M – 10

WATERMELON PLANT CLASSIFICATION BASED ON SHAPE AND TEXTURE FEATURE LEAF USING SUPPORT VECTOR MACHINE (SVM)

[1]Etriana Meirista, [2]Imam Mukhlash, [3]Budi Setiyono
[1]Master Student of Departement of Mathematics, [2,3]Departement of Mathematics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
etrianameirista47@gmail.com

Abstract

Nowadays, some efforts are used to increase results of agriculture production. One of those is utilizing herbisides to exterminate the weeds. However, there are some of the weeds having resemblance with the plant, with the result that we need to classify the plant and the weeds before utilizing herbisides as an extermination weeds. In this paper, we use watermelon plant classification as case study. The recognition of the plant owned by the similarity of leaves of these plants are divided into three phases. At the first phase we perform preprocessing to convert the RGB image into a grayscale images. Further, the grayscale images are changed into segmentation of edge detection using Canny operator. In the second, we use feature extraction to retrieve important informations for the recognition of those leaves. The last phase we classify that leaves as watermelon plants or weeds using Support Vector Machine (SVM) algorithm. The results of early trials indicate that this method has an accuracy of 91,3%.
Keywords : image, leaf, edge detection, feature extraction, and plant classification esults of early trials

INTRODUCTION

The farmers, botanist, researchers have tried some efforts to increase agricultural production. One of the ways is using herbicides to exterminate weeds. However, there are some weeds which have similarities with the plants, so that we need to identify the contours of the plants and the weeds, and also to classify them before exterminating with herbicides.

Plants are basically identified according to their morphological features such as number of ovaries in the fruit or number of stamens in the flower. A number of Manual and computer-aided keys for plant identification using morphological features are available in the literature. Identifying plants using such keys is a very time consuming task and has been carried out only by trained botanists. There are several other drawbacks in identifying plants using these features such as the unavailability of required morphological information and use of botanical terms that only experts can understand. Fortunately, in addition to the structures of reproductive organs, shape, size, texture and colour of the leaves also play an important role in plant identification (Bama, et al., 2011). Wang et al (2008) stated that "the leaf have two shapes dimensional so the structure to be simple". Therefore, based on the simple structure, leaves are used as a standards in classification and are also easily found in every season. Classification's plants based on the leaves by conventional way require a long time. Therefore, it is very important to develop accurate, fast and efficient sytem in identifying the various species of the plants and classifications.

Several researchers have used texture, shape, color, edge, and leaves veins features as the research object such as Gwo et al (2013), Devi and Pallavi (2014). Further, Gwo et al used the edge of the leaves to introduce the plants. Bayes theorem is used for matching rotary. Devi and Pallavi gave a new framework for recognizing and identifying plants based on shapes, veins, color and texture features. The shape and leaves texture features are only used in classification of plant because the leaves of watermelon and weeds have no different color.

There are lots of methods that can be used for classification plants based on the leaves such as Linier Discriminant Classification (LDC), Moving Center Hypershere (MCH), Particle Swarm Optimation (PSO), Probabilistic Neural Network (PNN), Support Vector Machine (SVM), etc. Among of these methods, SVM classifier is proved as a good method to solve problem of prediction, pattern recognition and classification (Qian, et al. 2010). Narayan and Subbarayan (2014) introduced Support Vector Machine (SVM) method for classification of plant leaves and tree to optimize the subset feature of selection using a Genetic Algorithm (GA). Features used are the leaf area, standard deviation, boundary, the amount of ripple and ripple number of pixels. The proposed method is shown to have a higher degree of accuracy than the K-Nearest Neighbors (KNN).

SVM is a classification method based on the margin (edge line), the largest space-based vector. The aim of this method is to find a decision boundary that has a maximum width of two support vector between the two categories in the training data. SVM find the best separation barrier by finding the maximum distance of all the support vector. The measurement errors can occur, but it is not a problem because it will not cause confusion categorize caused by the shape of a large margin resulting classification decisions with accuration of Cleaner. Based on the previous informations, this research of the SVM is used to classify watermelon plants and weeds.

