuracy analysis

The analysis conducted in this study is the accuracy of the system, the commission error and omission error. The accuracy of the overall system will be expressed in the form of tables. Experiments performed 10 times so that 10 tables were obtained from experimental results. Elements of this table will be added together to obtain the overall experiment results table. Accuracy can be searched by taking the diagonal and calculate the overall results table by using the following equation:

Commented [A22]: degree

Commented [A23]: to be detected

$$\frac{\text{number of diagonals of result table}}{100} \times 100\%$$

Commented [A24]: Not a proper way of writing formula

Omission error is an error that is obtained due to an error in the detection of an image. For example, plant type A is known as plant type B. This error is obtained from the minimum distance of the testing image of plant type A which is closer to the average vector of plant type B so that the system detects the testing image of plant type A as plant type B, while the Commission error is a detection error because plant type A gets additional detection errors from plant type B or other plant types.

### 3. Result

Commented [A25]: Result and Discussion

Based on the code formed in the programming language to analyze leaf similarity with PCA, a matrix is obtained from the identification of the 1st leaf to the 10th leaf from sample training and sample detection. The results can be seen in the table below

Table 1. Result of image identification for ten experiments

Trial	Image									
	1	2	3	4	5	6	7	8	9	10
	Identified as									
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	1	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	8	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	8	4	5	6	7	8	9	10

From Table 1 it can be seen that in experiments 2, 7 and 8, in the fifth image, the jackfruit leaf (*Artocarpus heterophyllus*) used as the testing image, this leaf is recognized as the first leaf or lime leaf (*Citrus aurantifolia*). In the ninth image, it is red betel leaf (*Piper crocatum*) and the third image, it is guava leaf (*Psidium guajava*), both are known as the eighth leaf or sapodilla (*Manilkara zapota*), the image recognition system created is not always able to recognize the image correctly this is because it is influenced by the shape of the leaves and the pattern of taking the image.

Commented [A26]: Table 1 shows that

From the results of each experiment in table 1, a new table can be created containing number 0 and 1. Number 1 showed the corresponding image. Tables 2 to 11 below show the results of the ten experiments that were carried out.

Commented [A27]: \* instead of giving a bunch of tables, just provide tables and figures that are relatef to article and then gives proper explanations

Table 2. Result of first image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0

<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 3. Result of second image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	1	0	0	0	0	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 4. Result of third image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 5. Result of fourth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0

<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 6. Result of fifth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 7. Result of sixth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 8. Result of seventh image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0



<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	1	0	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 9. Result of eighth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 10. Result of ninth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	1	0	0	0	0	0	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0
<b>6</b>	0	0	0	0	0	1	0	0	0	0
<b>7</b>	0	0	0	0	0	0	1	0	0	0
<b>8</b>	0	0	0	0	0	0	0	1	0	0
<b>9</b>	0	0	0	0	0	0	0	0	1	0
<b>10</b>	0	0	0	0	0	0	0	0	0	1

Table 11. Result of tenth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
<b>1</b>	1	0	0	0	0	0	0	0	0	0
<b>2</b>	0	1	0	0	0	0	0	0	0	0
<b>3</b>	0	0	0	0	0	0	0	1	0	0
<b>4</b>	0	0	0	1	0	0	0	0	0	0
<b>5</b>	0	0	0	0	1	0	0	0	0	0

6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

The experiment was carried out ten times, so information is also needed on how many times the system detects it accurately. The sum of each element from the experimental results table can be used to show the accuracy of the image detection system. Table 12 below is the result of the sum of all trials where the main diagonal shows the number of image detections correctly in 10 attempts

Table 12. Total of all experiments

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	10	0	0	0	0	0	0	0	0	0
2	0	10	0	0	0	0	0	0	0	0
3	0	0	9	0	0	0	0	1	0	0
4	0	0	0	10	0	0	0	0	0	0
5	1	0	0	0	9	0	0	0	0	0
6	0	0	0	0	0	10	0	0	0	0
7	0	0	0	0	0	0	10	0	0	0
8	0	0	0	0	0	0	0	10	0	0
9	0	0	0	0	0	0	0	1	9	0
10	0	0	0	0	0	0	0	0	0	10

**Commented [A28]:** Do not need to provide 10 tables, take a few tables only that can represent the data and revise the paragraph based on the representative tables, correct the title of the table as well and other sentences that contain the number of each table

**Accuracy Analysis**

Accuracy computation for leaf image detection can be counted base on sum of number diagonal in the table and divided by number of data and multiplied by 100%.

$$\begin{aligned}
 \text{Accuracy system} &= \frac{\text{Sum of number in diagonal}}{\text{Number of data}} \times 100\% \\
 &= \frac{97}{100} \times 100\% \\
 &= 97\%
 \end{aligned}$$

The accuracy of the leaf image detection system that has been built has an accuracy of 97%.

