Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Leaf classification with improved image feature based on the seven moment invariant

To cite this article: Febri Liantoni *et al* 2019 *J. Phys.: Conf. Ser.* **1175** 012034

IOP Conf. Series: Journal of Physics: Conf. Series **1175** (2019) 012034 doi:10.1088/1742-6596/1175/1/012034

Leaf classification with improved image feature based on the seven moment invariant

Febri Liantoni[1], Rifki Indra Perwira[2], Syahri Muharom[1], Riza Agung Firmansyah[1] and Akhmad Fahruzi[1]

[1]Institut Teknologi Adhi Tama Surabaya, Indonesia
[2]UPN "Veteran" Yogyakarta, Indonesia

[*]febri.liantoni@gmail.com

Abstract. Plants can be classified based on its leaves shape. The classification process is required well data extraction feature, so it needs fixing feature process at the pre-processing level. Combining median filter and image erosion is used for fixing the feature process. Whereas for feature extraction is used invariant moment method. In this research, it is used leaves classification based on leaves edge shape. K-Nearest Neighbor Method and Naïve Bayes Classifier are used for leaves classification process. Testing the result of leaves classification from an image which is on dataset has been built to get accuracy value about 84% using Naive Bayes classifier while using K-Nearest Neighbor the get accuracy value of about 84%.

1. Introduction

The science of plants is progressing rapidly, the field of knowledge that was previously only a branch of plant science alone now has become a field of science that stands alone. One of them is Plant Morphology which studies the shape and composition of the plant body. The leaf margin is used for reference leaf classification. Plants are used as a provider of oxygen for breathing, as food, fuel, medicine, cosmetics and more. The process of plant classification can be done by identifying the leaf shape image of the plant itself. How to take a leaf image of the plant, it can be done steps recognition of leaf pattern by recognizing the leaf structural characteristics such as the shape and texture of the leaf [1], [2].

The process of plant classification can be done by identifying the leaf shape image of the plant. In this way, leaf pattern recognition steps can be recognized by recognizing the structural characteristics of leaves such as the shape and texture of a leaf. The method for processing the input image with the utilization of digital image processing techniques is done to analyze the leaf structural characteristics.
Technological developments in image processing techniques are also growing rapidly. Various techniques are developed to facilitate human work, both as image processing, image analyst and image user for various purposes and purposes. Often the image used is not in ideal condition to be studied because of the many disturbances, it can be shadow, photo or blurred image, lack of clear appearance of the object so it can cause problems and affect the results and will affect the analysis and plan to be done, it required various processing techniques image to get the ideal image.

This digital image processing technique is done at the preprocessing stage of the image until it gets the shape of the edge and the structural features of each leaf. The method used in the extraction of this feature is the introduction of digital morphological features [3]. The feature extraction process used is the Moment Invariant method. After that performed feature extraction from the leaf image so obtained leaf structural information which then used as data grouping [4].

Automatic plant leaf recognition and classification system are very useful because it supports a fast classification of plants. Leaf is one of the unique plant characteristics and easily observable and representative enough so that it can be an object for the extraction of plant features. Extraction of the right features of the object greatly affects both the poor results of plant classification. Some research on the grouping of plant forms using an artificial neural network method [5], [6]. Artificial neural network methods are widely used because these methods are known to be substantially faster. However, the determination of the number of hidden layers used to affect the results, as well as required parameters an amount of epoch so it requires higher computation [5]. In the previous research, the authors conducted a research on the classification of mango leaves using K-Nearest Neighbor and Support Vector Machine method. In this research was presented Ant Colony Optimization method to improve the results of mango leaf image detection [7].

In this research, the preprocessing stage there will be feature improvement. Feature value based on seven moments invariant. For classification process using K-Nearest Neighbor method and Naïve Bayes classifier. The process of classification based on leaf margin features. Prior to the first classification stage, the pre-processed image stage and extraction of the leaf edge feature to obtain the appropriate input value for the leaf classification phase based on the leaf image.

2. Methods

2.1. Feature Invariant Moment Extraction

The feature extraction used is moment invariant. This process is performed to generate vector feature values from binary imagery. The feature used is seven-moment invariant which will generate seven values on the feature vector.