RESEARCH METHOD

Datasets are taken in digital image formats such as JPEG, and BMP. Meanwhile, the leaves of the plant watermelon are object of this research. Image types of this research are single image with a variety of conditions. The criteria of watermelon leaves image is an image of young leaves, old, clipped, disease, glare due to the illumination to sunlight and has a integral contour (See figure 1).

Picture 1. Leaf dataset with variety of conditions

1. The Image Preprocessing

Image segmentation is the process of partitioning a digital image into multiple regions or set of pixels. The segmentation is based on measurements taken from the image and might be grey level, colour, texture, depth or motion. The result of image segmentation is a set of segments that collectively cover the entire image. All the pixels in region are similar with respect to some characteristic or computed property, such as color, intensity, or texture (Gonzalez, 2008; Ramadevi, 2010). The image used is the image of the colored leaves of green and has a white background. Leaves on watermelon plants and weeds have the same color. Therefore, the color information of the image leaves is ignored so the next phase is to change the image of the Red Green Blue (RGB) into grayscale using the following equation:

$$Gray = 0.299{*}R + 0.587{*}G + 0.114{*}B \quad (1)$$

Further, the grayscale image is segmented by using the edge detection with Canny operation. Binary image is an image that has been through the process of separation of pixels based on the degree of gray-owned. The formation of a binary image requires grayscale boundary value to be used as a standart value. Pixels with gray level are greater than the value of the boundary that is given a value of 1 and gray pixels smaller than the limit value will be assigned the value 0.

2. Features Extraction

The image of the leaves that have been through a preprocessing input data on features extraction because it facilitates the process of shape and texture features extraction. Feature shape is a digital morphological features as shown in Table 2, which uses information from the basic features of the geometry as shown in Table 1 (Wu, et al. 2007).

Table 1. The Geometrical Features

No.	Features Name	Description
1.	Physiological Length (P_f)	The distance between the two termimals of the main vein leaf.
2.	Physiological Width (L_f)	Drawing a line passing through the two terminals of the main vein, which can plot infinite lines orthogonal to the line vein. The number intersection pairs between those lines and the leaf margin is also infinite. The longest distance between points of those intersection pairs is difined at the physiological widht.
3.	Leaf Area (A)	The value of leaf area is easy to evaluate, just counting the number of pixels of binary value 1 on smoothed leaf image.
4.	Leaf Perimeter (P)	Leaf perimeter is calculated by counting the number of pixels consisting leaf margin..

The overview of the length and width of the physiological relationship are shown in Figure 1.

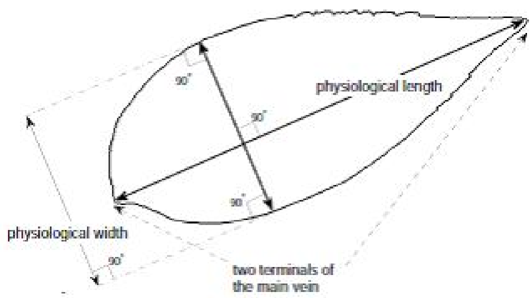

Figure 1. The relationship between physiological lenght and physiological width

Table 2. The Digital Morphology Feature

No.	Feature Name	Calculation
1.	Aspect Ratio / Slimness (A_r)	$A_r = \frac{P_f}{L_f}$
2.	Form Factor / Roundness (F)	$F = \frac{4\pi A}{P^2}$
3.	Rectangularity ($Rect$)	$Rect = \frac{P_f L_f}{A}$
4.	Perimeter Ratio of Physiological Lenght and Physiological Widht (P_{PLF})	$P_{PLF} = \frac{P}{P_f + L_f}$

Extracted texture features uses Gray Level Co-occurrence Matrices (GLCM). GLCM is very useful to obtain important information about the relative position of neighboring pixels in an image (Kadir et al, 2011). Co-occurrence features are noted on Gray Level Co-occurrence Matrices shown in Table 3.