**Omission dan Commission Error**

Omission error can be calculated by adding up each row of the resulting matrix, except for the elements on the main diagonal. The output results of the omission error are shown in Table 13 below

**Commented [A29]:** Simplify the order of the words and check the number of the table again

Table 13. **Table of** omission errors

Leaf	Leaf number	Omission Error
Citrus ( <i>Citrus aurantifolia</i> )	1	0
Durian ( <i>Durio zibethinus</i> )	2	0
Guava ( <i>Psidium guajava</i> )	3	1
Mango ( <i>Mangifera indica.</i> )	4	0
Jackfruit ( <i>Artocarpus heterophyllus</i> )	5	1
Avocado ( <i>Persea americana</i> )	6	0
Rambutan ( <i>Nephelium lappaceum</i> )	7	0
Sapodilla ( <i>Manilkara zapota</i> )	8	0
Red betel ( <i>Piper crocatum</i> )	9	1
Soursop ( <i>Annona muricata</i> )	10	0

Table 13 can be explained that the omission error occurred from 10 trials occurred on the 3rd, 5th and 9th leaves, or on the leaves of guava (*Psidium guajava*), jackfruit (*Artocarpus heterophyllus*) and red betel (*Piper crocatum*). As for the other leaves, it means that the system has successfully detected them correctly, such as the 1st leaf of Lime (*Citrus aurantifolia*), 2nd leaf of Durian (*Durio zibethinus*), 4th leaf of Mango (*Mangifera indica*), 6th leaf of Avocado (*Persea americana*), 7th rambutan (*Nephelium lappaceum*), 8th leaf of sapodilla (*Manilkara zapota*) and 10th leaf of soursop (*Annona muricata*).

Commission error can be calculated by adding up each column of the result matrix, except for the elements on the main diagonal. The output results from the commission error are shown in Table 14 below

Table 14. Commission error

Leaf	Leaf number	Commission Error
Citrus ( <i>Citrus aurantifolia</i> )	1	1
Durian ( <i>Durio zibethinus</i> )	2	0
Guava ( <i>Psidium guajava</i> )	3	0
Mango ( <i>Mangifera indica.</i> )	4	0
jackfruit ( <i>Artocarpus heterophyllus</i> )	5	0
Avocado ( <i>Persea americana</i> )	6	0
Rambutan ( <i>Nephelium lappaceum</i> )	7	0
Sapodilla ( <i>Manilkara zapota</i> )	8	2
Red betel ( <i>Piper crocatum</i> )	9	0
Soursop ( <i>Annona muricata</i> )	10	0

From Table 14 it can be explained that, out of 10 experiments conducted the system made commission errors 2 times on sapodilla leaves (*Manilkara zapota*) and 1 time on lime leaves (*Citrus aurantifolia*), meaning that the system detects other leaves as itself or in other words gets detection error from another leaf. While others never get detection errors from other leaves.

Commented [A30]: Describes/shows

Commented [A31]: Revise the sentence

#### 4. Conclusion

From the discussion that has been done, several conclusions can be drawn, namely the 1D-PCA technique can be used for dimension reduction. To provide 75% of the initial data, 14 feature vectors are needed. Of the 100 leaf images processed, the result was that in terms of accuracy, the 1D-PCA technique gave very good identification results during the test, successfully recognizing leaf images correctly with a percentage level of 97%. The number of training images used in the PCA method is directly proportional to the performance of the leaf recognition system, meaning that the more training images used for training, the better the recognition results.

**Commented [A32]:** Based on the experiment,

**Commented [A33]:** The result that is relatively accurate

**Commented [A34]:** No comparison with other methods? the article lacks a discussion on potential limitations or challenges that may arise during image processing, such as noise reduction or handling various image formats.

