Computer Vision for Plant Identification

Yeni Herdiyeni* and Julio Adisantoso

Department of Computer Science, Faculty of Mathematics and Natural Sciences,

Bogor Agricultural University, INDONESIA

Tel. +62 251 8625584, Fax. +62-251-8625584, E-mail: yeni.herdiyeni@ipb.ac.id

*Contact person and presenter

Abstract

Regarding to problems in plant identification, Botanist need a tool to assist them in finding the plant taxonomy. There are two goals of plant taxonomy are identification and classification. Computer vision (image understanding) is a discipline that studies how to reconstruct, interpret and understand a 3D scene from its 2D images in term of the properties of the structures present in the scene. We have extensively conducted some research to develop computer vision for plant identification based on image especially for Indonesian medicinal plants and house plant. In the experiment, we used three features for plant identification i.e. morphology, shape and texture. Also, we have conducted some research to classify the plant. There are several classifier algorithms that we used, i.e., Probabilistic Neural Network, Radial Basis Function Neural Network, and Support Vector Machine. To help user finding information easily about the plants, we also conducted research to develop search engine system based on article and research documents. The experiments show that our systems are very promising. In the future we need to explore more extensively mathematical model as function to model the plant in order to improve the accuracy in plant identification and classification.

I. Introduction

According to ((Damayanti E.K., Hikmat A., and Zuhud, E. 2011) the problems found in identification of medicinal plants are (1) limited number of local people who knows, recognizes, and understands medicinal plants diversity and usage that can identify plants in the field, (2) limited number of plant identification books/guides with many pages in each books/guide which is troublesome to be brought to the field, (3) limited number of authorized institution and facilities for plants identification, (4) limited number of taxonomists in each authorized institution that can scientifically identify enormous plant species, (5) time and money consuming, as effects of problems (1) to (4). They need a tool to assist them in plant identification.

Identification is simply the determination of the similarities or differences between two elements, i.e., two elements are the same or they are different. The comparison of an unknown plant with a named specimen and the determination that the two elements are the same also involves classification, i.e., when one correctly decides that an unknown belongs to the same group (species, genus, family., etc.) as a known specimen, the information stored in classification systems becomes available and applicable to the material at hand (Walters, D.R. & D.J. Keil, ..). Both of processes, identification and classification, involve comparison and judgment and require a definition of criteria of similarities. Plant Identification is determination of identity of unknown plant by comparison with previously collected specimen and plant classification is the process of finding the taxon to which a specimen belongs.

The methods of identification include **(1) expert determination, (2) recognition, (3)**

comparison, and (4) the use of keys and similar devices. In terms of reliability or accuracy the best method of identification is expert determination. Although of great reliability, this method *presents problems by requiring the valuable time of experts and creating delays for identification*. Recognition is the approaches expert determination in reliability. This is based on extensive, past experience of the identifier with the plant group in question. *In some groups this is virtually impossible*. A third method is by comparison of an unknown with named specimens, photographs, illustrations or descriptions. Even though this is a reliable method, *it may be very time consuming or virtually impossible-due to the lack of suitable materials for comparison.* The use of keys or similar devices (synopses, outlines, etc.) is by far the most widely used method and does not require the time, materials, or experience involved in comparison and recognition.

\

To overcome these problems, we have been conducted research to develop automatically system for plant identification based on image using computer or computer vision for plant identification. Plant identification system is done based on morphology, shape and texture. Plant image that we have used in the experiment is leaf of medicinal plant and house plant.

II. Computer Vision for Plant Identification

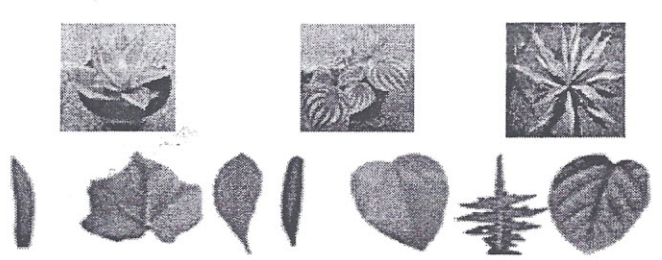
II.1. Computer Vision

Computer vision is concerned with modeling and replicating human vision using computer software and hardware. It combines knowledge in computer science, electrical engineering, mathematics, physiology, biology and cognitive science. It needs knowledge from all these fields in order to understand and simulate the operation of human vision system.