Table 3. The Co-ocucurrence Feature

No.	Features Name	Features Calculation
1.	Contrast	Contrast = $\sum_{i,j} \lvert i-j \rvert^2 f(i,j)$
2.	Correlation	Correlation = $\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)f(i,j)}{\sigma_i \sigma_2}$
3.	Energy	Energy = $\sum_{i,j} f(i,j)^2$
4.	Homogenecity	Homogenecity = $\sum_{i,j} \frac{f(i,j)}{1+\lvert i-j \rvert}$

3. Classification with SVM Method

The extraction of features is used to classify plants into the watermelon plant species or weeds by using Support Vector Machine (SVM). The success of SVM depends on the design of kernel and suitable parameters selection. Therefore, quadratic kernel is used as kernel because it is only suitable and a high accuracy for this proposed classification framework.

The process of SVM classifier is shown in Figure 2. Input image is a single image with foreground green and white background that use the format JPEG, BMP. The image is used as many as 56 images consisting of 36 images watermelon plant leaf and 20 weeds leaf with variety conditions. Then perform preprocessing the image by changing the RGB image into a grayscale image, then perform segmentation using edge detection Canny operator. In this proposed classification framework, there are two features. They are shape feature and texture feature. For shape feature extraction is done using morphological features digital (shown in Table 2), which previously had to be informed of the basic features geometry (shown in Table 1). While the extracted texture features using Gray Level Co-occurrence Matrices (GLCM) were determined by contrast, correlation, energy and homogenecity (see Table 3). The data is then divided into training data and testing data. The training data is used to build the model and the testing data is used to test the model that has been built. Input image is used as training data 1-33 and 34-56 images are used as the data 23 testing. The next stage is the process of classification using SVM method.

Diagram of the system of process or workin steps is shown this below.

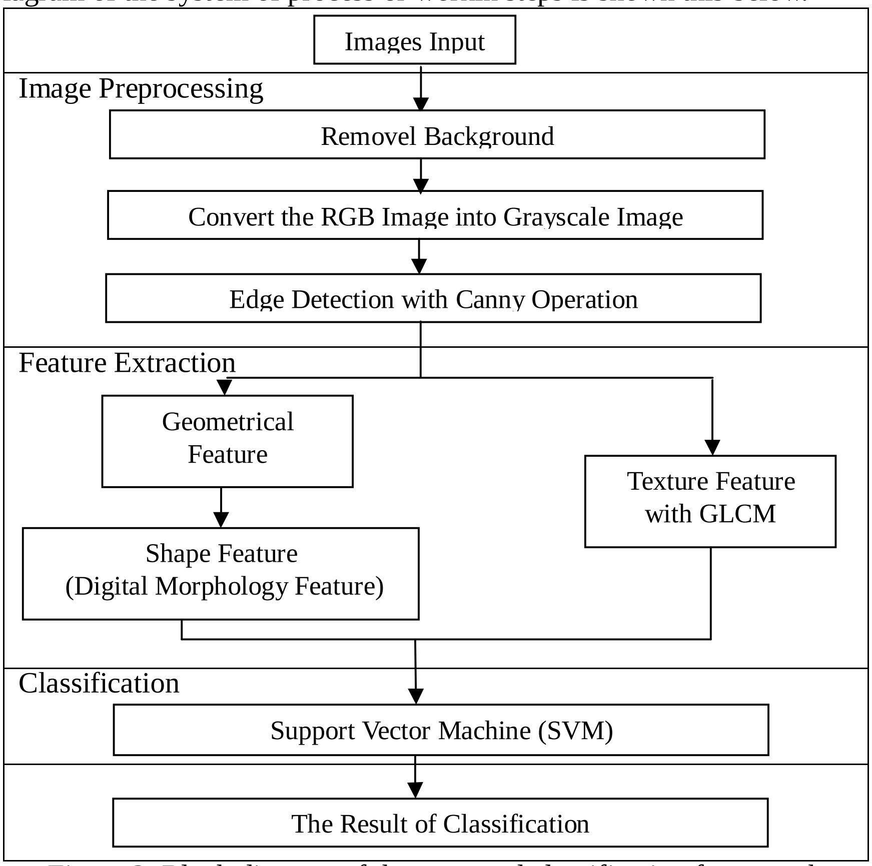

Figure 2. Block diagram of the proposed classification framework

RESULT AND DISCUSSION

We have conducted experiments to demonstrate the effectiveness of the proposed method. As mentioned previously, the first step in this research is the segmentation using edge detection with Canny opertaor. The dataset is the watermelon leaves and weeds that have a white background and preprocessed with segmentation (Figure 3). The image in this research is assumed that the image has no occlusion and noise.