**Commented [A35]:** Add more references and cite relevant article related to the article scope

#### References

- Abdi, H., Williams, L.J. 2010. Principal Component Analysis. Wiley Interdisciplinary Reviews: Computational Statistics. 2(4): 433-459.
- Dunlop P.J., Bignell, C.M., Jackson, J.F., Hibbert, D.B, 2000. Chemometric Analysis of Gas Chromatographic Data of Oils from Eucalyptus Species, Department of Chemistry, University of Adelaide, Adelaide.
- Ehsanirad, A. 2010. Plant Classification Based on Leaf Recognition. International Journal of Computer Science and Information Security (IJCSIS), Vol. 8, No. 4 / 2010.
- Han, J., Kamber M. 2006. Data Mining: Concept and Techniques. New York: Morgan Kaufmann Publisher.
- [IPB] Institut Pertanian Bogor. 2004. Diktat Dendrologi. Laboratorium Ekologi Hutan, Departemen Silviculture, Fakultas Kehutanan. Bogor (ID): IPB
- Laksmiana I, Hardiyeni Y, Zuhud E.A.M. 2013. *Genetic Programming for Medicinal Plant Family Identification System*, J.ICT Res.7(2), 117-136. (Jurnal) Mitchell T. 1997. *Machine Learning*. Portland: McGraw-Hill
- Rahmad A. 2012. Perbandingan Metode Ekstraksi ciri FFT, PCA dan FPE dalam Pengenalan Karakter Tulisan Tangan. Bogor: Institut Pertanian Bogor
- Santosa B., 2007, *Data Mining (Teori dan Aplikasi)*, Graha Ilmu, Yogyakarta.
- Smith I., 2002, *A tutorial on Principal Components Analysis*, Publisher John Wiley & Sons Inc.
- Sudiana D dan Elfa D. 2008. Analisis Indeks Vegetasi menggunakan Data Satelit NOAA/AVHRR dan TERRA/AQUA-MODIS. Seminar on Intelligent Technology and Its Applications (ID) ISBN 978-979-8897-24-5
- Tjitrosoepomo G. 2009. *Morfologi Tumbuhan*. Yogyakarta (ID): UGP Pr.

# IDENTIFICATION OF TROPICAL PLANTS LEAVES IMAGE BASE ON PRINCIPAL COMPONENT ANALYSIS

Indra Laksmiana<sup>\*1</sup>, Hendra<sup>1</sup>, Sri Aulia Novita<sup>1</sup>, Fithra Herdian<sup>1</sup>, Muhammad Riza Nurtam<sup>1</sup>,  
Perdana Putera<sup>1</sup>, Rosda Syelly<sup>2</sup>

<sup>1</sup>Department of Agricultural Engineering, Politeknik Pertanian Negeri Payakumbuh,  
Limapuluh Kota 26271, Indonesia

<sup>2</sup>Computer Engineering Department, Payakumbuh School of Technology, Payakumbuh  
26222, Indonesia

\*Corresponding author

E-mail: indra@politanipyk.ac.id

**Abstract.** *Difference and variation of leaves shape is usually used as primary identifier of the plant species. But some plants may have a similar leaf shape and thus require another more accurate identifier. This study applied principal component analysis (PCA) methods for identifying tropical plant species from the shape of the leaves. This method simplified the observed variables by reducing the dimensions of the information that is stored as much as 75%, so it did not eliminate important information and can save the data processing time. There were 100 images of leaves taken from several sides of the leaf in JPEG format with which the shape of leaves were look similar, like citrus (*Citrus aurantifolia*), durian (*Durio zibethinus*), guava (*Psidium guajava*), mango (*Mangifera indica*), jackfruit (*Artocarpus heterophyllus*), avocado (*Persea americana*), rambutan (*Nephelium lappaceum*), sapodilla (*Manilkara zapota*), red betel (*Piper crocatum*) and soursop (*Annona muricata*). Identification of those 10 kind plant leaves produced 97% accuracy rate. Measurement systems were designed using the K-fold Cross Validation with  $k = 10$ , the results of experiments shown omission error occurs on the leaves of guava, jackfruit and red betel while twice commission error were found on the leaves sapodilla and once on citrus leaves.*

**Keywords:** *leaves image, tropical plants identification, principal component analysis*

## 1. Introduction

The leaves are the important part of the plant as the leaf supply the needs of plants, through the conversion of light into chemical energy. Shape or structure of the leaf varies widely, depending on the species and environmental conditions of the plant life (Sudiana & Elfa, 2008). Leaves could be main identifier for identifying plant species because they are much more available compared to flower and fruits. There are several literatures directed to pattern identification that is used for plants identification and classification, 3D reconstruction of leaves, or species characterization where some tasks that could be tested in real applications (Dunlop et al., 2000; Siravenha & Carvalho, 2015). Generally, to identify the plant can be done in four-way: asking the identity of an unknown plant, directly to the expert, matching or herbarium specimens that have been identified,

matching with the pictures in books or flora monographs and lastly using key identification or key determination (IPB, 2004; Laksana *at al.*, 2018).

