# Herb Leaves Recognition using Gray Level Co-occurrence Matrix and Five Distance-based Similarity Measures

**R. Rizal Isnanto[1], Munawar Agus Riyadi[2], Muhammad Fahmi Awaj[3]**
[1,2]Department of Computer Engineering, Diponegoro University, Indonesia
[3]Department of Electrical Engineering, Diponegoro University, Indonesia

| Article Info | ABSTRACT |
|---|---|
| | Herb medicinal products derived from plants have long been considered as an alternative option for treating various diseases. In this paper, the feature extraction method used is Gray Level Co-occurrence Matrix (GLCM), while for its recognition using the metric calculations of Chebyshev, Cityblock, Minkowski, Canberra, and Euclidean distances. The method of determining the GLCM Analysis based on the texture analysis resulting from the extraction of this feature is Angular Second Moment, Contrast, Inverse Different Moment, Entropy as well as its Correlation. The recognition system used 10 leaf test images with GLCM method and Canberra distance resulted in the highest accuracy of 92.00%. While the use of 20 and 30 test data resulted in a recognition rate of 50.67% and 60.00%.<br><br> |

*Corresponding Author:*

R. Rizal Isnanto,
Departement of Computer Engineering,
Diponegoro University,
Jl. Prof. Soedarto, S.H., Tembalang, Semarang 50275, Indonesia.
Email: rizal_isnanto@yahoo.com

## 1. INTRODUCTION

Herb medical products derived from plants have been traditionally believed as an alternative option for treating various diseases, including for the treatment of chronic diseases such as cancer, heart disease, hepatitis, and also kidney and heart failure, at least in some Asian regions. The reason of using herbal leaves is the leaf can be easily found anywhere and processed in any form.

With today's technological advances, the demand for computer application also increases to provide benefits to human life. One of them is the need for applications that can recognize the herbal leaves pattern. The complexity in recognizing the leaf-based plants, due to various types and the different uses of herbal leaves, makes the herb leaves recognition difficult.

Various features related to the texture of the leaves were studied and the most appropriate features were used for leaf image-based plant classification. The developed system could be used to identify medicinal plants for particular diseases of human beings. The texture features have been extracted with using the GLCM and the PCA algorithms on the 390 images from 65 datasets and the new leaf or a defect to the test [1]. The PCA method comes out to be more efficient compared to the GLCM method by 98.46% accuracy. But the calculation time in the PCA method is time-consuming for example making the Eigenvector from considered leaves dataset almost took 2 hours. However The advantage of the method GLCM speed image recognition in just 5 seconds and weaknesses GLCM is very sensitive to any changes for images such as deforming or giving the new leaf image as a test.

In previous research, image-based retina recognition using the GLCM characterization (ASM, Contrast, IDM, and Entropy) and the Euclidean distance that produce retinal image recognition accuracy rate of 67.71%. The use of mean and standard deviation as the Euclidean distance threshold value in testing

against the external image that is not contained in the database generates a low level of recognition accuracy. The difference in rates due to the resulting accuracy of calculations using GLCM parameters greatly affected in image pixel position [2].

The experimental results showed that the proposed method features a combination of Gabor wavelet and Gray-Level Co-occurrence Matrix (GLCM) and Principal Component Analysis (PCA) algorithms gives the better image classification leaves [3]. The proposed method can be used to integrate simple and complex leaves for plant identification.

GLCM using SVM for linear and quadratic kernel functions gives classification accuracy of 96.15% and 100% with the execution time of 10.4848 seconds and 10.3837 seconds respectively. Medical practitioners burden will be reduced by this method when a large amount of data available. Meanwhile, the accuracy is 57.69% when PCA is implemented with the same kernel function [4].

GLCM Characterization combined with Snap Shot Method (SSM) of Eigenvalues using Principal Component Analysis (PCA) has been used on the face identification system. GLCM parameters used are Energy, Entropy, Contrast and Inverse Difference. Snap Shot Method is used to select 10 (ten) top image based on the classification using Euclidean Distance on the image of the extraction using GLCM. With an accuracy of 98.6%, this method is superior compared to the method using Wavelet + PCA = 94.5% or Curvelet + PCA = 96.6% [5].

Hashing method of feature extraction with GLCM was also used to identify the iris. By using the parameters of contrast, correlation, energy, entropy, and homogeneity, it is obtained that the recognition rate for GLCM angle of 0°, 90°, 45° and 135° are 70.49%, 51.43%, 76.24%, and 74.83%, respectively. The use GLCM angle (0° + 45° + 90° + 135°) have a higher level of recognition in the amount of 84.16% compared to the use of one GLCM angle and GLCM angle (45° + 135°) [6].