Computer vision overlaps significantly with the following fields: image processing, pattern recognition and photogrammetry. Image processing focuses on image manipulation to enhance image quality, to restore an image or to compress/decompress an image. Pattern recognition studies various techniques (such as statistical techniques, neural network, support vector machine, etc.) to recognize/classify different patterns. Photogrammetry is concerned with obtaining accurate and reliable measurement from images. It focuses on accurate mensuration.

Computer vision (image understanding) is a discipline that studies how to reconstruct, interpret and understand a 3D scene from its 2D images in term of the properties of the structures present in the scene. The computer vision consists of three levels, i.e.

1. Low level vision: process image for feature extraction (edge, corner, or optical flow).
2. Intermediate level vision: object recognition and 3D scene interpretation using features obtained from low level vision



3. High level vision: interpretation of evolving information provided by the intermediate level

vision as well as directing what intermediate and low level vision task should be performed. Interpretation may include conceptual description of scene like activity, intention and behavior.

We have been conducted extensive research to develop a computer vision for plant identification automatically. The plant identification system can be used to identify leaf of medicinal plant and house plant. Figure 1 show leaf of medicinal plant and house plant. Images were taken from Botanical Garden Bogor and conservation area at Gunung Leutik Bogor.

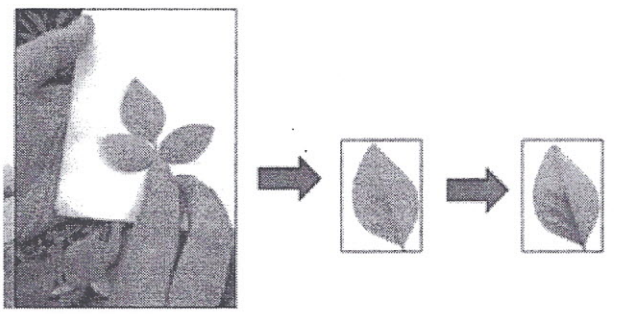


Figure 1. Sample images

II.2. Feature Extraction

Feature extraction is low level vision that providing information for later higher level analysis. It is an essential step in computer vision, performed either separately or jointly with learning process: preprocessing and feature construction are followed by feature selection. Figure 2 show preprocess of image acquisition and image enhancement of leaf. Feature selection is to select the best subset from the input space. Its ultimate goal is to select the optimal features subset that can achieve the highest accuracy results. In our system, we used three features for plant recognition i.e. morphology, texture, and shape.

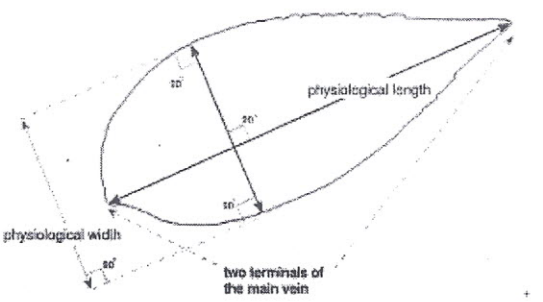


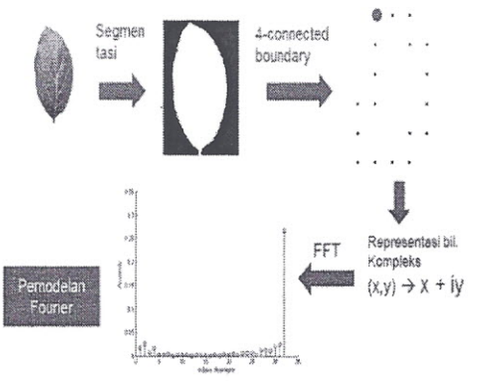
Figure 2. Image selection and enhancement

Morphology

Plants are basically identified according to their morphological features. We have been observed morphology of the leaf. There are two features of leaf morphology, i.e. basic and derivatives features. The basic features are the diameter, leaf area, and perimeter of leaves.