There are a lot of knowledge or information that can be derived from leaves of plants, like complete leaves, petiole, the leaf blade. However, incomplete leaves, part of the leaf, the layout leaves, leaf shape and others (Tjitrosoepomo, 2009). This will make difficulties in the process of plants classification manually, in addition to the most information concerning the description of the plants are still form of a document (Laksana *et al.*, 2013) so it takes a longer time to have a special understanding regarding the leaf or plant. So, it is needed a system to simplify the technique to extract data to identify plants using the image of leaf. Classification of data consists of two steps. The first is the training phase, which created a classification algorithm to analyse training data and represent it in a classification rule. The second process is the classification, in which the test data used to estimate the accuracy of the classification rule (Han & Kamper, 2006). One technique that is used to simplify the data that is used as a leaf image classifier is using Principal Component Analysis (PCA).

PCA can be used to reduce the dimension of the data without reducing significantly the characteristics of the data (Smith, 2002). According Santosa, (2007), Principal Component Analysis is a powerful technique for extracting the structure of a set of data with considerable dimensions. PCA projecting the image into its eigenspace by finding the eigenvector of every image and projected it into eigenspace obtained. Research of Ehsanirad, (2010) which used PCA method can classify plants with accuracy levels of over 90%. Dunlop *et al* (2000) used PCA to the identification of compounds typical of various species of eucalyptus plants and by using PCA, it obtained group of compounds that could clearly describes the specifications of each species. PCA probably the oldest but it is a useful tool for data modelling, compression, and visualization. PCA is the problem of fitting a low-dimensional affine subspace to a set of data points in a high-dimensional space (Vidal *et al.*, 2016).

Research on PCA has been done by many people like the research done by Destefanis. G *et al* (2000), Sinurat. S. (2014), Gniazdowski. Z. (2017), Nadler. B. (2008), Sanjaya. A., Widodo. D.W. (2018), Utami. E., Wulanningrum. R. (2014), etc. This study tried applying Principal Component Analysis to identify species based on images of leaves. Picture or image variable of the leaves can be simplified by shrinking (reducing) the image matrix into a new matrix as result of matrix transformation that has smaller dimensions but

average of each class (each leaf) from a transformed matrix, then look for the smallest distance between the image of the testing leaves and the average of each class. This is to obtain match information between the testing image and plant species. So that can be known the accuracy of the system and it is expected to help identify the types of plants quickly and precisely.

## 2. Methods

The data used in this study as many as 100 images, which consists of 10 types of plants each of which was taken 10 imagery. Capturing the image of leaves in the lighting conditions and relatively at the same distance. The leave consists of two parts, namely the upper and lower part, because leaves differences usually can be recognized on these parts such as color, bone spurs etc. These ten types of leaves are seem similar in general, like lime leaves (*Citrus aurantifolia*), leaves of durian (*Durio zibethinus*), guava leaves (*Psidium guajava*), mango (*Mangifera indica*), Jackfruit (*Artocarpus heterophyllus*), avocado (*Persea americana*), rambutan (*Nephelium lappaceum*), sapodilla (*Manilkara zapota*), red betel (*Piper crocatum*) and soursop (*Annona muricata*), the image of the leaves that are used can be seen in figure 2. All the leaves have a JPG image and RGB mode with image dimensions of  $100 \times 75$  pixels. These images were stored sequentially from the first to the tenth with name 10.jpg, 11.jpg, 12.jpg until 109.jpg. The first type of leaf *Citrus aurantifolia* named file 10.jpg, 11.jpg, 12.jpg until 19.jpg. Furthermore, second leave type, *Durio zibethinus*, named 20.jpg, 21.jpg, 29.jpg 22.jpg up to 29.jpg so on until the tenth leave type, *Annona muricata*, named 100.jpg, 101.jpg until 109.jpg

Stages in the research process consists of collecting data or image data acquisition, image processing, applying Principal Component Analysis, and lastly analysing accuracy. These stages are shown in Figure 1.

### Image Reading and Data sharing

The first process is to read an image in a matrix form. Previous image data in RGB mode converted into greyscale mode. The matrix of an image stored into a row vector with size  $1 \times 7500$ . The process is repeated for the entire image and the image data are joined to obtain a combined matrix that is the matrix X and O. Matrix X is a collection of vector lines for image training, while the matrix O is a collection of vector lines for testing image. The division of the image using the k-fold cross validation. The value of k was 10. The so it is called *10-fold cross validation*. In the 10-fold cross validation, data is divided into two parts, nine images as training set and one image as testing set. Images as training data ~~take nine of the leaf information will be used as the image of 75% as well as finding the~~

images to the training data were  $9 \times 10$  or 90 images. So that there are 10 images left as testing images. The tests were performed 10 times by applying k-fold cross validation with value of  $k = 10$ .

