

Classification Of Plants And Weeds In Multi-Leaf Image With Support Vector Machine Based On Leaves Shape And Texture Features

Etriana Meirista¹, Dessy Riski Suryani⁴

Mathematics Education Departement
Universitas Musamus
Merauke, Indonesia

¹etrianam@unmus.ac.id, ⁴drsuryani7@gmail.com

Imam Mukhlash², Budi Setiyono³

Mathematics Departement
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

²imamm@matematika.its.ac.id, ³budi@matematika.its.ac.id

Evy Nurvitasari⁵

Chemistry Education Departement
Universitas Musamus
Merauke, Indonesia

⁵evy_fkip@unmus.ac.id

Abstract—We often find plants that have similarities in terms of shape and texture. In agriculture if we want to plant a plant species, other plants will be called weeds as they can inhibit plant growth. We can easily classify plants and weeds using image processing. Watermelon plants were the objects of this research, so other plants besides watermelons will be considered weeds. The recognition of plants based on the similarity of the plant leaves used digital imagery which was divided into three stages. At the first stage, preprocessing was done by cropping the image, resizing the image, separating the background and foreground, and doing edge detection segmentation using the Canny operator. The second stage was feature extraction to retrieve important information for leaf recognition. The features used were features of shape and texture. Then we classified the leaves as leafy plants and weeds using the algorithm of Support Vector Machine (SVM). The SVM method is proven to have good accuracy for classifying plants and weeds based on the shape and texture features in a multi-leaf image using quadratic kernel. The average accuracy is 73.95%.

Keywords—Classification of plants and weeds, Leaves shape and texture features, Multi-leaf image.

I. INTRODUCTION

We often find plants that have similarities in terms of shape and texture. In agriculture, if we want to plant, other plants will be called weeds as they can inhibit plant growth. Weeds are undesired plants in the cultivation. The object of this research was watermelon plants, thus weeds are considered to be all plants that grow in addition to watermelon plants, for example, melons that grow on the sidelines of watermelon monoculture plants can be considered as weeds. There are several types of plants known as main weeds, such as grass, nut grass, and reed.

There are several parts of plants such as stems, flowers, fruits, leaves and roots that can be used in classifying plants. The leaves have two-dimensional shape so they have a simple structure [1]. Leaves are also easily found and collected in each season. Therefore in this research, the part of the plant used as a benchmark in the classification of plants is the leaves. Classification of plants based on leaves using conventional methods takes a long time. Therefore, it is very important to develop accurate, fast, and efficient systems in identifying various plant species and classifying them. We can use leaves as research objects to identify plants based on texture features, shape features, color features, venous features and leaf edges [2][3]. The most used image of the leaves is green or brown. Therefore, color features cannot be used in leaf image queries so that the features used to classify watermelon plants are only based on the leaf shape and texture features.

Many methods can be used for classification of leaf-based plants such as Linear Discriminant Classification (LDC), moving center hypersphere (MCH), Particle Swarm Optimization (PSO), probabilistic neural network (PNN), Support Vector Machine (SVM), and others. Among these methods, the SVM classifier is proven to be a good method for solving prediction, pattern recognition and classification problems. Like the research conducted by Zhiliang Kang and Peng Huang, they used 400 samples to examine leaf classification algorithms based on images. The results obtained were that BP Neural Network had an accuracy of 87.22%, PNN had an accuracy of 88.95% and SVM had an accuracy of 95.15% in classifying leaves [4]. The Support Vector Machine method is proven to classify single leaf images well with an accuracy rate of 91.3% [5]. The SVM method for classification of plant and tree leaves is by

optimizing the selection of feature subsets using Genetic Algorithm (GA). The features used are leaf area, standard deviation, boundary, number of ripples and number of ripple pixels. The proposed method is proved to have a higher level of accuracy than K-Nearest Neighbors (KNN) [6].

SVM is the largest classification method based on margins (edge lines) in accordance with the vector space. The aim is to find a decision boundary that has a maximum width of two support vectors between two categories of training data. The SVM method finds the best dividing limit by finding the maximum distance from all support vectors. Measurement calculation errors may occur, but that is not a problem because it will not trigger categorical errors caused by large shape of margins resulting in high classification accuracy decisions. Based on the description above, in this research SVM is used to classify plants and weeds in an image which has more than one leaf.

The steps taken in this research are segmentation of edge detection and extraction of shape and texture features. The shape features to be extracted are the basic features of geometry and digital morphological features which include diameter, physiological length, physiological width, leaf area, perimeter, ratio aspect, form factor, rectangularity, narrow factor, perimeter ratio of leaves, perimeter ratio of physiological length and physiological width, while the texture feature will be extracted using Gray Level Co-occurrence Matrices (GLCM).