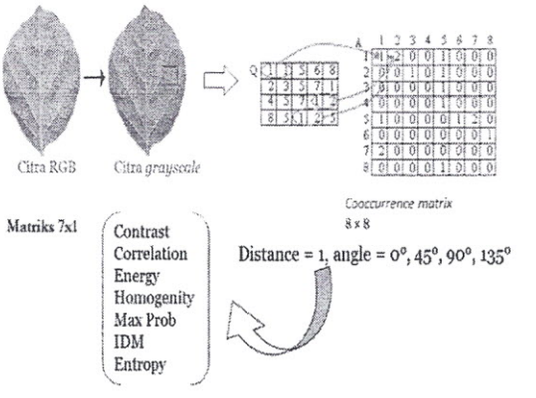
Diameter is the furthest point between the two points from boundary of the leaf. Area is number of pixels that are inside the leaf edge. Perimeter is the number of pixels that are on the edge of the leaf. From three basic features, it can be combined to getting eight derived features such as smooth factor, form factor, rectangularity, perimeter ratio of diameter, perimeter ratio of physiological length and physiological width and vein features.

Figure 3. Physiological length and physiological width



Shape

Shape of plant leaf is one of the most important features for characterizing various plant species. We have used Fourier transformation to extract the shape of leaf (Rahmadhani, M., 2009). Fourier transformation can provide useful descriptions of shape that independent of position, orientation, and size of an image object. For shape extraction, Fourier transformation needs a binary image. Binary image is used to find set of points in leaf curve. Suppose points (x, y) represented as complex number x + iy, N rows of complex number is called boundary. N



complex numbers is transformed to obtain Fourier descriptor an object. Figure 4 show shape

extraction using Fourier Transformation.

Figure 4. Shape extraction using Fourier Transformation

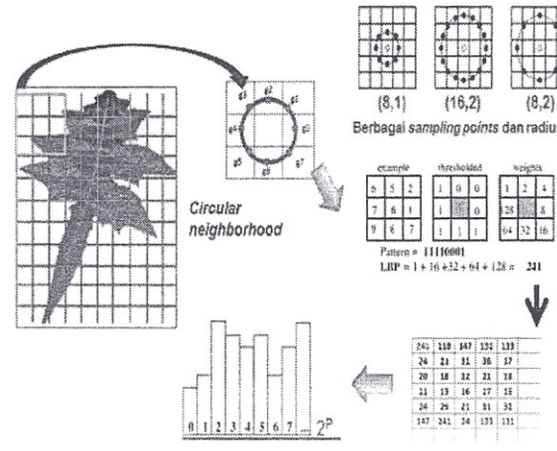
Texture

Co-occurance matrix

Texture analysis plays an important role in many image analysis applications. We have analyzed texture of the leaf using co-occurence matrix (Nurafifah, 2009) and local binary pattern (LBP) (Kulsum, L, 2009). Co-occurrence matrix is a popular representation for texture image using statistical approach. Co-occurrence matrix contain a count of number of times gray level occurs in particular spatial relation to another given features. These matrices well suited for discrimination that are structurally based.

Figure 5. Co-occurrance Matrix

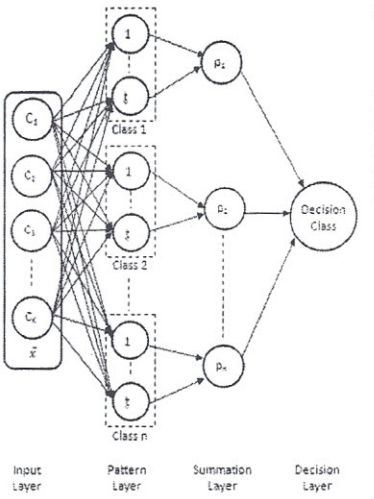
Figure 5 show co-occurance matrix to extract shape of leaf. The position p is determined by the angle θ and distance d. Suppose that in Figure 5, p (1.0), d = 1 and θ = 0°, then A (i, j) is expressed by frequency of occurrence along the i-th gray level g(i) at position p with g