Digital Image Processing implementation was used in the leaf recognition preprocessing. The image that is meant here is a still image (photo) and the moving image (video) coming from a web camera. While digital here means that the image processing is done digitally using a computer [7].

The leaf recognition process begins with a leaf image data processing using segmentation process with the intensity normalization process, feature extraction with Gray-Level Co-occurrence Matrix (GLCM) and is identified using Chebyshev distance compared with other distance, i.e. City-Block, Minkowski, Euclidean, as well as Canberra Distances, so the resulting best among the five distances.

In leaf recognition system, there are 17 types of feature extraction that are done consist of 5 basic geometric features as well as 12 digital morphological features. Five basic geometry features are the diameter, length, width, outside and leaves perimeter [8], while 12 digital morphological features are smooth factor, aspect ratio, form factor, rectangularity, narrow factor, the perimeter ratio of the diameter, the perimeter ratio with leaf's length and width, and 5 kinds of vein feature [9].

Results of a new research with a new approach using a combination of Gray-Level Co-occurrence Matrix, lacunarity with Shen features and Bayesian classifier showed that the system provides a level of 97.19% accuracy when using Flavia dataset [10]. PCA method successfully recognizes 98% to classify 13 types of plants with 65 new or defect leaves as test images of the algorithm is trained by 390 leaves. GLCM method gives only 78% accuracy with the same object [11].

The recognition and identifying plants using shape, vein, color, texture features, combined with the apparent movement of Zernike. Radial basis probabilistic neural network (RBFNN) has been used as a classifier. To train RBFNN using double stage training algorithm that significantly improves the performance of the classifier. 95.12% accuracy rate for the simulation results on a dataset Flavia leaves [12].

## 2.    RESEARCH METHOD
### 2.1.  Materials

In this research, there are 30 types of herb leaf that had been trained and tested. All leaves observed are the well-known herb leaves by Indonesian people. The 30 types of herb leaf, both in binomial (scientific) as well as in local (Indonesian) names can be shown in Table 2. The table completed with images of all herb leaf observed can be seen in Attachment. While, six samples of herb leaf image with unique texture can be shown in Figure 1.

Table 1. List of 30 Types of Herb Leaf Observed

| No | Binomial (Scientific) name | Local name |
|----|----------------------------|------------|
| 1  | Pluchea indica | *Beluntas* |
| 2  | Ficus benjamina | *Beringin* |
| 3  | Anredera cordifolia | *Binahong* |
| 4  | Ricinus communis | *Jarak* |
| 5  | Hemigraphis alternate | *Sambang Darah* |
| 6  | Alpina galangal | *Lengkuas* |
| 7  | Strobilanthes crispus | *Keji Beling* |
| 8  | Polyscias scutellaria | *Mangkokan* |
| 9  | Morinda citrifolia | *Mengkudu* |
| 10 | Mussaenda pubescens | *Nusa Indah* |
| 11 | Stachytarpheta jamaicensis | *Pecut Kuda* |
| 12 | Carica papaya | *Pepaya* |
| 13 | Syzygium polyanthum | *Salam* |
| 14 | Piper ornatum | *Sirih Merah* |
| 15 | Annona muricata | *Sirsak* |
| 16 | Psidium guajava | *Jambu Merah* |
| 17 | Zamioculcas zamifolia | *Dolar* |
| 18 | Ixora javanica | *Soka* |
| 19 | Lobelia inflata | *Lobelia* |
| 20 | Piper betle | *Sirih Hijau* |
| 21 | Citrus hystrix | *Jeruk Purut* |
| 22 | Sauropus androgynus | *Katuk* |
| 23 | Clinacanthus nutans Lindau | *Dandang Gendis* |
| 24 | Curcuma longa | *Kunir* |
| 25 | Catharanthus roseus I | *Tapak Dara Merah* |
| 26 | Averrhoa bilimbi | *Belimbing Wuluh* |
| 27 | Annona squamosa | *Srikaya* |
| 28 | Catharanthus roseus II | *Tapak Dara Putih* |
| 29 | Boesenbergia rotunda | *Kunci* |
| 30 | Pandanus amaryllifolius | *Pandan* |



(a) Ricinus communis  (b) Stachytarpheta jamaicensis  (c) Carica papaya  (d) Piper ornatum  (e) Obelia inflata  (f) Catharanthus roseus II

Figure 1. Six samples of herb leaf observed

There are 3 (three) scenarios in this research to observe the herb leaf recognition rate, i.e. using 10, 20, and 30 types of leaf respectively. When we use 10 and 20 types of leave, then the types leaf used in training and testing process are listed in Table 2.