The process of recognizing an object in an image after the segmentation process often bumps on object positioning problems, object axis rotation, and scale changes of objects. The position of a shifted or rotating object or its size smaller or larger than can cause errors in the recognition or identification of an object.

In the use of the calculation of the value of 2 (two) dimensions of the sample moment of the image M x M of the continuous function f (x, y), (x, y=0, ... M-1) obtained Equation 1.

$$m_{pq} = \sum_{=0}^{x=M-1} \sum_{y=0}^{y=M-1} (x)^p \cdot (y)^p \cdot f(x,y) \quad (1)$$

Moments can represent an object in terms of area, position, orientation and other undefined parameters. By obtaining a certain amount of moment information, either the moment of the level to zero (m00) and to the one (m10) and (m01) or the central moment, and the moment at the level of ≥ 2 or the moment invariant of an object, the object can be identified even if it has a shift (translation), rotation (rotation) and scale changes.

From moment f (x, y) will be translated by the value (a, b) to obtain a new calculation like Equation 2.

$$\mu_{pq} = \sum_x \sum_y (x+a)^p \cdot (y+b)^q \cdot f(x,y) \quad (2)$$

From the main central moment is m_pq or μ_pq computed through substitution process to value a = -x and value b = -y it will get the calculation in Eq 3.

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p \cdot (y-\bar{y})^q f(x,y) \quad (3)$$

Where: $\bar{x} = \frac{m_{10}}{m_{00}}$ dan $\bar{y} = \frac{m_{01}}{m_{00}}$

When the normalization process then the scaling value used in the calculation changed, as shown in Equation 4.

$$\eta_{pq} = \mu_{pq}/\mu_{pq}{}^{\gamma} \quad (4)$$

For value: $\gamma = [(p+q)/2] + 1$

Then from the process then got the value of seven moments invariant with the calculation in Equation 5.

$$Hu_1 = \eta_{20} + \eta_{02}$$
$$Hu_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}{}^2$$
$$Hu_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
$$Hu_4 = (\eta_{30} + \eta_{12})^2 + (3\eta_{21} + \eta_{03})^2$$
$$Hu_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{21} + \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$
$$Hu_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$Hu_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (5)$$

2.2. Preprocessing

In this research the data used in the form of a leaf image with a white background. The dataset used is a green leaf image consisting of 5 types (5 classes). Image data used 100 image. Leaf image classification system using K-Nearest Neighbor method consists of several processes including pre-processing, feature extraction and classification stage. The sample data used as shown in Figure 1.

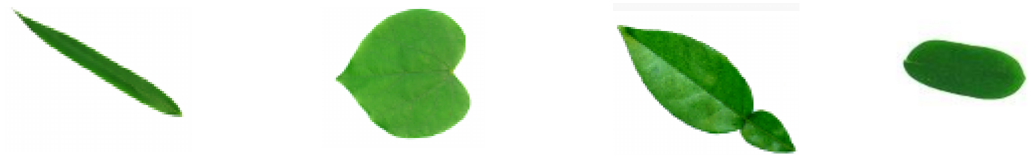

Figure 1. Sample leaf

Preprocessing is done in order to process input data so that it can be used for feature extraction process. There are several steps taken for preprocessing the image conversion in the RGB color space to the grayscale color space, opening the median filter, creating binary imagery, and fixing some image pixels using the image erosion method. From the preprocessing stage will be obtained the binary image that will be used for feature extraction. The steps performed on the pre-processing are shown in Figure 2.

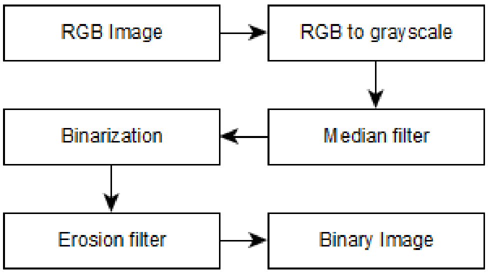

Figure 2. Preprocessing stages

The initial stage of preprocessing by converting an image into grayscale. This processing is done to convert an image pixel domain to an 8-bit grayscale. For such conversion is used as shown in Eq 6.

$$gray = \frac{red*299+green*587+blue*114}{1000} \quad (6)$$

2.3. Classification

This research used K-Nearest Neighbor (KNN) and Naïve Bayes Classifier method. KNN is a method to classify objects based on learning data closest to the object [8]. Learning data is projected into many-dimensional spaces, where each dimension represents a feature of the data [9], [10]. Dimensional space is divided into sections based on the classification of learning data. The best value of K for this algorithm depends on the data, in general, the high K value will reduce the noise effect on the classification, but make the boundary between each classification becomes more blurred.