II. RESEARCH METHOD

The object used was leaves from watermelon plants. The type of image used in this research was single-leaf image and multi-leaf image. Single-leaf image was used as training data which would later function as a classifier, while multi-leaf image acted as testing data.

A. Image Preprocessing

The image used was the image of a green leaf and which has a white background. Leaves on watermelon and weed plants have the same color so that the color information from leaf images was ignored. The initial stage in image preprocessing was cropping and resizing image. After cropping and resizing the image, there was a shadow on white paper so that the background and foreground separation stage was needed. The aim was to keep the foreground colored Red Green Blue (RGB) and the background colored white [7][8].

This separation was done by changing the image to Hue Saturation Value (HSV). The intensity component used was only Hue and Saturation. Then the image obtained was used as a binary image. Binary image is an image that has gone through a process of separating pixels based on gray degrees with a threshold of 0.5. The next process was to negate and make a frame on the image for making the background colored black and foreground colored and for eliminating noise. Then the image was merged then the opening morphology was done with radius 1 to eliminate unwanted parts (stems, flowers, and fruit). Furthermore, the segmentation of edge detection process was carried out using the Canny operator.



Fig. 1. Watermelon's Single-Leaf Image



Fig. 2. Weed's Single-leaf Image

B. Features Extraction

Leaf image that had been preprocessed was input data on feature extraction to facilitate the extraction of shape and texture features. The shape feature taken was a digital morphology feature (Table II) using information from 6 basic geometry features as shown in Table I and Fig. 3. Digital morphology features used were ratio aspect (slimness), form factor (roundness), rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of physiological length and width.

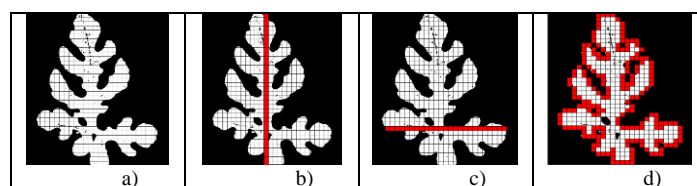


Fig. 3. Form Extraction Process. a) Area, b) Physiological Length, c) Physiological Width, and d) Perimeter

TABLE I. SIX BASIC GEOMETRY FEATURES

Name of Feature	Explanation
Diameter	A scalar that determines the diameter of a circle with the same area as the region is calculated by the formula: $\sqrt{\frac{4 \text{ Area}}{\pi}}$
Physiological Length (P_l)	the distance between the two terminals of the main vein of the leaf
Physiological	A line drawing through the two main vein

Name of Feature	Explanation
Width (L_f)	terminals, which can plot infinity orthogonal lines to the venous line. The number of intersecting pairs between these lines and leaf margins is also infinite. The longest distance between points of intersection is defined by physiological width.
Leaf Area (A)	The value of the Leaf Area is easily evaluated, that is, only by calculating the number of pixels of the binary which is worth 1 in the image of the smooth margin leaf.
Perimeter (P)	Calculation of the estimation of the number of pixels containing leaf margins.

TABLE II. DIGITAL MORPHOLOGY FEATURES

Name of feature	Feature Calculation
Ratio Aspect / Slimness (RA)	$RA = \frac{P_f}{L_f}$
Form Factor / Roundness (FF)	$FF = \frac{4\pi A}{P^2}$
Rectangularity (R)	$R = \frac{P_f L_f}{A}$
Narrow Factor (NF)	$NF = \frac{D}{P_f}$
Perimeter Ratio of Diameter (RPD)	$RPD = \frac{P}{D}$
Perimeter Ratio of Physiological Length and Physiological Width (RPPLF)	$RPPLF = \frac{P}{P_f + L}$

Texture features were extracted using Gray Level Co-occurrence Matrices (GLCM). GLCM is very useful for obtaining valuable information about the relative position of neighboring pixels in the image [9]. The Co-occurrence feature noticed in Gray Level Co-occurrence Matrices was shown in Table III.

TABLE III. CO-OCCURRENCE FEATURE

Name of Feature	Feature Calculation
Contrast	$\sum_{i,j} i - j ^2 f(i, j)$
Correlation	$\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) f(i, j)}{\sigma_1 \sigma_2}$
Energy	$\sum_{i,j} f(i, j)^2$
Homogeneity	$\sum_{i,j} \frac{f(i, j)}{1 + i - j }$

C. Classification With SVM Method

The results of feature extraction can be used to classify the plants into watermelon plants or weed using the Support

Vector Machine (SVM) method. SVM's success lies in matching kernel design and parameter selection. Therefore, the kernel used is a quadratic kernel because it is suitable and has high accuracy.