Table 2. List of 10 and 20 Types of Leaf used in Testing

| No | Binomial (Scientific) name | Local name |
|----|----------------------------|------------|
| | List of 10 types of leaf used in testing | |
| 1  | Pluchea indica | *Beluntas* |
| 2  | Ficus benjamina | *Beringin* |
| 3  | Anredera cordifolia | *Binahong* |
| 4  | Ricinus communis | *Jarak* |
| 5  | Strobilanthes crispus | *Keji Beling* |
| 6  | Alpina galangal | *Lengkuas* |
| 7  | Polyscias scutellaria | *Mangkokan* |
| 8  | Morinda citrifolia | *Mengkudu* |
| 9  | Mussaenda pubescens | *Nusa Indah* |
| 10 | Stachytarpheta jamaicensis | *Pecut Kuda* |

| List of 20 types of leaf used in testing | | |
|---|---|---|
| No | Binomial (Scientific) name | Local name |
| 11 | Pluchea indica | *Beluntas* |
| 12 | Ricinus communis | *Jarak* |
| 13 | Alpina galangal | *Lengkuas* |
| 14 | Polyscias scutellaria | *Mangkokan* |
| 15 | Morinda citrifolia | *Mengkudu* |
| 16 | Mussaenda pubescens | *Nusa Indah* |
| 17 | Stachytarpheta jamaicensis | *Pecut Kuda* |
| 18 | Carica papaya | *Pepaya* |
| 19 | Hemigraphis alternate | *Sambang Darah* |
| 20 | Piper ornatum | *Sirih Merah* |
| 21 | Annona muricata | *Sirsak* |
| 22 | Psidium guajava | *Jambu Merah* |
| 23 | Zamioculcas zamifolia | *Dolar* |
| 24 | Ixora javanica | *Soka* |
| 25 | Lobelia inflata | *Lobelia* |
| 26 | Piper betle | *Sirih Hijau* |
| 27 | Citrus hystrix | *Jeruk Purut* |
| 28 | Sauropus androgynus | *Katuk* |
| 29 | Clinacanthus nutans Lindau | *Dandang Gendis* |
| 30 | Catharanthus roseus II | *Tapak Dara Putih* |

## 2.2. System design

In this herb identification system, there are two main processes i.e. training and recognition processes. The flow diagram of the identification system process, in general, can be illustrated in Figure 2. The figure shows that identification system consists of two main processes namely training and recognition processes. The training process serves to register and store the image features of the leaves in database that will be used as a comparison to determine whether the image of leaves recognizable or not on the recognition process. While in the recognition process, the features from the training process are compared with the test images to obtain the recognition results. The methods of image preprocessing, feature extraction, as well as 5 (five) distance-based similarity measures will be explained in 3 (three) separate sections.
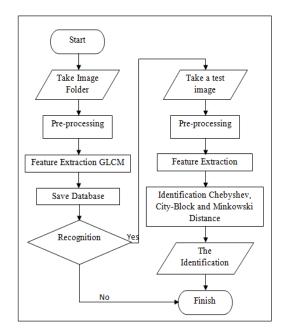


Figure 2. Flowchart of system

## 2.2. Image preprocessing

The first process in image preprocessing is conversion of color images into gray-level images to make the programming process easier because the number of bits in the gray-scale image (8-bit) is less than the number of bits in the RGB image (24-bit).

The following equation is used for converting the RGB image into a gray-scale image [13].

Gray-scale = 0,299*R + 0,587*G + 0,114*B                                            (1)

Where:

Gray-scale: value of levels of gray after conversion
R: matrix value of red component
G: matrix value of green component
B: matrix value of blue component

The second process in image preprocessing is image segmentation to separate object from the background. When the image is too bright as well as too dark, the histogram equalization is required. The histogram equalization is to obtain a histogram by equalizing the grayscale intensity values in an image. The objective is to obtain a distribution of the histogram with equalized intensity so that each degree of gray has a number of pixels that are relatively equal.

The equation for calculating histogram equalization on an image with *k*-bits gray-scale is:

$$K_0 = round\left(\frac{c_i(2^k-1)}{wh}\right)$$                                            (2)

Where:

$C_i$ : cumulative distribution of $i^{th}$ gray-scale value of the original image
$K_o$: gray-scale as a result of histogram equalization
$w$ : image width
$h$ : image height

### 2.3.  Feature extraction

Gray Level Coocurrence Matrix (GLCM) is used to extract the features of herb leaf images. The result of GLCM is some pair of pixels that each has a certain intensity value. The pixel pair pattern is spaced by distance *d*, and direction *θ*. The distance is expressed in pixels and the orientation angle is expressed in degrees. Adjacency of pixels in GLCM feature extraction method can be illustrated in a four-directions with intervals of 45°, i.e. 0°, 45°, 90°, and 135°.