Naïve Bayes Classifier is the fastest and simplest method of Bayesian Learning [10]. This is derived from Bayes's theorem and the hypothesis of freedom, resulting in an opportunity-based statistical classification. The Bayes method is a statistical approach for inducing inference induction on classification problems.

3. Results and Discussion

The test is done at the pre-process stage and classification stage. The results of the preprocessors are used as data on the classification method. This test aims to determine the success of the system in leaf classification. The trial scenarios were performed by means of data separated into two parts ie 75% used as training data and 25% used as test data. Train data is used for feature extraction while test data is used to test the accuracy of the system in performing leaf classification. The classification results will be recorded and compared with the actual classification. From the classification results are then calculated the accuracy.

3.1. Preprocessing Results

In the process preparing the image processing for several stages. Includes grayscale process, median filter, binarization and erosion of image. Grayscale image of the conversion results then made the process of the median filter with the aim to remove the noise from the image.

The grayscale image binarization process is converted a black and white image. The next step is the erosion process. This process aims to improve the resulting image of the binarization process. This process is done by removing the leaf structure so as to obtain a clear object.

The result of image erosion process is done a reverse operation to get the image with white color with a black background. The overall preprocessing process is shown in Figure 3. After the pre-processed step is completed then leaf feature extraction is done using invariant moment method.

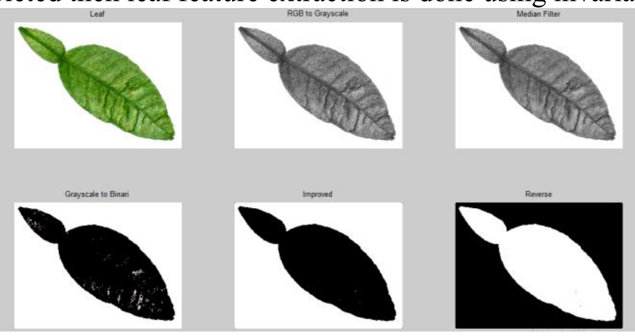

Figure 3. Preprocessing Result

3.2. Feature Extraction Results

The feature extraction stage has succeeded in producing a feature vector containing seven values from the invariant moment. Examples of values on feature extraction generated by the system are shown in Table 1.

Table 1. The result of extraction data.

Data	Seven Moment Invariant						
	hu1	***hu2***	***hu3***	***hu4***	***hu5***	***hu6***	***hu7***
Leaf1	162.787839479	0.10442111865	0.25563241080	0.00070082711	0.00000000373	0.00000336590	0.00000000860
Leaf2	162.752889166	0.03050220396	0.17991796911	0.00019704409	-0.0000000007	0.00000000790	-0.0000000009
Leaf3	712.263937048	480.065934073	3.32494417946	2.96713929571	0.00931942558	2.04119278500	-0.0000630451
Leaf 4	661.335726428	410.302965262	1.29919715329	1.10734941945	0.00132820298	0.68937913577	0.00000090850
Leaf 5	223.372663363	24.2855879465	0.00023134486	0.00040517418	0.00000000008	0.00005863919	0.00000000008
Leaf 6	227.918679497	26.2228166097	0.02282850949	0.00378480935	0.00000003062	0.00045822595	0.00000001731

Leaf 7	233.496841301	28.7418562294	0.10310639903	0.03168439226	0.00000180680	0.00528332967	-0.0000001227

3.3. Classification Results

Classification testing was conducted on 25 test data. Test data will be classified into training data. The classification process is carried out using the Naïve Bayes Classifier and K-Nearest Neighbor (KNN) method. The results of leaf classification testing based on the moment invariant value are shown in Table 2.

Table 2. The result of each classification test.