The diagram of system process was used to describe the system work steps that will be created.

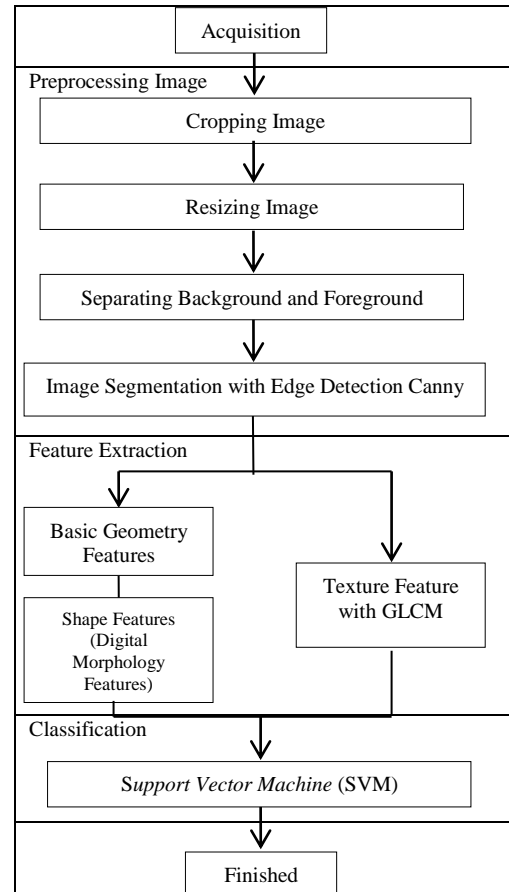


Fig. 4. Diagram of Classification

The classification process using the SVM method is shown in Fig. 5. The input image used was a single image for data training and multi-leaf image for data testing with colored foreground based on the color of the leaf and white background using jpg format. The image for training used was 189 images consisting of 156 images of watermelon plant leaves and 33 weeds aged 15 and 40 days. Then image preprocessing was done by cropping and resizing the image. After that the background separation was done by changing the Red Green Blue (RGB) image to the Hue Saturation Value (HSV) image, then taking the intensity component H and S. Furthermore, segmentation of edge detection was carried out using the Canny operator. In this research, there are two features used, which are the shape and texture features. For shape features, the extraction was performed using digital morphology features, which previously must know information from 6 basic geometry features. The texture feature was extracted using Gray Level Co-occurrence Matrices (GLCM) which was determined by looking at the contrast, correlation, energy and homogeneity (see Table III).

Data training was used to build models and data testing was used to test the models that have been built. The next stage was to do the classification process using the SVM method.

III. RESULT AND DISCUSSION

We have conducted an experiment to show the effectiveness of the method proposed. As previously explained, the initial steps taken in this research were cropping and resizing images, separating the background and foreground and eliminating unwanted parts, and segmentation using edge detection of Canny operator. The dataset used was watermelon leaf and green weed which has a white background and it is assumed that the image used was without occlusion and had no noise.

After cropping and resizing the image, there were watermelon leaves, flowers, fruit, stems and shadows on white paper. Because the object used in this research was the leaves, the leaves were the foreground while the watermelon flowers, fruit, watermelon stems and shadows were considered to be the background that must be changed to white. This process is called separating the background and foreground. The digital image taken was the image of Red Green Blue (RGB). The initial stage which separated the background and foreground was to change the RGB image to Hue Saturation Value (HSV) image and then took the color components H and S from the HSV image to convert it to black and white images. The next was to negate the black and white image and combine the two images to produce an image with a black background and white foreground (Fig. 5).

Image segmentation using edge detection is needed to find information on texture feature values. The operator used in this edge detection was the Canny operator obtained from the image of multiplication result between the original image converted into a grayscale image and the image from background separation. Image from edge detection with canny operator (Fig. 5. f).

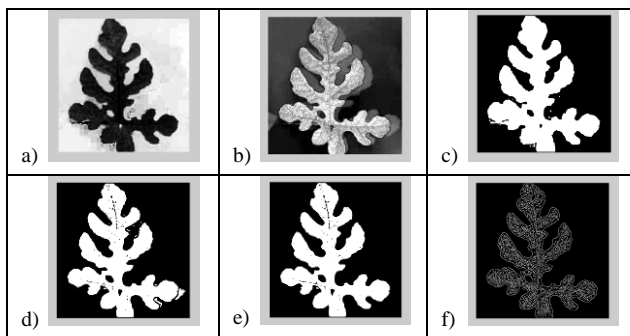


Fig. 5. The image after resizing process (a) Taking Hue Color Components, (b) Taking Saturation Color Components, (c) Changing Hue Color Components to Negated Black and White, (d) Changing the Saturation Color Components to Negated Black and White Images, (e) Merging Black and White Images from Hue and Saturation Color Components, (f) Segmentation with Canny Operator Edge Detection Method.