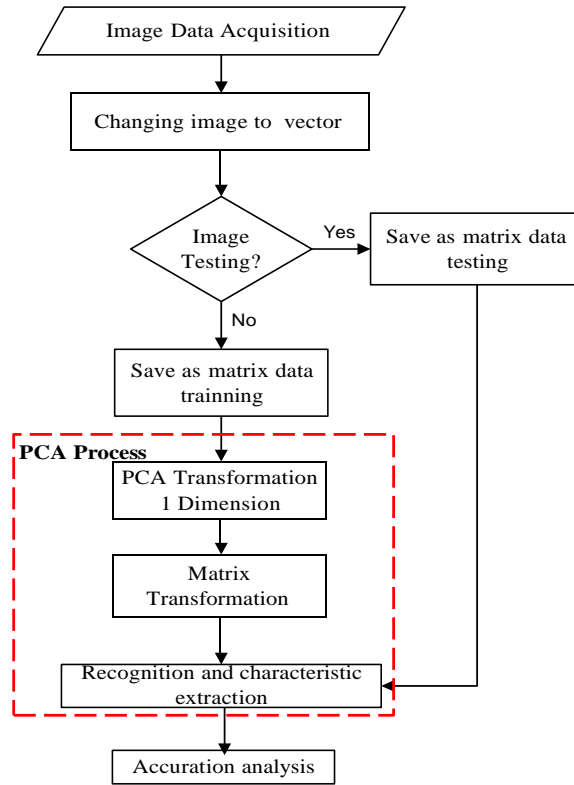


Figure. 1 Research methodology



Figure 2. Leaf images



During the experiment, the size  $90 \times 7500$  for the image of testing in a trial is added to data matrix X, because there is one image from each class (leaf) will be the image of the training or who did not O is a combined image of the testing matrix that has size  $10 \times 7500$ .

### **The process of Principal Component Analysis (PCA)**

PCA is the process by eliminating the correlation among independent variables by transforming initial independent variables into a new independent variables without any correlation. This new variable contains values of major components or principal component (PC) of the variable. PCA procedure is basically aimed to simplify the observed variables by reducing its dimensions (Abdi, 2010). So that the PCA can be used for pattern or image recognition and also for image reconstruction. Digital image consists of the smallest element that is called pixels. Pixel store information in the form of color image intensity at those coordinates. The image can be translated as matrix and pixels is the matrix elements (Rahmad, 2012).

In this study, the detection of images of the leaves was conducted through several stages. The process of reading the image, determining the covariance matrix, determining the vector characteristics and root characteristics that is corresponding to the covariance matrix, selecting the largest root characteristic that has contribution to give desired information and determine the transformation matrix by using vector characteristic that corresponding to the characteristic of root, deriving the matrix from transformation of leaves images, determine the average vector which yield of transformation for each class (plant species), determine the distance Euclidean of testing image with each average vector class, the next image recognition based on distance Euclidean minimum, If the minimum distance vector image was obtained on average the same class then the image is recognized.

After obtaining matrix X, next process was the covariance calculation, in order to obtain the matrix sigma. This covariance calculations using the variance equations  $ragam(y_i) = \sigma_{y_i}^2 = \mathbf{a}_i^T \Sigma \mathbf{a}_i$ . To perform the calculation of covariance, the data matrix X type or unit, was converted into double so that the range of values of matrix elements change from 0-255 into 0.0-.

Having obtained the covariance matrix sigma, next step was to determine the vector and the root characteristic of the covariance matrix. The resulting output is a matrix v and d which show the vector and root characteristic, respectively and also determines the diagonal of the matrix d. The output of the diagonal matrix d is lambda that is root characteristic values of the covariance matrix sigma. Furthermore, after obtaining values of

lambda roots characteristic value and vectors characteristic  $v$ , extraction image characteristic and feature process was carried out. Root characteristics were large enough to represent the desired information. In this study, the information retrieved is 75% of the initial information. In the tenth experiment, which bring out the tenth image of the each class/ leaf. The number of feature vector is taken to obtain information up to 75% in total are 14 vector characteristic. Vectors characteristic that corresponds to the root characteristics are arranged as a column vector of the matrix A. The matrix A is a transformation matrix that will be used to reduce matrix X. Since there were 14 vector characteristic that was taken as a column vector for the matrix A, then the matrix A shrink to  $7500 \times 14$ . The percentage information value for tenth trial tenth p is equal to 75.333%.

By using the transformation matrix A, the matrix Y was obtained as result of multiplication the matrix X with the transformation matrix A. Reduction of the matrix X into matrix Y will store the information that reached 75% of the information held by the matrix X. matrix Y size was  $90 \times 14$ , obtained from:

$$Y_{90 \times 14} = X_{90 \times 7500} \cdot A_{7500 \times 14}$$

Beside the reduction of testing image, the matrix X, reduction also applied to training image, the matrix O. Result of transformation matrix O to matrix A is the matrix U which has size  $10 \times 14$ , obtained from:

$$U_{10 \times 14} = O_{10 \times 7500} \cdot A_{7500 \times 14}$$

The matrix U also will provide information to reach 75% of the information held by the matrix O.