Data	Class			Result		Data	Class			Result	
	Ground truth	KNN	Naïve Bayes	KNN	Naïve Bayes		Ground truth	KNN	Naïve Bayes	KNN	Naïve Bayes
Leaf1	1	1	1	TRUE	TRUE	Leaf14	3	3	3	TRUE	TRUE
Leaf2	1	2	2	FALSE	FALSE	Leaf15	3	3	3	TRUE	TRUE
Leaf3	1	1	1	TRUE	TRUE	Leaf16	4	4	4	TRUE	TRUE
Leaf4	1	2	1	FALSE	TRUE	Leaf17	4	2	2	FALSE	FALSE
Leaf5	1	1	1	TRUE	TRUE	Leaf18	4	4	4	TRUE	TRUE
Leaf6	2	2	2	TRUE	TRUE	Leaf19	4	4	4	TRUE	TRUE
Leaf7	2	1	2	FALSE	TRUE	Leaf20	4	4	4	TRUE	TRUE
Leaf8	2	2	2	TRUE	TRUE	Leaf21	5	5	5	TRUE	TRUE
Leaf9	2	2	2	TRUE	TRUE	Leaf22	5	5	5	TRUE	TRUE
Leaf10	2	2	1	TRUE	FALSE	Leaf23	5	5	5	TRUE	TRUE
Leaf11	3	3	3	TRUE	TRUE	Leaf24	5	5	5	TRUE	TRUE
Leaf12	3	5	5	FALSE	FALSE	Leaf25	5	5	5	TRUE	TRUE
Leaf13	3	3	3	TRUE	TRUE						

Table 2 shows the true results representing the corresponding values based on the actual class and also corresponding to the system whereas the false value represents the improper between actual classes and classes in the system. By using the Naïve Bayes Classifier method obtained 21 leaf images classified correctly according to the actual class. From the results of this test, it is obtained system accuracy of 21/25 = 84%. While using K-Nearest Neighbor method obtained 20 leaf pictures are classified correctly according to the actual class. Then obtained the system accuracy of 20/25 = 80%.

4. Conclusion

From the results of tests that have been done include preprocessing, feature extraction and classification obtained conclusions.

1) Usage of median filter and erosion of image can improve leaf image, so it can simplify the process of feature extraction and classification.
2) Classification testing using Naive Bayes Classifier method obtained the accuracy value of 84% and K-Nearest Neighbor (KNN) obtained an accuracy of 80%.

References

[1] S. Wu, F. Bao, E. Xu, Y.-X. Wang, Y.-F. Chang, and C.-L. Shiang, “A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network,” in *ISSPIT 2007 - 2007 IEEE International Symposium on Signal Processing and Information Technology*, 2007.

[2] A. Kadira and L. Nugroho, “Leaf Classification Using Shape, Color, and Texture Features”, International Journal of Computer Trends and Technology,” *International Journal of Computer Trends and Technology*, 2011.

[3] Z. Husin, A. Shakaff, A. Aziz, and R. Farook, “Embedded portable device for herb leaves recognition using image processing techniques and neural network algorithm,” *Computers and Electronics in Agriculture*, vol. 89, pp. 18–29, Nov. 2012.
[4] J. Chaki and R. Parekh, “Plant Leaf Recognition using Shape based Features and Neural Network classifiers,” *International Journal of Advanced Computer Science and Applications*, vol. 2, pp. 41–47, Oct. 2011.
[5] C. A. Priya, T. Balasaravanan, and A. S. Thanamani, “An efficient leaf recognition algorithm for plant classification using support vector machine,” in *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, 2012, pp. 428–432.
[6] S. Theodoridis and K. Koutroumbas, “Pattern Recognition,” in *Pattern Recognition (Fourth Edition)*, Boston: Academic Press, 2009.
[7] F. Liantoni and L. Agus Hermanto, “Adaptive Ant Colony Optimization on Mango Classification Using K-Nearest Neighbor and Support Vector Machine,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 3, p. 75, Oct. 2017.
[8] Chapman and Hall, “The Top Ten Algorithms in Data Mining,” *CRC Press*, 09-Apr-2009.
[9] L. Rokach and O. Maimon, “Clustering Methods,” in *Data Mining and Knowledge Discovery Handbook*, Springer, Boston, MA, 2005, pp. 321–352.
[10] G. Goujon, Chaoqun, and W. Jianhong, *Data Clusterin Theory, Algorithms, and Applications,*. Virginia, 2007.