MEDICINAL PLANT SPECIES IDENTIFICATION SYSTEM USING TEXTURE ANALYSIS AND MEDIAN FILTER

[a]Prihastuti Harsani, [b]Arie Qurania, [c]Triastinurmiatiningsih
[a,b]Program Studi Ilmu Komputer FMIPA Universitas Pakuan , Bogor 16143
[c]Program Studi Biologi, FMIPA Universitas Pakuan, Bogor 16143
E-mail: prihastuti.harsani@unpak.ac.id

Abstract

Identification of plants can be done through objects - objects in plants by asking an expert or through a specimen (herbarium) that have been identified previously. Identification is done by matching the pictures in the book of flora or monograph. Computer-aided identification can be done using digital image processing methods which utilize digital image matching object plant with a picture on the book. Identification key that is used is the image of the leaves. This study develops previous research has identified using the method of fractal and Euclidian Distance. Accuracy obtained in each of the identification system for the fractal dimension and fractal code is of 68% and 51%. Improved accuracy is the main objective of this study. The proposed method is a method of texture analysis and median filter. Texture analysis is used as feature extraction technique while the median filter is image enhancement techniques. Based on the trials, the results of the identification of texture analysis method and median filter to increase to 78%. Median filter is used as a technique to improve the image quality leaves. The use of an identification system to be tested in the web application of information systems of medicinal plants.

Keywords: identification, medicinal plant, image processing, fractal, texture analysis, median filter

INTRODUCTION

Identification system is a mathematical representation of the modeling process of an object based on the experimental data. Identification is required when there are one or more objects have characteristics in common, so it takes a differentiator between the two objects. Object identification can be done through the morphology and characteristics that appear. In plants, the identification can be easily distinguished by flowers, fruits and seeds. The identification of this object is plagued by problems arising by season. Other objects that can be used for identification are the leaves. Another obstacle is the use of the leaves as the identification of several different plants have a similar leaf shape. This will affect the accuracy of the identification results, especially if the identification is done manually. The experts identify through several ways, including: by asking the other experts and through the identification of key plant and match objects with the image on the flora book / monograph.

Indonesia is a country rich in medicinal plants, and the potential for development, but has not been managed optimally. Natural wealth of plants in Indonesia covering 30,000 species of plants of a total of 40,000 species of plants in the world, 940 types of which are medicinal plants (this amount represents 90% of the number of medicinal plants in Asia) [1].

Indonesia's wealth of medicinal plants species opens opportunities for new medicinal plant research. Identification of medicinal plants is necessary to note that the recording species of medicinal plants in Indonesia can be recorded and identified properly. The use of plants as medicines generated the need for plant identification. Identification is performed to determine whether a particular plant is included in the medicinal properties of plants or not. Identification of plant species as well as medicinal plants can be done through digital image processing method. Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it [2]. Method of digital image do the matching digital data objects plant database with input data. This method will extract the main characteristic

feature of the image to be a factor. These factors will be the distinguishing traits in each species of plant. After characterizing factor is obtained, the classifier will identify the plants. Feature image can consist of colors, shapes and textures. In contrast with the color and the shape, the texture of the image can be taken through the matrix co-occurrence intensity of the image. Co-occurrence intensity matrix illustrates how often the appearance of a couple of two pixels with a certain intensity within a certain distance and direction that occurs in the image [3].

The study of plant species identification through the image of the leaves has been done by applying the method of fractal as feature extractor [4]. The resulting identification accuracy of 68% for the fractal dimension and 51% for the fractal code. To improve the accuracy value, there are several alternative steps that are used to improve this accuracy, including improvement of image quality and other traits using extraction techniques eg texture analysis.

Feature extraction used is the analysis of texture. Texture analysis is used because the texture of the leaves are hard to see by naked eye so it needs to be processed using digital image processing techniques. The method used in this research is the statistical method. Statistical methods using statistical calculations degrees of gray distribution (histogram) by measuring the level of contrast, granularity, and the roughness of a region of adjacency relationship between pixels in the image. These statistical paradigm use is not limited, so it is suitable for natural textures unstructured of sub patterns and the set of rules (microstructure).

One technique to obtain the characteristics of statistics is to calculate the probability of adjacency relationship between two pixels at a certain distance and angular orientation. This approach works by forming a matrix co-occurrence of the image data, followed by determining the characteristics as a function of the temporary matrix [5]. Identification is done by measuring the closeness between plants based on characteristics acquired. The measurement method used is Euclidian Distance.