(j). Matrix A can be calculated based on the values of texture features. Co-occurrence matrix is computed in four directions consist 0°, 45°, 90°, 135°. Each image will produce four co-occurrence matrix. Seven texture features are obtained based on the co-occurrence matrix, such as energy, inverse different moment, entropy, maximum probability, contrast, homogeneity.

Local Binary Pattern

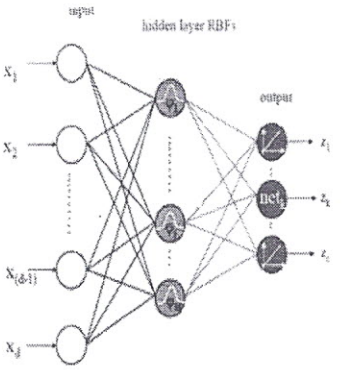
The local binary pattern (LBP) operator was developed as a gray-scale invariant pattern measure adding complementary information to the "amount" of texture in images. According to our experiment, LBP has shown excellent performance in many comparative studies, in terms of both speed and discrimination performance (Kulsum L., 2009). The LBP approach is bringing together the separate statistical and structural approaches to texture analysis, opening a door for the analysis of both stochastic microtextures and deterministic macrotextures simultaneously. The LBP operator can be made invariant against rotation, and it also supports

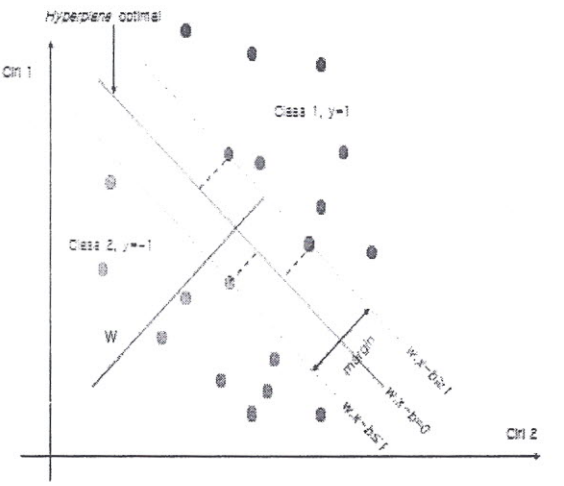


multi-scale analysis.

Figure 6. Local Binary Pattern

Every pixel in an image generates a single LBP code. Then a histogram is created to collect up the occurrences of different LBP codes from all pixels in the image. This histogram forms the LBP feature vector, which characterize the image texture. To be able to deal with textures at different scales, the LBP operator was later extended to use neighborhoods of different sizes. Defining the local neighborhood as a set of sampling points evenly spaced on a circle centered at the pixel to be labeled allows any radius and number of sampling points.





II.3. Feature Classification

Two goals of plant taxonomy are identification and classification. For plant classification we

have been extensively conducted research for plant classification using several algorithms. We have used classification several classification algorithm for example Probabilistic Neural Network (PNN), Radial Basis Function Neural Network (RBFNN), and Support Vector Machine (SVM). Structure of PNN is shown in Figure 7.
Figure 7. Structure of PNN

PNN consist of four layer namely input layer, pattern layer, summation layer, and decision/output layer. Also we compared performance of Probabilistic Neural Network to Radial Basis Function Neural Network (Figure 8 show structure of RBFNN) and Support Vector Machine (Figure 9 show structure of SVM).