Having obtained the results of the transformation matrix Y, then the average for each class will be calculated. The average of the first class (first leaf) obtained from calculating the average of the first until the ninth row of the matrix Y. The second class flats (second leaf) obtained from calculating the average of the tenth to the eighteenth row of the matrix Y, and so on until the tenth class (tenth leaf). The results of the average of first until the tenth class was in the form of vectors that was stored in the matrix C as a vector line. So that the matrix C size will be  $10 \times 14$ . To perform the calculation of the average of each of these classes use for initial value is 1, which shows the calculation starting from the first row of the matrix Y. The addition of step at each iteration was 9 to show that there are 9 rows matrix Y will be processed to be calculated the average value in each class. This move stops up to the value 90 or row 90 of the matrix Y. Here, each row of the matrix Y is the image data for one leaf.

Euclidean distance is used to measure the degree of similarity leaf whose image was detected, by calculating the Euclidean distance between the average vectors of each class with the testing image. Average vector of each class is declared as a vector row of the matrix C and the testing image in a trial declared a row vector of the matrix U. Distance Euclidean one testing image with an average vector of the class is in the form of a scalar. So that the Euclidean distance calculation results of the testing with a vector image of the average of each class, obtained 10 scalar value which is stored as a column vector of the matrix D. The matrix D had dimensions  $10 \times 10$ .

Image recognition is done by looking for the most minimum Euclidean distance matrix D. Since each column vector of the matrix D is the Euclidean distance between an image of testing for a class and an average vector of all classes. This is to determine the testing image closer (more similar) to which class as an image identification testing Result minimum Euclidean distance matrix D is stored in the matrix M and I. The matrix M contains a minimum Euclidean distance for each class in a trial, so that the matrix M is  $1 \times 10$ . While the matrix I contains information about the minimum distance are in which class. From this matrix I can recognize an image of the testing were detected in where class the leaf was.

### **Accuracy Analysis**

The analysis conducted in this study is the accuracy of the system, the commission error and omission error. The accuracy of the overall system will be expressed in the form of tables. Experiments performed 10 times so that 10 tables were obtained from experimental results. Elements of this table will be added together to obtain the overall experiment results table. Accuracy can be searched by taking the diagonal and calculate the overall results table by using the following equation:

$$\frac{\text{number of diagonal of result table}}{100} \times 100\%$$

Omission error is an error obtained for fault detection of an image. For example, species A species is known as B. This error image appears from a minimum distance of testing of plant species A which is closer to the average vector species B so that the system detects the image of testing a plant species as species B, while the commission error is an error for A plant species get extra error detection of species B or other plant species.

### **3. Result and Discussion**

Based on the code that is formed in the programming language to analyze the similarity of the leaves with the PCA derived matrix from the identification off the 1<sup>st</sup> leaf

to 10<sup>th</sup> leaf from the training sample and sample detection. The results can be seen in each of Table 1.

Table 1. Result of image identification for ten experiments

Trial	Image									1
	2	3	4	5	6	7	8	9	10	
Identified as										
1	1	2	3	4	5	6	7	8	9	10
2	1	2	3	4	1	6	7	8	9	10
3	1	2	3	4	5	6	7	8	9	10
4	1	2	3	4	5	6	7	8	9	10
5	1	2	3	4	5	6	7	8	9	10
6	1	2	3	4	5	6	7	8	9	10
7	1	2	3	4	5	6	7	8	8	10
8	1	2	3	4	5	6	7	8	9	10
9	1	2	3	4	5	6	7	8	9	10
10	1	2	8	4	5	6	7	8	9	10

Table 1 shows that 2<sup>nd</sup>, 7<sup>th</sup> and 10<sup>th</sup>, that fifth image i.e. jackfruit leaf (*Artocarpus heterophyllus*) which recognized as testing image. This leave is recognized as first leaf or citrus leaf (*Citrus aurantifolia*). On the ninth image i.e. red betel leaf (*Piper crocatum*) and third image i.e Guava leaf (*Psidium guajava*), both are recognized as eighth image or sapodilla (*Manilkara zapota*), Leaf image processing system is not always recognize corresponding image accurately because of effect of leaf form and how the image pattern taken.

From the result of each experiment in table 1, new table can be extracted that is contain number 1 and 0. Number 1 showed the corresponding image. Table 2 until Table 5 as follow describe result of 10 experiments.