MATERIAL AND METHODS

The method used in this study using the methods of systems development and image processing. In general, system architecture consists of: database of medicinal plants, medicinal plants information systems and web-based systems for identification of medicinal plants. There are four (4) steps being taken to produce a database of medicinal plants and identification applications as shown in Figure 1.
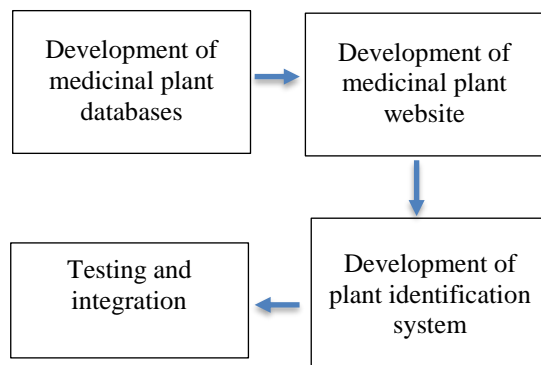
Development of medicinal plant databases → Development of medicinal plant website → Development of plant identification system → Testing and integration

Figure 1. Steps of system development

Making the database planting

The database is used as the basis of information to be displayed on the web. Source of data derived from books on medicinal plants [medicinal plants IPB] [6] and direct observation Medicinal Gardens in Bogor, the Medicinal Plant Garden Sringanis.

Making of web-based medicinal plant encyclopedia

Web medicinal plants was prepared as an information provider medicinal plants in Indonesia. Information provided include: general information medicinal plants, properties and processes. Making a web of medicinal plants through the following phases: identification, design implementation and testing.

Making an identification system

Application identification is an application that is assigned into the information systems of medicinal plants. Stages of manufacture of medicinal plant identification applications using image processing stages as presented in Figure 2.
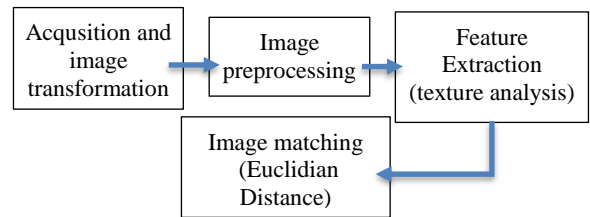
Acqustion and image transformation → Image preprocessing → Feature Extraction (texture analysis) → Image matching (Euclidian Distance)

Figure 2. Digital image leaves processing

Image acquisition

The image of the leaves is taken directly through two ways, outdoor and indoor. The leaves image that are used are the leaves without background. The image data captured by 20 of image data, each leaf is taken 30 times on the different leaves. 20 data is used as training data to generate a classifier and 10 data will be used as test data to determine the accuracy of the classifier.

Transformation image

Digital image leaf transformed into a gray scale image data using Luminosity equation method to the Equation (1) :

$$0.21 \, R + 0.72 \, G + 0.07 \, B \qquad (1)$$

luminosity method is used for converting color images into gray level image due to the nature of luminosity methods that are closer to human visual perception [7].

Preprocessing

Preprocessing undertaken to improve image quality, including noise removal, sharpness enhancement, color correction, etc.

This study uses median filter technique, which uses a 3x3 filter window [8], the algorithm is as follows:

Read f (y, x), the pixel at the position (y, x).
Sort 9 pixel values.
Find the median value of the pixel values to be changed with other 8 neighborhood Change the pixel value f (y, x) to g (y, x) based on the median value of the pixel D.

Feature Extraction

Feature extraction classified into three types: low-level, middle-level and high-level. Low-level feature is a feature extraction based on visual content such as color and texture, is a

middle-level feature extraction based on the specified image region segmentation, while the high-level feature extraction feature is based on the semantic information contained in the image [9].

Feature extraction used is to use texture analysis with statistical methods Order 2. Usage statistics of order 2 is made to obtain an image differentiator. The distinguishing feature of this can not be obtained through statistical order 2. In this study there were four second-order statistics used are: Angular Second Moment (Energy), Contrast, Inverse Difference Moment (homogeneity), and Entropy. Based on the previous observations made by Mohaniah in 2013 [10] show that these texture features have high discrimination accuracy, requires less computation time and hence efficiently used for real time Pattern recognition applications.

Angular Second Moment (ASM)/Energy

ASM or energy states uniformity levels of pixels of an image. The higher the energy, the more uniform texture.