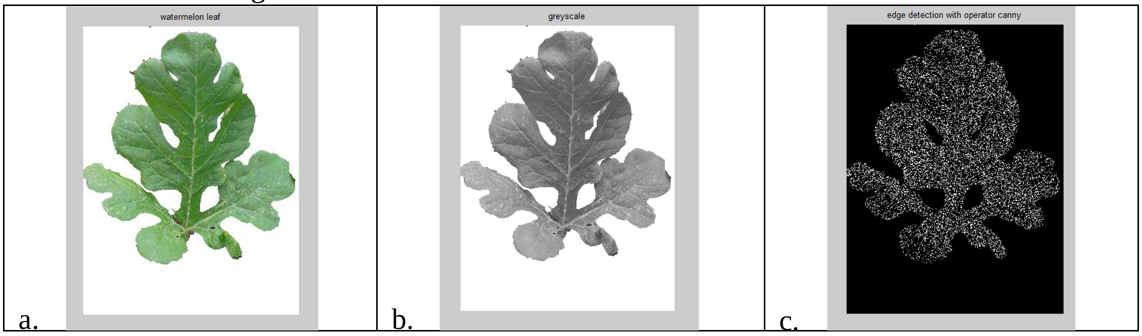

Figure 3. (a) original image, (b) a grayscale image, and (c) edge detection image by Canny operator.

In the feature extraction process, input data is the image of the leaves that have been through the process of segmentation. The segmented image of leaves is to make easy in the process of feature extraction form. Image leaves with green leaves and white background is the data input on texture extraction process.

The initial step of this method is to determine the SVM classifier. Information from leaf image of watermelon plants and weeds that have preprocessed (see Figure 3) is used to extract features. There are 33 training data, each of represented data with 12 features (such as in Table 4) are then formed by the method of SVM classifier.

Table 4. The Feature Watermelon Plants and Weeds

No.	Contrast		Correlation		Energy		Homogenecity		Ar	F	Rec	Pplt	Class
	Min	max	Min	Max	min	max	Min	max					
1	0,035	0,004	0,98	0,998	0,907	0,91	0,994	0,998	0,123	0,001	94,02	3,173	Weeds
2	0,054	0,036	0,99	0,994	0,692	0,694	0,98	0,984	0,223	0,004	201	1,494	Weeds
3	0,042	0,021	0,99	0,995	0,803	0,804	0,991	0,993	0,134	0,004	125,1	1,633	Weeds
4	0,043	0,028	0,99	0,995	0,731	0,732	0,987	0,989	0,135	1E-03	8,889	12,16	Weeds
5	0,054	0,053	0,99	0,987	0,648	0,648	0,979	0,979	0,134	0,002	14,35	7,327	Weeds
6	0,118	0,118	0,98	0,982	0,352	0,351	0,947	0,946	0,922	0,004	552,3	1,232	Weeds
7	0,049	0,05	0,99	0,993	0,429	0,428	0,977	0,976	0,422	0,003	40,76	4,503	Weeds
...	...	...	...	...	...	...	...	...	...	...	...	...	...
31	0,032	0,032	0,97	0,972	0,864	0,864	0,991	0,991	0,631	0,001	95,47	4,986	Watermelon
32	0,014	0,013	0,95	0,955	0,937	0,937	0,995	0,995	0,914	0,006	55,85	2,948	Watermelon
33	0,012	0,012	0,93	0,936	0,97	0,97	0,996	0,996	0,881	0,008	29,98	3,62	Watermelon

Further, there are 23 classifications of SVM for testing process. First testing data is preprocessed, it is then taken its information. The feature information is tested one by one with a classifier that has been obtained from the training data. Similar feature values are classified using SVM method. The results of testing are processed with SVM method as shown in Table 5.