Table 2. Result of first image as testing image

Origin	Identified as									
	1	2	3	4	5	6	8	9	10	
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Table 3. Result of second image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Table 4. Result of seventh image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	0	0	1

Table 5. Result of tenth image as testing image

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	1	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0
4	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	0	0	0	0	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	1

Experiments were carried out ten times. For the 3<sup>rd</sup>, 4<sup>th</sup>, 5<sup>th</sup>, 6<sup>th</sup>, 8<sup>th</sup> and 9<sup>th</sup> experiments have the same result with the first experiment as shown in the Table 2 above. The summation of each element of the table experimental results, it is used to demonstrate the accuracy of the image detection system. The following Table 6 is the sum of all the

experiments where the main diagonal indicates the number of the 10 attempt of image detection correctly.

Table 6. Result of all experiments

Origin	Identified as									
	1	2	3	4	5	6	7	8	9	10
1	10	0	0	0	0	0	0	0	0	0
2	0	10	0	0	0	0	0	0	0	0
3	0	0	9	0	0	0	0	1	0	0
4	0	0	0	10	0	0	0	0	0	0
5	1	0	0	0	9	0	0	0	0	0
6	0	0	0	0	0	10	0	0	0	0
7	0	0	0	0	0	0	10	0	0	0
8	0	0	0	0	0	0	0	10	0	0
9	0	0	0	0	0	0	0	1	9	0
10	0	0	0	0	0	0	0	0	0	10

### Accuracy Analysis

Accuracy computation for leaf image detection can be counted base on sum of number diagonal in the table and divided by number of data and then multiplied by 100%.

$$Accuracy\ system = \frac{Sum\ of\ number\ in\ diagonal}{Number\ of\ data} \times 100\% = \frac{97}{100} \times 100\% = 97\%$$

### Omission and Commission error

Omission error is derived by summing of each row from result matrix except in diagonal. Result of omission error is shown in the table 7 as following

Table 7. Omission error

Leaf	Leaf number	Omission Error
Citrus ( <i>Citrus aurantifolia</i> )	1	0
Durian ( <i>Durio zibethinus</i> )	2	0
Guava( <i>Psidium guajava</i> )	3	1
Mango ( <i>Mangifera indica.</i> )	4	0
Jackfruit ( <i>Artocarpus heterophyllus</i> )	5	1
Avocado ( <i>Persea americana</i> )	6	0
Rambutan ( <i>Nephelium lappaceum</i> )	7	0
Sapodilla ( <i>Manilkara zapota</i> )	8	0
Redbetel ( <i>Piper crocatum</i> )	9	1
Soursop ( <i>Annona muricata</i> )	10	0

Table 7 describe omission error on Guava leaf (*Psidium guajava*), Jackfruit (*Artocarpus heterophyllus*) dan Red Betel (*Piper crocatum*). While for the others, system could detect the leaf correctly

Commission error can be derived by summing each column from result matrix except the main diagonal. Output result of commission error as described in Table 8.

Table 8. Commission error

Leaf	Leaf number	Comission Error
Citrus ( <i>Citrus aurantifolia</i> )	1	1
Durian ( <i>Durio zibethinus</i> )	2	0
Guava ( <i>Psidium guajava</i> )	3	0
Mango ( <i>Mangifera indica.</i> )	4	0
Jackfruit ( <i>Artocarpus heterophyllus</i> )	5	0
Avocado ( <i>Persea americana</i> )	6	0
Rambutan ( <i>Nephelium lappaceum</i> )	7	0
Sapodilla ( <i>Manilkara zapota</i> )	8	2
Red Betel ( <i>Piper crocatum</i> )	9	0
Soursop ( <i>Annona muricata</i> )	10	0

Accuracy of image detection of the system as shown in table 8, it was found twice commission error on the leaf sapodilla (*Manilkara zapota*) and once on citrus leaves (*Citrus aurantifolia*), in the other word the system got fault detection of other leaves. While others got no error detection from the leaves of others.

#### 4. Conclusion

Base on the experiment can be conclude that 1D-PCA may applied for dimension reduction. To provide information as much as 75% of the initial data required 14 vector characteristics. From the 100 leaves images, the result that is relatively accurate, 1D-PCA technique provides results for very good identification is managed correctly identify the image of leaf with a percentage rate of 97%. Number of training images used on the PCA method is directly proportional to the performance of the leaf recognition system, the more training images used for training, the better the results of its introduction.