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p^2(i,j) \qquad (2)$$

p is normalized gray-tone cooccurrence matrix *defined by*:

$$p(i,j) = \frac{C(i,j)}{\sum_i \sum_j C(i,j)} \qquad (3)$$

$C(i,j)$ is the cooccurence matrix with G dimension.

Contrast

Contrast iIndicates the size of the deployment (moment of inertia) elements of the image matrix. If located far from the main diagonal, the value of great contrast. Visually, the contrast value is a measure of the variation between the degree of gray an image area.

$$Contrast = \sum_{i=1}^{G} \sum_{j=1}^{G} (i-j)^2 p(i-j) \qquad (4)$$

Homogenity

Homogeneity expressed proximity of each element of the cooccurrence matrix.

$$Homogenity = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i,j)}{1+(i-j)^2} \qquad (5)$$

Entropy

Entropy Showed a measure of disorder form. ENT price is great for the image with uneven degrees of gray transitions and image of little value if the structure is irregular (variable). Entropy equation shown in Equation (5).

$$Entropy = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) log_2 p(i,j) \qquad (6)$$

Test of Classifier

The method used is k-fold cross validation. K-fold method classifier is a method for measuring the strength of the model.

K-fold cross-validation is one of the techniques to evaluate the accuracy of the model, with characteristics [9]:
1. Partitioning the data randomly into k sets / fold namely D1, D2, ..Dk Each class has nearly the same amount.
2. At iteration is used in a set of test data and the other as the training data.
3. Conduct training and testing as much as k times.
4. Calculating the accuracy of the formula.

Identification

Based on the characteristics of the image, identification is done by measuring the distance between the data. Identification using Euclidean Distance is used to measure the distance between the input data with measured data.

The accuracy of the identification results calculated using confusion matrix (Table 1). Rated accuracy is obtained by measuring how percent of the data that has been identified appropriately based classifier.

Table 1. Matriks Confussion

	Classified Positive	Classified Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

The performance of the system is calculated based on Equation (7) :

$$Accuration = \frac{(TP+TN)}{(TP+FN+FP+TN)} \qquad (7)$$

TP = True Positive
FN = False Negative
FP = False Positive
TN = True Negative

RESULT AND DISCUSSION

This study develops applications early identification of medicinal plants through the leaves image. Data medicinal plants collected from several sources, including Medicinal Gardens Sringanis, Medicinal Plant Garden State University Pakuan. There are 45 types of plants which are divided into 45 classes. Each plant consists of 10 images of leaves. Eight image leaves into training data and two image of the leaves into test data.

These web-based applications built using MySQL database and PHP programming language. Identification application consists of several modules, including medicinal plants database module, feature extraction module and the module identification.

Image Acqusition

Imagery obtained by making directly in the field. Plants taken from different trees. The image of the leaves will be taken front look. Next process is cropping the background so that the only remaining image of its leaves. For some plants are quite rare, the image of the leaves are not taken from the stem but leave the white paper as a screen in the image.

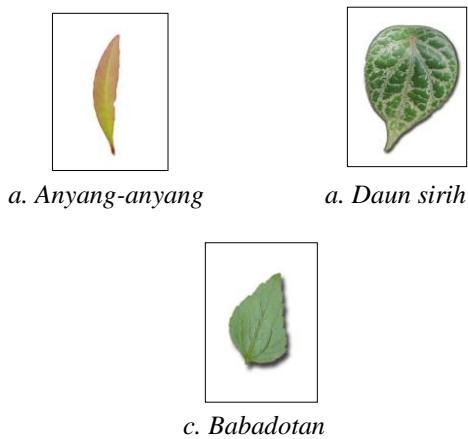The image of the leaves are stored in the database module (Figure 3).

a. Anyang-anyang   *a. Daun sirih*

c. Babadotan

Figure 3. Examples of image leaves

Median Filter Module

Median filter is one technique for image enhancement. Using the median filter in the identification system is used to sharpen the image contrast.

Feature Extraction Module

Feature extraction module is a module that is used to perform feature extraction that produces identifier. Feature extraction technique used is the analysis of texture. In texture analysis obtained some identifier be a major element for the identification process. The identifier are: entropy, energy, contrast and homogeneity. Entropy measures the randomization of the intensity distribution of the image pixels. Energy is a feature to measure the concentration of the intensity in the pair co-occurrence matrix. The greater the energy value if the spouse is eligible matrix pixel intensities are concentrated in a few coordinates co-occurrences and shrink when lying spread. Another feature is the contrast that is used to measure the strength of the difference in intensity in the image. Values greater contrast if the variation in the image of high intensity and low intensity decreases when variations. The opposite of contrast is intensity variations in the image. Values greater homogeneity if the variation in the intensity of the image decreases and vice versa. The results of the database image feature extraction of the leaves is as shown in Figure 4.