Tablel 5. The Results of Testing with SVM Method

No	Class	Prediction	Conclusion
34	Weeds	Weeds	True
35	Weeds	Watermelon	False
36	Weeds	Weeds	True
37	Weeds	Weeds	True
38	Weeds	Weeds	True
39	Weeds	Weeds	True
40	Weeds	Weeds	True
41	Weeds	Weeds	True
42	Weeds	Weeds	True
43	Weeds	Weeds	True
44	Watermelon	Watermelon	True
45	Watermelon	Watermelon	True

No	Class	Prediction	Conclusion
46	Watermelon	Watermelon	True
47	Watermelon	Watermelon	True
48	Watermelon	Watermelon	True
49	Watermelon	Watermelon	True
50	Watermelon	Weeds	False
51	Watermelon	Watermelon	True
52	Watermelon	Watermelon	True
53	Watermelon	Watermelon	True
54	Watermelon	Watermelon	True
55	Watermelon	Watermelon	True
56	Watermelon	Watermelon	True

Testing of the data as shown in Table 5, there is a prediction error in the input image of the 35^{th} and 50^{th}. The error rate is 8,7%, so the level of accuracy of the SVM method in this study is 91,3%

CONCLUSION AND SUGGESTION

In this research, we has been conducted segmentation edge detection with Canny operator to get shape features of digital images. Further, the texture features are extracted by using GLCM and GLCM is only obtained from the grayscale images. Training data are used as many as 33 images consisting of 23 images of watermelon plant leaf and 10 weeds images. By using 23 images of leaves randomly testing the data, the results of the classification watermelon plants and weeds using SVM based on the shape and texture features showed high accuracy. The accuracy is 91,3%. For further research, we can be conducted on plant classification is not a single image, so in one image artifacts consisting a few pictures of watermelon plants and weeds leaves, and then we can estimate the amount of herbicide to be used on the agricultural land.

REFERENCES

[1] Bama, B. S., Valli, S. M., Raju, S., and Kumar, V. A. (2011), “Content Based On Leaf Image Retrieval (CBLIR) using Shape, Color, and Texture Features”, *Indian Journal of Computer Science and Engineering (IJCSE)*, Vol. 2, No. 2.

[2] Devi, V and P, Pallavi. (2014), “Leaf Recognition Based on Feature Extraction and Zernike Moments”, *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 2.

[3] Gonzalez, R. C., dan Woods, R. E. (2008), *Digital Image Processing*, 3^{rd} edisi, Pearson Eduacation, Inc., New Jersey.

[4] Gwo, C. Y., Wei, C. h.,and Li, Y. (2013), Rotary Matching of Edge Feature for Leaf Recognition, *Computer and Electronics in Agriculture*, Vol. 91, hal. 124-134.

[5] Kadir, A., Nugroha, L.E., Susanto, A., and Santosa, P. I. (2011), “Neural Network Application on Foliage Plant Identification”, *International Journal of Computer Applications* ,Vol. 29, No.9, hal. 15 - 22.

[6] Kalyoncu, C. dan Toygar, Ö. (2014), “Geometric leaf classification”, *Computation Vision and Image Understanding*. http://dx.doi.org/10.1016/j.cviu.2014.11.001

[7] Qian, H., Mao, Y., Xiang, W., and Wang, Z. (2010), “Recognition of Human Activities using SVM Multi-class Classifier”, Pattern Recognition Letters 31 (2010) 100–111.

[8] Ramadevi, Y., Sridevi, T., Poornima, B., and Kalyani, B. (2010), “Segmentation on Object Recognition using Edge Detection Techniques”, *International Journal of Computer Science & Information Technology (IJCSIT)*, Vol. 2, No. 6.

[9] Wang, X. F., Huang, D. S., Du, J. X., Xu, H., and Heutte, L. (2008), “Classification of Plant Leaf Images with Complicated Background”, *Applied Mathematics and Computation*, Vol. 205, hal. 916-926.

[10] Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., and Xiang, Q.L. (2007), A Leaf Recognition Algorithm for Plant Calssification Using Probabilistic Neural Network.