#### References

- Abdi, H., & Williams, L.J. (2010). Principal Component Analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433-459.
- Arun, C.H., Sam Emanuel, W.R., & Christopher Durairaj, D. (2013). Texture feature extraction for identification of medicinal plants and comparison of different classifier. *International Journal of Computer Applications*, 62(12), 1-9.
- Destefanis. G., Barge, M.T., Brugiapaglia, A., Tassone. S. (2000). The use of principal component analysis (PCA) to characterize beef. *Meat Science*, 56(3), 255-259
- Dunlop, P.J., Bignell, C.M., Jackson, J.F., & Hibbert, D.B. (2000). Chemometric Analysis of Gas Chromatographic Data of Oils from Eucalyptus Species, Department of Chemistry. Adelaide: University of Adelaide.
- Ehsanirad, A. (2010). Plant classification based on leaf recognition. *International Journal of Computer Science and Information Security (IJCSIS)*, 8(4).
- Gniazdowski. Z. (2017). New Interpretation of Principal Components Analysis. *Zeszyty Naukowe WWSI*, 11(16), 43-65
- Han, J., & Kamber M. (2006). *Data Mining: Concept and Techniques*. New York: Morgan Kaufmann Publisher.

- [IPB] Institut Pertanian Bogor. (2004). *Diktat Dendrologi*. Laboratorium Ekologi Hutan, Departemen Silviculture, Fakultas Kehutanan. Bogor (ID): IPB.
- Laksmana, I., Hardiyeni, Y., & Zuhud, E.A.M. (2013). Genetic programming for medicinal plant family identification system. *J.ICT Res.* 7(2), 117-136.
- Laksmana, I., Syelly, R., Tazar, N., Putera, P. (2018). A genetic programming study on classification of cassava plant. *Journal of Agricultural Informatics.* 9(1), 49-61  
ISSN 2061-862X
- Nadler. B. (2008). Finite Sample Approximation Results For Principal Component Analysis: A Matrix Perturbation Approach. *The Annals of Statistics* 36(6), 2791–2817. DOI:10.1214/08-AOS618
- Rahmad, A. (2012). Perbandingan Metode Ekstraksi ciri FFT, PCA dan FPE dalam Pengenalan Karakter Tulisan Tangan. Bogor, Indonesia: Institut Pertanian Bogor.
- Sanjaya. A., Widodo. D.W. (2018). Identifikasi Tulisan Tangan Huruf Hijaiyah. *Jurnal Ilmiah NERO.* 4(1). 23-29
- Santosa, B. (2007). *Data Mining (Teori dan Aplikasi)*. Yogyakarta, Indonesia: Graha Ilmu.
- Sinurat. S. (2014). Analisa sistem pengenalan wajah berbentuk citradigital dengan algoritma principal components analysis. *Jurnal Informasi dan Teknologi.* 3(1)  
ISSN : 2339-210X
- Siravenha, A. & Carvalho, S.R. (2015). Exploring the Use of Leaf Shape Frequencies for Plant Classification. *Conference on Graphics, Patterns and Images.*
- Smith, I. (2002). *A tutorial on Principal Components Analysis*, Publisher John Wiley & Sons Inc.
- Sudiana, D. & Elfa, D. (2008). Analisis Indeks Vegetasi menggunakan Data Satelit NOAA/AVHRR dan TERRA/AQUA-MODIS. *Seminar on Intelligent Technology and Its Applications (ID)* ISBN 978-979-8897-24-5.
- Tjitrosoepomo, G. (2009). *Morfologi Tumbuhan*. Yogyakarta (ID): UGP Pr.
- Utami. E., Wulanningrum. R. (2014). Use of Principal Component Analysis and Euclidean Distance to Identify Signature Image. *IPTEK-KOM.* 16(1). 1-16. ISSN1410-334
- Vidal, R., Ma, Y. & Sastry, S.S. (2016). *Generalized Principal Component Analysis (GPCA)*. Verlag, New York, Springer.





Date : 3th January 2020  
Ref. No: 156/LoA/JAAST/2020

**Indra Laksana\*, Hendra, Sri Aulia Novita, Fithra Herdian, Mohamad Riza Nurtam,  
Perdana Putera, Rosda Syelly**

<sup>1</sup>Department of Agricultural Engineering, Politeknik Pertanian Negeri Payakumbuh,  
Limapuluh Kota 26271, Indonesia

<sup>2</sup>Computer Engineering Department, Payakumbuh School of Technology, Payakumbuh  
26222, Indonesia

\*Corresponding author  
Email: indra@politanipyk.ac.id

Dear colleague,

**Acceptance of the article publication In the journal**

Thank you for submitting an article entitled:

**“Identification of Tropical Plants Leaves Image Base on Principal Component  
Analysis”**

to the Journal of Applied Agricultural Science and Technology.

We are pleased to inform you that the reviewer has decided that your paper has been  
**ACCEPTED** for publication in this journal.



Dr. Edi Syafri  
Editor in chief