Some steps in image processing with GLCM method are Gray-level pixel quantization to a matrix, to create a GLCM matrix working area, to determine the spatial relationship between the reference pixel and neighboring pixel for the value *d* and the angle *θ*, to compute the number of pairs of pixels that have the same intensity and incorporating them into the GLCM matrix working area, resulting in a co-occurrence matrix, to convert the co-occurrence matrix into symmetrical matrix by adding the matrix to its transpose, and finally, to normalize the GLCM symmetrical matrix into a form of probability numbers.

There are five GLCM parameters to calculate the second order statistical characteristics of image:
a.  ASM (Angular Second Moment)

ASM or energy is used to measure the concentration of pairs of pixels with particular gray intensity in the matrix GLCM. ASM value would be greater if the variation in the intensity of the image decreases. Function to calculate ASM is shown by the following equation:

$$ASM = \sum_{i=1}^{L}\sum_{j=1}^{L}\left(GLCM(i,j)\right)^2$$                                            (3)

b.  Contrast

Contrast is a feature that is used to measure the difference in intensity or variations of gray pixels in the image. The following equation is used to measure the contrast of an image.

$$Contrast = \sum_{i=1}^{L} n^2\left\{\sum_{|i-j|=n}^{L} GLCM(i,j)\right\}$$                                            (4)

c.  IDM (Inverse Different Moment)

IDM represents a local homogeneity in the image that has similar shades of gray in the co-occurrence matrix. IDM value will be greater when couples of pixels that have the eligible intensity of co-occurrence matrix are concentrated in a few coordinates and will shrink when spread.

$$IDM = \sum_{i=1}^{L}\sum_{j=1}^{L}\frac{(GLCM(i,j)^2)}{1+(i-j)^2} \tag{5}$$

d. Entropy

Entropy on GLCM measures the disorder distribution of gray levels of an image on a co-occurrence matrix. Entropy is high when the elements of GLCM have relatively equal values.

$$Entropy = \sum_{i=1}^{L}\sum_{j=1}^{L}\Big(GLCM(i,j)\log\big(GLCM(i,j)\big)\Big) \tag{6}$$

e. Correlation

Correlation is a measure of linear dependence between the values of gray levels in the image. Correlation function can be seen in the equation:

$$Correlation = \frac{\sum_{i=1}^{L}\sum_{j=1}^{L}(i,j)\Big(GLCM(i,j)-\mu_i'\mu_j'\Big)}{\sigma_i'\sigma_j'} \tag{7}$$

## 2.4. Similarity measures using 5 (five) distances

After the feature extraction process, similarity tests between testing image features and registered image features. There are 5 (five) distances implemented, those are Chebyshev, Cityblock, Minkowski, Canberra, and Euclidean distances.
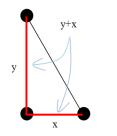
City-block distance is defined as follows:

$$d(v_1, v_2) = \sum_{K=1}^{N}|v_1(k)-v_2(k)| \tag{8}$$

Where $v_1$ and $v_2$ are two vectors whose distances will be calculated and $N$ denotes the length of the vector. If the vector has two values, city-block distance can be imagined as a horizontal plus vertical distance from the first vector to the second vector, which is illustrated in Figure 3(a).
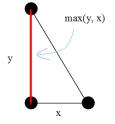
Distance box chess or also known by the name of Chebyshev distance is defined as follows.

$$d(v_1, v_2) = \max_{k=1\rightarrow N}(|v_1(k)-v_2(k)|) \tag{9}$$

The equation above illustrates that $v_1$ and $v_2$ are two vectors that distance will be calculated and N denotes the length of the vector. If the vector has two values, the distance can be envisioned as the longest distance between the horizontal distance and vertical distance, which is depicted in Figure 3(b).

(a) City-block Distance illustration                  (b) Chebyshev Distance illustration

Figure 3. Comparison between City-block and Chebyshev distances

Minkowski distance is defined as follows.

$$d(v_1, v_2) = \sqrt[p]{\sum_{k=1}^{N}|v_1(k)-v_2(k)|^p} \tag{10}$$

If the value $p = 1$, then the equation becomes the City-block distance. For the value $p = 2$, it

becomes the Euclidean distance, which can be stated as the formula:

$$d(v_1, v_2) = \sqrt{\sum_{k=1}^{N} |v_1(k) - v_2(k)|^2} \tag{11}$$

While Canberra distance can be expressed as:

$$d(v_1, v_2) = \sum_{k=1}^{n} \frac{|v_1(k) - v_2(k)|}{|v_1(k)| + |v_2(k)|} \tag{12}$$

Referring to the explanation uses leaf pattern as its characteristic, this research is to design a system that can recognize the types of herbal leaves. The leaf image recognition process begins with an image data processing using segmentation process with the intensity normalization process, feature extraction with Gray-Level Co-occurrence Matrix (GLCM) method and its identification using distance-based similarity measures. Therefore, this research aims to compare the recognition method using distance-based similarity measures, i.e. Chebyshev, City-Block, Minkowski, Canberra, and Euclidean distances.