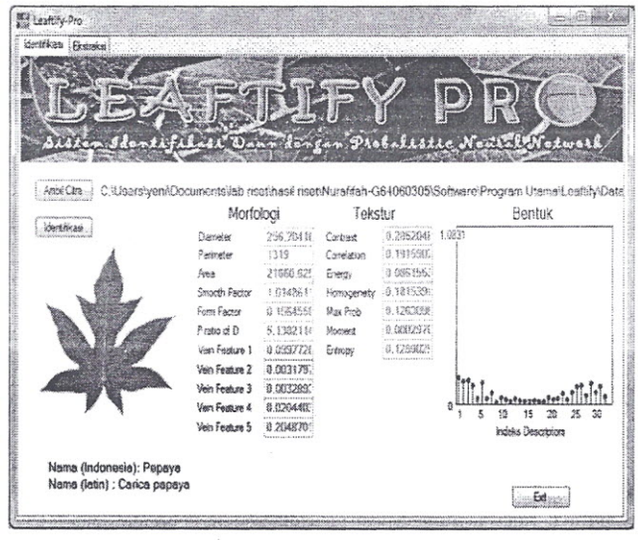
Figure 7. Structure of Radial Basis Function

Figure 8. Structure of Support Vector Machine

II.4. Research Results

Several research results we have conducted are:

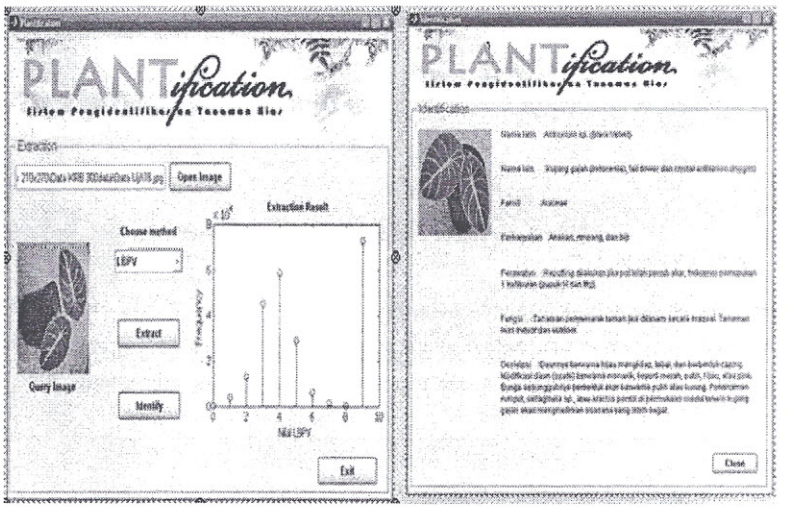
1. Feature extraction of morphological and texture of leaf images for leaf image retrieval (Annisa, 2009). This research use three primary morphological features and three



derivatives morphological features. The primary features are diameter, perimeter, and

area. The derivates features are smooth factor, form factor, ratio of diameter, and five vein features. Co-occurrence matrix is used for texture extraction and Fourier descriptor is used for shape extraction.

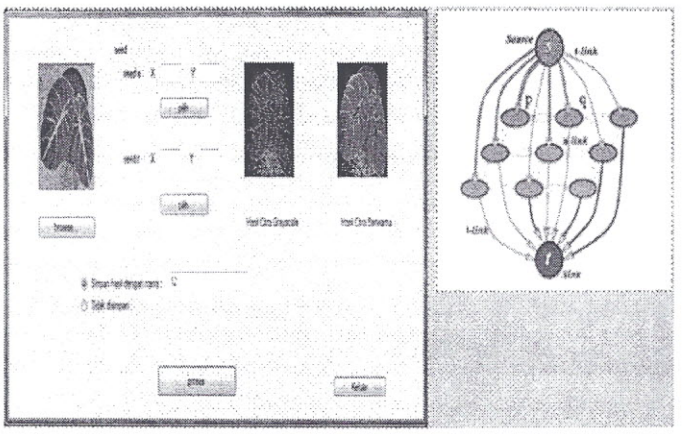
2. Shape and vein feature extraction of leaf using Fourier Descriptor and B-Spline Curve (Rahmadhani, M, 2009). In this research, we used Hough Transform and Fourier Descriptor to extract shape and vein of leaf. Experiments result show that leaf shape represented by Fourier Transform descriptor is more effective than represented by Hough Transform.