TABLE IV. SHAPE FEATURES OF WATERMELON LEAVES

No.	RA	FF	R	NF	RPD	RPPLF
1	0.391	16.242	4.706	0.077	0.074	0.077
2	0.454	12.542	4.135	0.065	0.062	0.065
3	0.563	9.8781	3.843	0.079	0.082	0.079
...
154	0.526	11.278	4.036	0.103	0.104	0.103
155	0.297	20.148	4.791	0.047	0.046	0.047
156	0.404	27.934	8.039	0.108	0.106	0.108

In the feature extraction process, the form of input data used was the image of the leaf that had separated the background and foreground. This leaf image facilitated the extraction process of shape feature. Leaf image with green leaves and white background that had been gray scale was then multiplied by the results of the background and foreground separation which was the input data in the texture extraction process.

TABLE V. SHAPE FEATURES OF WEED LEAVES

No.	RA	FF	R	NF	RPD	RPPLF
1	0.74	9.819	3.85	0.076	0.075	0.076
2	0.667	10.82	3.983	0.06	0.057	0.06
3	0.664	9.546	3.647	0.063	0.061	0.063
...
31	0.699	9.383	3.416	0.038	0.038	0.038
32	0.709	9.463	3.76	0.045	0.044	0.045
33	0.679	8.48	3.381	0.048	0.049	0.048

Then it was preceded to the classification using SVM for data training so as to produce a classifier. Then the classification was carried out using the SVM method. The same process was performed the same as classifying data training using a single image. Data testing was first preprocessed then the feature information was taken. After the feature information was obtained, the feature was tested one by one with the classifier that had been obtained from the above data training. Feature values that share the similarities were classified using the SVM method. This process is shown in the diagram block of system design.

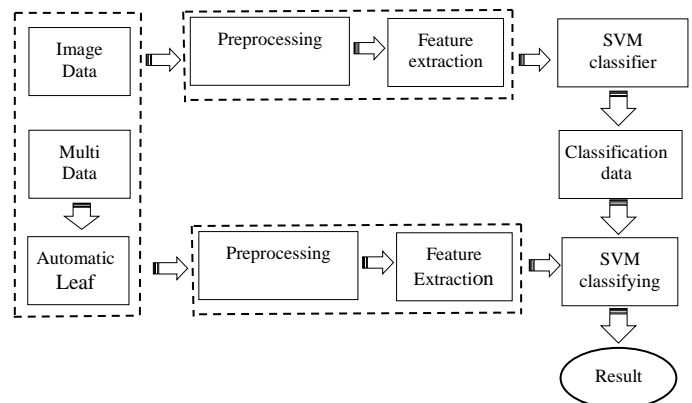


Fig. 6. System Design

Fig. 7 presents the original image, multi leaf image used in this research. It is consist of watermelon and weed leaves image. Fig. 8 shows the results of leaf image detection by making a rectangle on the detected leaf. As a variation, we display the image of the leaf that is detected using a different color presented in Fig. 10. Each different color shows different leaves. If an error occurs in detecting the leaf image due to occlusion (stacking), then the detected leaf is not included in the classification. Multi-leaf image classification process with the SVM method is the same as the training process.



Fig. 7. Multi-Leaf

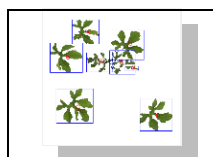


Fig. 8. Single Leaf Detected from Multi-Leaf Image

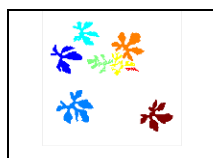


Fig. 9. Single Leaf Detected and Given Different Colors

Multi-leaf image classification was done by detecting the watermelon and weed leaves further classified by observing the classification of a single image from the training results. The initial step in the multi-leaf image classification process was detecting watermelon and weed leaves, and after that the testing process was carried out one by one using 189 single image data that had been classified during the training. Leaf images in Fig. 9 show leaves that are classified by giving different colors to each detected leaf and giving S code if the leaf is identified as watermelon leaves and G code if the leaf is identified as weeds. The classification of the results of the data testing presented shows 3 watermelon leaves and 5 weeds. The results of the data test displayed have 100% accuracy. However, not all data testing was detected with a 100% accuracy rate as shown in Fig. 10