## 3. RESULTS AND ANALYSIS
### 3.1. GLCM feature extraction
There are five characteristics as the results of GLCM feature extraction, i.e. ASM, contrast, IDM, entropy, and correlation. These five parameters can be used as distinguishing features for each leaf image of medicinal plants. The examples from GLCM characteristic value of the test images by the two types of leaves, i.e Pluchea indica and Ficus benjamin of medicinal plants at a distance of 1 pixel and an orientation angle of 0° can be seen in Table 3.

Table 3. Samples of GLCM Value of Test Image at Angle $0^0$ and Distance 1 Pixel

| Image Name | ASM | Contrast | IDM | Entropy | Correlation |
|---|---|---|---|---|---|
| Pluchea indica  (1) | 0.0012 | 129.6602 | 0.2426 | 7.9795 | 4.25e-04 |
| Pluchea indica (2) | 0.0027 | 102.3739 | 0.2929 | 7.6757 | 3.78e-04 |
| Ficus benjamin (1) | 0.0033 | 95.2215 | 0.362 | 7.6097 | 2.03e-04 |
| Ficus benjamin (2) | 0.0026 | 123.8462 | 0.337 | 7.7655 | 1.89e-04 |

### 3.2. Training image recognition
Training image recognition is conducted on the image that has previously existed in database. Testing with this training image aims to validate whether the resulting Chebyshev distance is zero or not. The Table 4 shows three samples of leaf as testing results of training image.

Table 4. Samples of Training Image Test

| Test Data Name | Chebyshev distance | Recognized as | Explanation |
|---|---|---|---|
| Ricinus communis_012 | 0 | Ricinus communis | True |
| Sauropus androgynus_064 | 0 | Sauropus androgynus | True |
| Zamioculcas zamifolia_050 | 0 | Zamioculcas zamifolia | True |

Table 4 is the result of Chebyshev distance in training image testing is zero as the result of the difference between the feature vector of the training image being tested and the training image in the database. Thus, it is known that the recognition program works properly because the train image being tested is exactly the same as the existing image in the database. If the train image testing does not produce the Chebyshev distance of zero, it means that the recognition program does not work properly.

### 3.3. Test image recognition
The test image recognition of the herb leaves is performed to find the percentage of successful recognition from 30 types of leaves of medicinal plants. Five images from each type of leaves are taken as the test images that are not going through the training stage, then tested against 5 variations of GLCM orientation angle and 5-pixel distances. To calculate the percentage of the success rate of recognition (*p*), it

can use the Equation (13). Table 5 is an example of the test results on 3 (three) variation distance-based similarity measures of angle 90° and distance of 1 pixel.

$$p = \frac{\text{Number of test images recognized correctly}}{\text{Number of all test images}} \qquad (13)$$

Table 5. Samples of Testing Result of Test Images with Angle $90^0$, $D = 1$

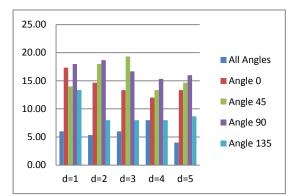| Test Data Name | Chebyshev | Cityblock | Minkowski |
|---|---|---|---|
| Pluchea indica (1) | True | False | False |
| Pluchea indica (2) | True | False | False |
| Pluchea indica (3) | True | False | False |
| Pluchea indica (4) | True | False | False |
| Pluchea indica (5) | True | False | False |
| Ficus Benjamina (1) | False | False | False |
| Ficus Benjamina (2) | False | True | True |
| Ficus Benjamina (3) | False | False | False |
| Ficus Benjamina (4) | False | False | False |
| Ficus Benjamina (5) | False | False | False |
| Anredera Cordifolia (1) | True | False | False |
| Anredera Cordifolia (2) | True | False | False |
| Anredera Cordifolia (3) | False | False | False |
| Anredera Cordifolia (4) | True | False | False |
| Anredera Cordifolia (5) | False | False | False |
| Ricinus Communis (1) | False | False | True |
| Ricinus Communis (2) | False | True | True |
| Ricinus Communis (3) | False | False | False |
| Ricinus Communis (4) | False | True | True |
| Ricinus Communis (5) | False | False | False |
| Hemigraphis Alternata (1) | False | False | False |
| Hemigraphis Alternata (2) | False | False | False |
| Hemigraphis Alternata (3) | False | False | False |
| Hemigraphis Alternata (4) | False | False | False |
| Hemigraphis Alternata (5) | False | False | False |
| Alpinia Galanga (1) | True | True | True |
| Alpinia Galanga (2) | True | True | True |
| Alpinia Galanga (3) | True | True | True |
| Alpinia Galanga (4) | True | True | True |
| Alpinia Galanga (5) | False | False | False |

Results of recognition rate throughout the testing that has been conducted on 150 test images with all variations are shown in Table 6.