Figure 4. Feature extraction

The median filter applied in the plant leaf image will result in a single value to be identified in the system. The result of the application of the median filter on the leaf plant image is presented in Figure 5 and Figure 6.
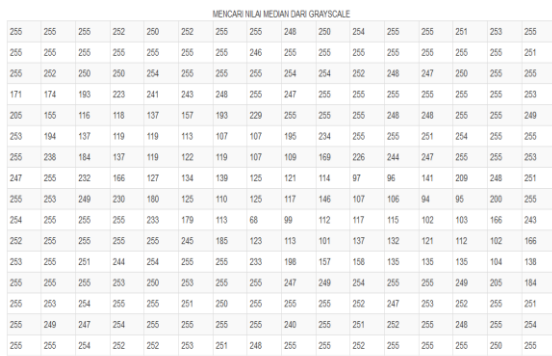
Figure 5. Identification Result

Figure 6. Pixel Result of Median Filter

Figure 6 is the pixel value of one of the leaf images. This value is the result of a median filter with a 3x3 window size.

Identification Module

Image identification is done using Euclidean Distance. This method of classification is based on its closest neighbors. Euclidian distance will measure the distance or difference between two objects into the test data identification.

The smallest difference value means it has the closest and most appropriate distance as the two objects are the most suitable.

The results of the identification system is presented in Figure 5.
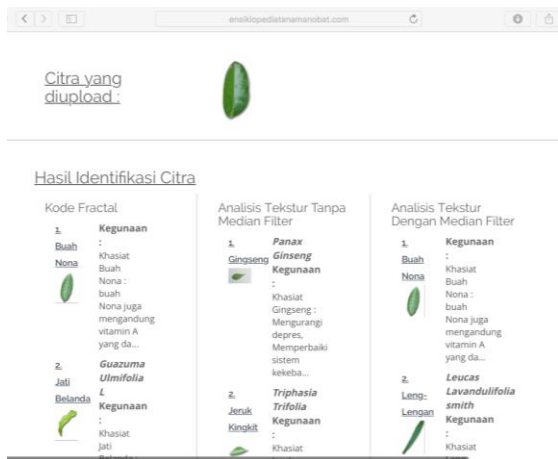

Figure 5. Identification Result

Trial identification is done by entering keywords leaf image that class has been known. The system will identify and examine the accuracy provided by the system. Measurement accuracy is accomplished by using the matrix confusion (7) for feature extraction method with a median filter is obtained as shown in Table 2.

Table 2. Results of Test Identification

	Classified Positive	Classified Negative
Actual Positive	35	7
Actual Negative	13	37

Based on the results in Table 2, it appears that accuracy for identification by Equation (7) is 60%.

Texture analysis is one of the feature extraction method used to acquire the image based on color texture, surface, and homogeneity.

Some feature extraction techniques have been studied to obtain the characteristics of each image factor leaves. The methods used include the fractal dimension and fractal code.

The results showed no significant accuracy, at 68% for the fractal dimension and 51% for the fractal code [2][11]. Repair accuracy value can be done through several stages in the identification of the image, including: a preprocessing stage, feature extraction and change the stage classifier for identification. Median filter is one of image enhancement techniques. This method performs the detection of the presence of noise and reduce it by taking the median value of neighboring pixels. The use of median filter is made to see the effectiveness of preprocessing stages in determining the value of accuracy.

Test accuracy is done with two scenarios as follows:
1. Trial to determine the accuracy of the characteristics using the k-fold cross validation.
2. Making a test to determine the accuracy of the identification of plant. Table 3 shows the accuracy of each method.

Table 3. Accuracy with Feature Extraction Techniques

	Classes	Identified classes	Accuracy
Fractal dimension	45	31	68%
Fractal code	45	23	51%
Texture analysis	45	27	60 %
Texture analysis	45	35	78%

Feature Accuracy for Each Class

Feature accuracy obtained by k-fold cross validation. There are 45 classes of data plants, each plant data consists of 10 images of leaves. Each class will be divided into 7 training data and test data 3. Based on the test results, the obtained value of the highest accuracy is 3 fold as much as 5 times.