On test data 1, there is an identification of objects that should be 2 and 3 leaves but are identified as 1 leaf. But because the two leaves are watermelon leaves, the leaves are defined as watermelon leaves. Likewise, for weed leaves, if 2 weed leaves are identified with 1 leaf then the leaf is defined as weed leaves. If there is an introduction to the object of watermelon leaves and weeds identified as 1 leaf, then the object is not counted in the results of the classification. The results of the test 2 data classification consist of 3 watermelon leaves and 5 weed leaves. There is an introduction of objects that should be 2 and 3 leaves but are identified as 1 leaf. But

because the two leaves are watermelon leaves, the leaves are defined as watermelon leaves. Likewise for weed leaves, if 2 weed leaves are identified with 1 leaf then the leaf is defined as weed leaves. If there is an introduction to the object of watermelon leaves and weeds known as 1 leaf, then the object is not counted in the classification results. The test data is then interpreted as the previous interpretation.

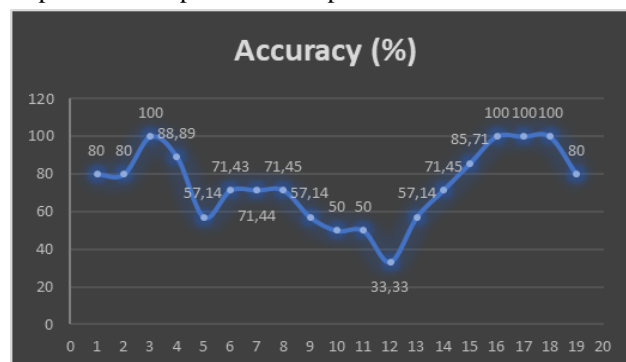


Fig. 10. Accuracy of Multi-Leaf Image

The results of the classification with the SVM method presented in Fig. 10 states that the SVM method is proven to have good accuracy for classifying plants and weeds based on shape and texture features in multi-leaf image using quadratic kernel. This is indicated by the average accuracy of 73.95%.

IV. CONCLUSION

Based on training data used were 189 pictures consisting of 156 watermelons leaves images and 33 weeds images. By using 19 multi-leaf images for data testing, the classification results between plants and weeds using SVM looking at the shape and texture features show good accuracy of 73.95%.

REFERENCES

- [1] X.F. Wang, D.S. Huang, J.X. Du, H. Xu, and L. Heutte, "Classification of plant leaf images with complicated background," *Appl. Math. Comput.*, vol. 205, no. 2, pp. 916–926, 2008.
- [2] C.Y. Gwo, C.H. Wei, and Y. Li, "Rotary matching of edge features for leaf recognition," *Comput. Electron. Agric.*, vol. 91, pp. 124–134, 2013.
- [3] P. Pallavi and V. Devi, "Leaf Recognition Based on Feature Extraction and Zernike Moments," *Int. J. Innov. Res. Comput. Commun. Eng. (An ISO Certif. Organ.)*, vol. 32972, no. 2, pp. 67–73, 2007.
- [4] Z. Kang and P. Huang, "Research on Image Classification Algorithm Based on the Image," vol. 83, no. Snce, pp. 835–839, 2018.
- [5] E. Meirista, I. Mukhlash, and B. Setiyono, "Watermelon Plant Classification Based on Shape and Texture Feature Leaf Using Support Vector Machine (Svm)," no. May, pp. 17–19, 2015.
- [6] V. Narayan and G. Subbarayan, "An Optimal Feature Subset Selection Using GA for Leaf Classification," *Int. Arab J. Inf. Technol.*, vol. 11, no. 5, pp. 447–451, 2014.
- [7] K. ElDahshan, M. Youssef, E. Masameer, and M. Hassan, "Comparison of Segmentation Framework on Digital Microscope Images for Acute Lymphoblastic Leukemia Diagnosis Using RGB and HSV Color Spaces," *J. Biomed. Eng. Med. Imaging*, vol. 2, no. 2, 2015.
- [8] T.A. Kulkarni and D.S. Bhosale, "A Robust Morphological Analysis of Normal and Abnormal Leukemic Cells Populations in Acute Lymphoblastic Leukemia," vol. 18, no. 8, pp. 978–93, 2014.
- [9] A. Kadir, L.E. Nugroho, A. Susanto, and P.I. Santosa, "Neural Network Application on Foliage Plant Identification," *Int. J. Comput. Appl.*, vol. 29, no. 9, pp. 975–8887, 2011.