Table 6. Recognition Percentage of All Variations of GLCM Parameters (Chebyshev)

| Orientation | Pixel Distance | | | | |
|---|---|---|---|---|---|
| Angle | d=1 | d=2 | d=3 | d=4 | d=5 |
| All Angles | 6.00 | 5.33 | 6.00 | 8.00 | 4.00 |
| Angle 0 | 17.33 | 14.67 | 13.33 | 12.00 | 13.33 |
| Angle 45 | 14.00 | 18.00 | 19.33 | 13.33 | 14.67 |
| Angle 90 | 18.00 | 18.67 | 16.67 | 15.33 | 16.00 |
| Angle 135 | 13.33 | 8.00 | 8.00 | 8.00 | 8.67 |

Figure 5 and Table 6 illustrate the recognition success rates graph of 150 test images from 30 types of herb leaf using Chebyshev distance measurement. After testing be done, this recognition system with GLCM method and Chebyshev distance produce the highest accuracy rate of 19.33% which is at a distance of 3 pixels and using a single orientation angle of 45°. While the lowest accuracy rate is 4.00% when using an orientation all angle within 5 neighboring pixels.

Table 7 and Figure 6 depict the recognition success rates graph of 150 test images using City-block distance measurement. After testing be done, this recognition system with GLCM method and Cityblock distance produce the highest accuracy rate of 18.67% which is at a distance of 1 pixels and a single orientation angle of 0°. While the lowest accuracy rate is 4.00% when using an orientation all angle within 5 neighboring pixels and using an orientation all angle.
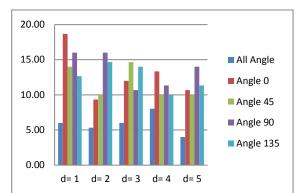
Figure 5. Success Rates Graph of Test Image
Recognition using Chebyshev distance



Figure 6. Success Rates Graph of Test Image
Recognition using City-block

Table 7. Recognition Percentage of All Variations of GLCM Parameters (City-block)

| Orientation | Pixel Distance | | | | |
|---|---|---|---|---|---|
| Angle | d=1 | d=2 | d=3 | d=4 | d=5 |
| All Angles | 6.00 | 5.33 | 6.00 | 8.00 | 4.00 |
| Angle 0 | 18.67 | 9.33 | 12.00 | 13.33 | 10.67 |
| Angle 45 | 14.00 | 10.00 | 14.67 | 10.00 | 10.00 |
| Angle 90 | 16.00 | 16.00 | 10.67 | 11.33 | 14.00 |
| Angle 135 | 12.67 | 14.67 | 14.00 | 10.00 | 11.33 |

Table 8. Recognition Percentage of All Variations of GLCM Parameters (Minkowski)

| Orientation | Pixel Distance | | | | |
|---|---|---|---|---|---|
| Angle | d=1 | d=2 | d=3 | d=4 | d=5 |
| All Angles | 6.00 | 5.33 | 6.00 | 8.00 | 4.00 |
| Angle 0 | 16.67 | 11.33 | 13.33 | 15.33 | 10.67 |
| Angle 45 | 15.33 | 12.67 | 14.00 | 10.67 | 10.00 |
| Angle 90 | 18.67 | 18.00 | 11.33 | 12.67 | 14.67 |
| Angle 135 | 14.00 | 16.67 | 14.67 | 11.33 | 10.67 |

Table 8 and Figure 7 depict the recognition success rates graph of 150 test images using Minkowski distance measurement. After testing be done, this recognition system with GLCM method and Minkowski distance produce the highest accuracy rate of 18.67% which is at a distance of 1 pixels and a single orientation angle of 90°. While the lowest accuracy rate is 4.00% when using an orientation all angle within 5 neighboring pixels and using an orientation angle.
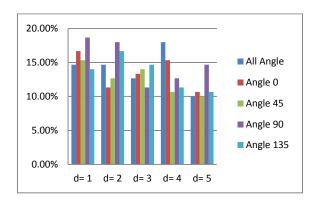


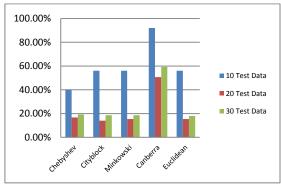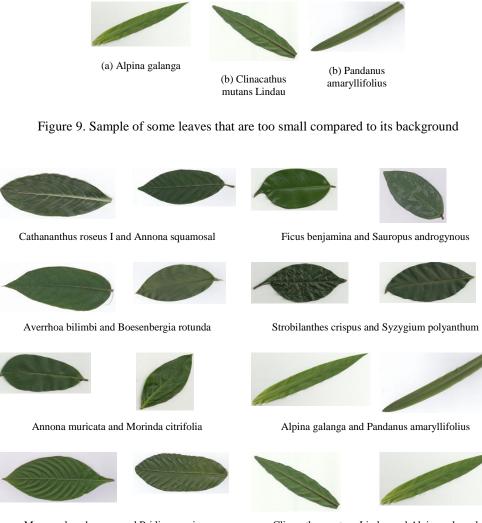Figure 7. Success Rates Graph of Test Image
Recognition Minkowski