Based on test results obtained that characterize the contrast produces the highest accuracy values. It is addressed that the contrast characteristic has the most impact on plant identification.
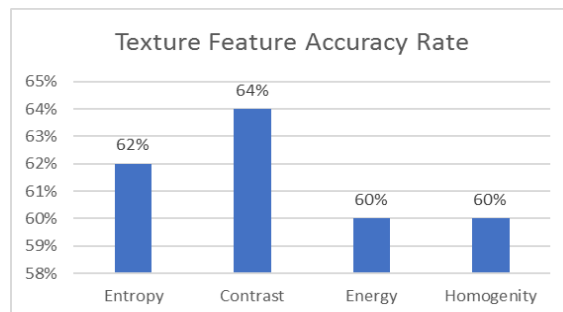

Figure 6. Accuracy each feature

Based on trial results shows that the application of the median filter on the identification of plant species provide most excellent accuracy value. Analysis of texture when compared with the analysis of the characteristics also still produce the most good accuracy values.

These results indicate that the texture analysis feature extraction techniques that are GCLM Grey Level Co-occurrence Matrices can produce better when combined with other techniques.

This is also seen in a study conducted by Zhaobin in 2017 [12]. This study analyzed the methods for the identification of plants through 16 shape features, 11 texture features, 4 color features. The method of classification used 8 methods. In the research it was proposed that the recognition rate through two features gave better results compared to the use of 1 feature. Based on the research also the highest rate of plant identification is given through texture features found in Average and Correlation features. The four features used in this study received recognition rates of 6.43, 6.49, 8.6, 9.5. To improve the accuracy it is necessary to add texture features and feature shapes and colors.

CONCLUSION

Identification of the medicinal plants through plant leaves image object identification is a stage of early identification of a plant species. Digital image-based identification can be through a characteristic identifier including shapes, textures and colors. Identification through identifier forms has been carried out using the method of fractal and produce two grades of accuracy for fractal dimension and code that is respectively 68% and 51% [4]. Accuracy increased to 78% after use texture analysis as extraction techniques characteristics.

Application of the techniques of image enhancement has been made to show the effectiveness of the identification results. It is seen that the results showed an increase for the accuracy of the image that are not applied to the median filter to the application of the median filter is equal to 30%.

REFERENCES

[1] Departemen Kehutanan, "Lokakarya Nasional Tanaman Obat Indonesia," in *Sinergi Multipihak Dalam Budidaya, Pelestarian Dan Peningkatan Kualitas Tanaman Obat Indonesia*, Gedung Manggala Wanabakti, 2010.

[2] Rafael C. Gonzalez and Richard E. Woods, *Digital Image Processing*. New Jersey: Prentice Hall, 2002.

[3] P. Jackman and D.-W. Sun, "Recent advances in image processing using image texture features for food quality assessment," *Trends Food Sci. Technol.*, vol. 29, pp. 35–43, Jan. 2013.

[4] Prihastuti Harsani, Iyan Mulyana, and Prasetyorini, "Application of Image Retrieval Using Fractal Dimension to Identify Medicinal Plant," presented at the International Conference on Science Technology Innovation, Jakarta, 2012.

[5] A. Materka, M. Strzelecki, and others, "Texture analysis methods–a review," *Tech. Univ. Lodz Inst. Electron. COST B11 Rep. Bruss.*, pp. 9–11, 1998.

[6] Gagas Ulung, *Sehat alami dengan herbal*. Kompas Gramedia, 2014.

[7] D. Biswas, "Novel Gray Scale Conversion Techniques Based On Pixel Depth," *J.*

Glob. Res. Comput. Sci. Approv. J., vol. 2, no. 6, pp. 118–121, Jul. 2011.

[8] B. Yuwono, "Image Smoothing Menggunakan Mean Filtering, Median Filtering, Modus Filtering dan Gaussian Filtering," *Telematika*, vol. 7, no. 1, 2015.

[9] O. Marques and B. Furht, *Content-Based Image and Video Retrieval*. Springer Science & Business Media, 2002.

[10] P. Mohanaiah, P. Sathyanarayana, and L. Guru Kumar, "Image texture feature extraction using GLCM approach," *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, p. 1, 2013.

[11] Prihastuti H, Arie Q, and Triastinurmiatiningsih, "Sistem Identifikasi Tanaman Obat Menggunakan Kode Fraktal," presented at the Seminar Nasional Kualitas Hidup Melalui IT dan Manajemen, Palembang, 2014.

[12] Z. Wang, H. Li, Y. Zhu, and T. Xu, "Review of Plant Identification Based on Image Processing," *Arch. Comput. Methods Eng.*, vol. 24, no. 3, pp. 637–654, Jul. 2017.