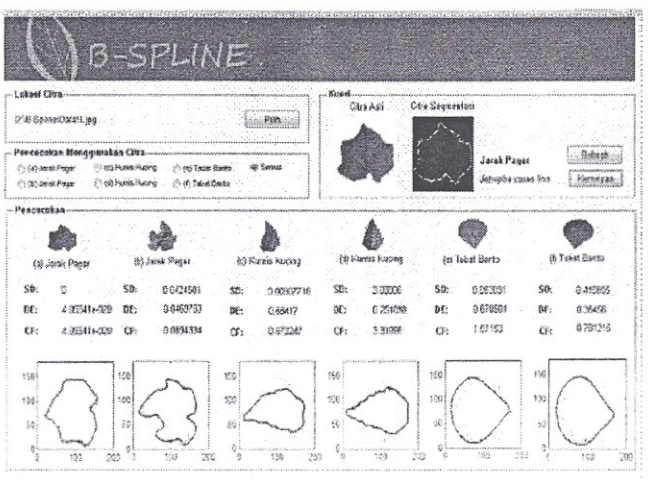
3. A combine morphological, texture and shape feature based approach for leaf identification using Probabilistic Neural Network (Nurafifah, 2010). This research proposed a new system for leaf identification using classifier combination and Probabilistic Neural Network (PNN). The features that used for classification are morphology, texture, and shape of leaves. This research used co-occurrence matrix and Fourier descriptors to extract texture and shape of leaves respectively. After feature extraction, the feature is classified using Probabilistic Neural Network. Classifiers combination is used to combine these features. The experiment data consist of thirty species flora from Bogor Botanical Garden, Indonesia. The experiment results showed that classification without classifier combination has the accuracy of 79.05%, and using classifier combination the accuracy increased to 83.33%. Hence, the proposed system is promising for leaf identification and supporting plant biodiversity in Indonesia.

Figure 10. Plant leaf identification using combination leaf features

4. Automatically plant house identification using Local Binary Patterns Descriptor and



Probabilistic Neural Network (Kulsum, L, 2010). This research proposed a new method for a house plant identification using Local Binary Patterns (LBP) descriptor and



Probabilistic Neural Network (PNN) (Figure 1). There are three kinds of LBP descriptor used in this research, i.e. LBPriu, AR, and LBPV. Database of 300 house plant images belong to 30 different types of house plant in Indonesia are extracted using LBPriu, VAR, and LBPV based on texture feature and classified using PNN. The experimental result showed that LBPV has the best accuracy to indentify house plants with accuracy 73.33%. The proposed system is promising because system is capable to identify house plants species efficiently and accurately.

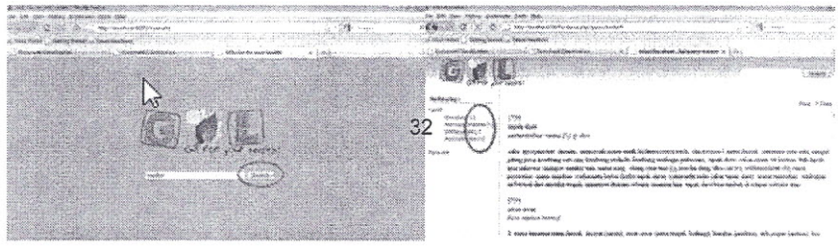
Figure 11. Automatically plant house identification using Local Binary Patterns Descriptor and Probabilistic Neural Network

5. House plant image segmentation using Boycov Method and Kolmogorov Max Flow/Min Cut graph (Wibowo, R. W, 2010). The objective of this research is to remove the complex background of image using Boycov Method and Kolmogorov Max Flow/Min Cut graph. Segmentation is used to increase the accuracy of identification especially for image with the complex background.