Figure 8. Graph of Success Rate Introduction of 10,
20 and 30 Leaf Types

Figure 8 depicts a graph of the success rate of 10, 20 and 30 leaf images. From the test using 10 (ten) test image data with various distances used for identification, it can be shown that testing using Canberra distance is the best with recognition rate 92%.

There are some most probably reasons that influence the recognition rates are not too good results. *First,* some objects are too small if compared to its background, for example: Alpina galanga, Clinacathus mutans Lindau, and Pandanus amaryllifolius. This situation tends to give false recognition results. Figure 9 depicts some leaves with its object is too small if compared too its background. *Second,* some leaves have almost similar texture as the others. This similarity tends to give false recognition results too. Figure 10 shows eight pairs of different leaves with similar texture.



(a) Alpina galanga

(b) Clinacathus mutans Lindau

(b) Pandanus amaryllifolius

Figure 9. Sample of some leaves that are too small compared to its background



Cathananthus roseus I and Annona squamosal

Ficus benjamina and Sauropus androgynous

Averrhoa bilimbi and Boesenbergia rotunda

Strobilanthes crispus and Syzygium polyanthum

Annona muricata and Morinda citrifolia

Alpina galanga and Pandanus amaryllifolius

Mussaenda pubescens and Psidium guajava

Clinacathus mutans Lindau and Alpina galangal

Figure 10. Eight pair of leaves with similar texture

Not all leaf images can be well recognized by the application because the recognition process of the leaf image to be recognized still depends on the number of leaf test images. Normally, the more data tested the tendency of recognition results to increase. However, when we add more data to be tested which its texture is similar to other types of leaf, the recognition will be false. Therefore, in this research, the addition of test data even lowers the recognition rate. Also, the characteristics of the leaves can be seen from other aspects such as color, size, shape, texture, and odor for which they are not be involved in this research. Therefore, some research involving the other aspects mentioned above to recognize the herb leaves are suggested to produce the best recognition results.

## 4.    CONCLUSION

In this research, the classification based on the recognizing the leaves images with extracted texture features was proposed and performed.  The texture features have been extracted with using the GLCM, on the 150 images in dataset leaf images for the test. The recognition system using 10 leaf test image data with GLCM method and Canberra distance resulted in the highest accuracy of 92.00%. While the use of 20 and 30 test data, the recognition rates are 50.67% and 60.00%, respectively. The lowest percentage rate in this recognition system is 4.67% when using the City-block distance while on the 20 test data. Moreover, in the future works, next researchers can either use more images or other methods in order to compare the results of the current research with their results, for example Principal Component Analysis method, Haar Feature Selection, and Independent Component Analysis.

## REFERENCES

[1]    A. Ehsanirad, and S. Kumar, "Leaf recognition for plant classification using GLCM and PCA methods", *Oriental Journal of computer Science & Technology*, vol. 3, no. 1, pp. 31-36, 2010.
[2]    R.R. Isnanto, *et al.*, "Identifikasi Retina Mata Menggunakan Jarak Euclidean Dengan Pencirian Matriks Kookurensi Aras Keabuan", *Jurnal Sistem Komputer*, vol. 6, no. 1, 2016.
[3]    A.P. Akshay and K.S. Bhagat, "Plants Identification by Leaf Shape using GLCM, Gabor Wavelets, and PCA," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 37, no. 3 July 2016.
[4]    D. Singh, and K. Kaur, "Classification of Abnormalities in Brain MRI Images Using GLCM, PCA, and SVM," *International Journal of Engineering and Advanced Technology (IJEAT)*, ISSN: 2249 – 8958, vol. 1, no. 6, Agust 2012.
[5]    M. Madhu and R. Amutha, "A Combinatorial Approach to Human Face Recognition and Expression Identification", *Academic Journals,* vol. 8, no. 31, pp. 1464-1470, 2013
[6]    R.R. Isnanto, *et al.*, "Implementation of Gray Level Cooccurrence Matrix (GLCM) for Human Iris Recognition System", *Proceedings of Conference on Information Technology and Electrical Engineering (International Session) CITEE 2010*, Juli 2010, Yogyakarta, pp. 95-101
[7]    McAndrew dan Alasdair, 2004, "An Introduction to Digital Image Processing with Matlab", *Notes for SCM2511 Image Processing 1, School of Computer Science and Mathematics Victoria University of Technology*.
[8]    J.R. Parker, 1994. *Practical computer vision using C*. John Wiley & Sons, USA.
[9]    S.G. Wu, *et al.* "*A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network*", Chinese Academy of Sciences, Beijing, 2007.
[10]   A. Kadir, "A Model of Plant Identification System Using GLCM, Lacunarity, and Shen Features", *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, vol. 5, no. 2, 2014, pp. 1-10.
[11]   A. Ehsanirad, "Plant Based on Leaf Recognition", *(IJCSIS) International Journal of Computer Science and Information Security,* vol. 8, no. 4, July 2010.
[12]   A.H. Kulkarni, *et al.*, "A Leaf Recognition System for Classifying Plants Using RBPNN and Pseudo Zernike Moments", *International Journal Latest Trends in Engineering and Technology*, 2011, vol. 2, no. 1, pp. 6-11.
[13]   Y.J. Zhang, Image Engineering, vol. 1, De Gruyter, Tsinghua University Press, Beijing, 2017 pp. 246.

## BIOGRAPHIES OF AUTHORS

**R. Rizal Isnanto** received the Bachelor and Master degrees in Electrical Engineering of Gadjah Mada University, Yogyakarta – Indonesia, in 1994 and 2002 respectively. While, the Doctoral degree is received in Electrical Engineering and Information Technology Department at the same university. As a lecturer, now he is an Associate Professor in Computer Engineering Department, Faculty of Engineering, Diponegoro University, Semarang – Indonesia. His research interests include biomedical image processing and biometrics as well as in pattern recognition.

**Munawar Agus Riyadi** received the Bachelor and Master degrees in Electrical Engineering from Bandung Institute of Technology, Bandung – Indonesia, in 2000 and 2002 respectively. While, the Doctoral degree is received from Universiti Teknologi Malaysia in 2012. As a lecturer, now he is an Associate Professor in Electrical Engineering Department, Faculty of Engineering, Diponegoro University, Semarang – Indonesia. His research interests include Electronics and Telecommunication.

**Muhammad Fahmi Awaj** received the Bachelor in Computer Engineering Department, Faculty of Engineering, Diponegoro University, Semarang, Indonesia in 2014. While, his Master degree was received in 2017 from Electrical Engineering Departement, in the same university he received the Bachelor degree. His research interests include Embedded Systems as well as in Image Processing and Pattern Recognition.

**ATTACHMENT**

Table A. Images of All Herb Leaves Observed

| No. | Binomial (Scientific) name | Indonesian (local) name | Picture of leaf |
|---|---|---|---|
| 1 | Pluchea indica | *Beluntas* | |
| 2 | Ficus benjamina | *Beringin* | |
| 3 | Anredera cordifolia | *Binahong* | |
| 4 | Ricinus communis | *Jarak* | |
| 5 | Hemigraphis alternata | *Sambang Darah* | |
| 6 | Alpina galanga | *Lengkuas* | |
| 7 | Strobilanthes crispus | *Keji Beling* | |
| 8 | Polyscias scutellaria | *Mangkokan* | |
| 9 | Morinda citrifolia | *Mengkudu* | |
| 10 | Mussaenda pubescens | *Nusa Indah* | |
| 11 | Stachytarpheta jamaicensis | *Pecut Kuda* | |
| 12 | Carica papaya | *Pepaya* | |
| 13 | Syzygium polyanthum | *Salam* | |

| No. | Binomial (Scientific) name | Indonesian (local) name | Picture of leaf |
|-----|---------------------------|-------------------------|-----------------|
| 14 | Piper ornatum | *Sirih Merah* | |
| 15 | Annona muricata | *Sirsak* | |
| 16 | Psidium guajava | *Jambu Merah* | |
| 17 | Zamioculcas zamifolia | *Dolar* | |
| 18 | Ixora javanica | *Soka* | |
| 19 | Lobelia inflata | *Lobelia* | |
| 20 | Piper betle | *Sirih Hijau* | |
| 21 | Citrus hystrix | *Jeruk Purut* | |
| 22 | Sauropus androgynus | *Katuk* | |
| 23 | Clinacanthus nutans Lindau | *Dandang Gendis* | |
| 24 | Curcuma longa | *Kunir* | |
| 25 | Catharanthus roseus I | *Tapak Dara Merah* | |
| 26 | Averrhoa bilimbi | *Belimbing Wuluh* | |
| 27 | Annona squamosa | *Srikaya* | |
| 28 | Catharanthus roseus II | *Tapak Dara Putih* | |
| 29 | Boesenbergia rotunda | *Kunci* | |
| 30 | Pandanus amaryllifolius | *Pandan* | |