6. Leaf Matching using B-Spline Curve (Hikmawati, I. N , 2010). This research proposed an approach that incorporates deformable mapping (deformation energy) and geometric characteristics (strain difference) of spline curve to measure leaf similarity. This method was implemented on leaf image (Figure 3). Experiment result showed that same species leaves have low total cost as compared to different species leaves.

Figure 12. House plant image segmentation using Boycov Method and Kolmogorov Max Flow/Min Cut graph

Figure 13. Leaf matching using B-Spline curve



7. Medicinal Plant Identification Based on Morphology, Shape and Texture Using Probabilistic Neural Network (Nurfadillah, E., 2011). This research proposed a new system for Indonesian medicinal plant identification. The experiments result show that combination of plant features can improve the accuracy of plant identification until 74.67%.
8. Classifier Combination for Leaf Identification (Heriningtyas, P., 2011). This research compares a several classifier algorithms for plant leaf classification. We compared three classifiers i.e. Probabilistic Neural Network, Radial Basis Function Neural Network and Support Vector Machine. The Experiments result show that classifier Probabilistic Neural Network is superior to classify plan leaf image for identification.

II.5. Search Engine

Some plants for medicinal use have been documented in the form of books, reports and papers by researchers, students, and practitioners. With increasing number of research on medicinal plants, the number of books, reports, and papers will be increasing. This will make searching

information is not easy. So it is need to store and retrieve written information using computer. Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).

The field of information retrieval also covers supporting users in browsing or filtering document collections or further processing a set of retrieved documents. Given a set of documents, clustering is the task of coming up with a good grouping of the documents based on their contents. It is similar to arranging books on a bookshelf according to their topic. Given a set of topics, standing information needs, or other categories (such as suitability of texts for different age groups), classification is the task of deciding which class(es), if any, each of a set of documents belongs to. It is often approached by first manually classifying some documents and then hoping to be able to classify new documents automatically.

We have been conducted some research in information retrieval, especially for document indexing, classification, clustering, and question answering system (Paskianti, K, 2011; Ramadhina, A., 2011; Indriyani, W., 2011, Herawan, Y., 2011, Pramujadi, A, 2009).. In term of computer vision technology for plant identification, our research in information retrieval is very important to support identification and to provide informations. Figure 1 show information retrieval for medicinal plant document.

Figure 14. Information retrieval system for medicinal plant

III. Conclusion

Computer vision (image understanding) is a discipline that studies how to reconstruct, interpret and understand a 3D scene from its 2D images in term of the properties of the structures present in the scene. We have extensively conducted some research to develop computer vision for plant identification based on image especially for Indonesian medicinal plants and house plant. In the experiment, we used three features for plant identification i.e. morphology, shape and texture. Also, we explored several classifier to classify the plant. The classifier that we used, i.e., Probabilistic Neural Network, Radial Basis Function Neural Network, and Support Vector Machine. To help user finding information easily about the plants, we also conducted research to develop search engine system based on article and research documents. The experiments show that our systems are very promising. In the future we need to explore more extensively mathematical model as function to model the plant in order to improve the accuracy in plant identification and classification.

REFERENCES

Annisa. 2009. Morphology and Texture Feature Extraction For Leaf Image retrieval. Skripsi. Department of Computer Science, Faculty Mathematics and Natural Sciences. Bogor Agricultural University.

Herawan, Y. 2011. Feature Extraction of Medicinal Plants Using Chi-Square with Naïve Bayes Classifier. Skripsi. Department of Computer Science, Faculty Mathematics and Natural Sciences. Bogor Agricultural University. Unpublished

Heriningtyas. 2011. A Classifier Combination for Plant Leaf Identification. Skripsi. Department of Computer Science, Faculty Mathematics and Natural Sciences. Bogor Agricultural University. Unpublished

Hikmawati, I., N., 2010. Leaf Matching Using B-Spline Curve.. Skripsi. Department of Computer Science, Faculty Mathematics and Natural Sciences. Bogor Agricultural University.

Indriyani, W., 2011. Pseudo-Relevance Feedback for Information Retrieval using Sentence