

Learning Vector Quantization Image for Identification Adenium

Resty Wulanningrum*, Bagus Fadzerie Robby

Department of Information Technology, University of Nusantara PGRI Kediri, Kampus 2, Gang 1 Mojoroto Kediri

*Corresponding author, e-mail: resty0601@gmail.com

Abstract

Information and technology are two things that can not be separated and it has become a necessity for human life. Technology development at this time was not only used for intelligence purposes only, but has penetrated the world of horticulture. Adenium is one of the plants are much favored by ornamental plants lovers. Many of cultivation adenium who crosses that appear new varieties that have the color and shape is similar to each other. From this case, then made an application that can identify the type of adenium based on the image of that flower. Learning Vector quantization is one of the algorithms that used for clustering. Based on test scenarios were performed, image identification applications Adenium petals produce an accuracy of 86.66% with a number of training dataset of 135 images and datasets with a test as many as 45 images max epoch 10 and learning rate between 0.01 to 0.05.

Keywords: learning vector quantization, image, adenium

Copyright © 2016 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Adenium Obesum (japanese adenium) as known as desert rose have more than 100 species spread in Indonesia and have become the one of trade commodities with a competitive prices [1]. No wonder many of ornamental plants lovers cultivation the adenium become a commercial. To get the result of crosses with beautiful flower and great stump, we should to identify what type of Adenium is to be crossed, because more better the result of crosses more higher the prices that offered, but the limitation of the human left and right brain ability to remember names and pattern matching flower shape with the same as well as the absence of a program that helps in the identification of each type of Adenium, is a several factors that cause inaccuracies in the identification of adenium species. Therefore we need a method that can identify adenium types with a high degree accuracy.

The research to identify a type of object has been a lot performed, one of them using Learning Vector Quantization. LVQ algorithm selected because this algorithm can automatically learn to classify input vectors [2]. Previous research has been performed by Andri [3] of STMIK Mikroskil extracting characteristic patterns of bottles using Learning Vector Quantization with a size of 20x20 boxes with each box consists of black and white representing a value of 0 and 1 which then bits and weights value was trained as knowledge in the process of familiarization with the result 88.88%.

Learning Vector quantization is a method that can help to identify the characteristic image. This method is one method that is unsupervised learning so competitive layer will automatically learn to classify the input vector that given. If some input vector has a very close distance, then the input vectors will be grouped in the same class. This method is the right path when used to help identify the types of adenium to obtain results in line with expectations.

For the learning process using Learning Vector Quantization, flower image trained first, then proceed through the process of Euclidean Distance. The output of the learning phase is the weight of flowers image that will be stored for subsequent use as a reference in the testing process. From this process will be result a identification of adenium species.

2. Research Method

This research uses a several image processing to obtain the results of analysis and testing. The method used is as follows:

1. Input image

In the early stages, input image the training to be processed. Training image is a image that will be used for the training process before introduction process. Image or images that will be used is image of adenium petals.

2. Grayscale

At this stage, the input image will be converted to grayscale. Which at this stage is intended to normalize the color of pixels in the image.

3. Edge Detection

Use of edge detection is to clarify the image pattern adenium, edge detection is the division of the two regions have different intensities of gray value, so the image of Adenium petals will be clearly with the addition of edge detection

4. Classification

In classification process or clustering algorithms using Learning Vector Quantization (LVQ). LVQ will perform the introduction before the input image pattern that should be presented in the form of vector in order to know his class.

5. Match

Matching method that use is Euclidean Distance. The purpose of using Euclidean Distance method is to search distance from image training to image testing.

2.1. Digital Image

A digital image is a discrete two-dimensional function $f(x, y)$, which has been quantized over its domain and range. Without ss of generality, it will be assumed that the image is rectangular, consisting of Y rows and columns. The resolution of such an image is written as $X \times Y$. By convention, $f(0, 0)$ is ken to be the top left corner of the image ($X - Y -$) the bottom right corner [4].

Automatic (machine) recognition, description, classification, and grouping of patterns are important problems in a variety of engineering and scientific disciplines such as biology, psychology, medicine, marketing, computer vision, artificial intelligence, and remote sensing. A pattern could be a fingerprint image, a handwritten cursive word, a human face, or a speech signal. Given a pattern, its recognition/classification may consist of one of the following two tasks: 1) supervised classification (e.g., discriminant analysis) in which the input pattern is identified as a member of a predefined class, 2) unsupervised classification (e.g., clustering) in which the pattern is assigned to a hitherto unknown class. The recognition problem here is being posed as a classification or categorization task, where the classes are either defined by the system designer (in supervised classification) or are learned based on the similarity of patterns (in unsupervised classification). These applications include data mining (identifying a "pattern", e.g., correlation, or an outlier in millions of multidimensional patterns), document classification (efficiently searching text documents), financial forecasting, organization and retrieval of multimedia databases, and biometrics. The rapidly growing and available computing power, while enabling faster processing of huge data sets, has also facilitated the use of elaborate and diverse methods for data analysis and classification. At the same time, demands on automatic pattern recognition systems are rising enormously due to the availability of large databases and stringent performance requirements (speed, accuracy, and cost). The design of a pattern recognition system essentially involves the following three aspects: 1) data acquisition and preprocessing, 2) data representation, and 3) decision making. The problem domain dictates the choice of sensor(s), preprocessing technique, representation scheme, and the decision making model. It is generally agreed that a well-defined and sufficiently constrained recognition problem (small intraclass variations and large interclass variations) will lead to a compact pattern representation and a simple decision making strategy. Learning from a set of examples (training set) is an important and desired attribute of most pattern recognition systems. The four best known approaches for pattern recognition are: 1) template matching, 2) statistical classification, 3) syntactic or structural matching, and 4) neural networks [5].

2.2. Edge Detection

The edge detection algorithms can be generally classified based on the behavioural study of edges with respect to the operators. Different edge-detection approaches can be broadly classified under Classical or Gradient based edge detectors (first derivative), Zero crossing (second derivative) and Optimal edge-detector [6].

LVQ is a method for training the layers of supervised competitive. Competitive layer will automatically learn to classify the given input vector. If some input vector has a very close distance, then the input vectors will be grouped in the same class [2].

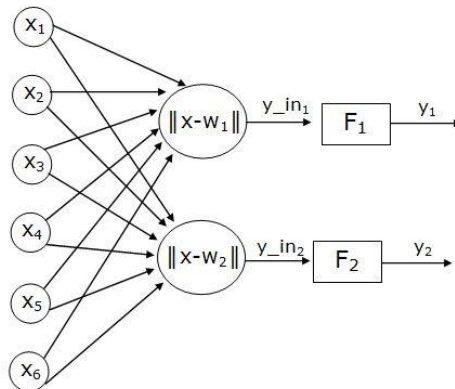


Figure 1. Example architecture LVQ

Euclidean method is the method of classification of its closest neighbors to calculate the distance between two objects; this method is also called the Euclidean distance [7]. Distance calculation formula is written as follows:

$$d_e = \sqrt{\sum_{k=1}^m (fd_{i,k} - k_j)^2}$$

$$d_e = \sqrt{\left(\begin{pmatrix} fd_{11} & fd_{12} & \dots & fd_{1m} \\ fd_{21} & fd_{22} & \dots & fd_{2m} \\ \dots & \dots & \dots & \dots \\ fd_{m1} & fd_{m2} & \dots & fd_{mm} \end{pmatrix} - \begin{pmatrix} k_1 & k_1 & \dots & k_1 \\ k_2 & k_2 & \dots & k_2 \\ \dots & \dots & \dots & \dots \\ k_m & k_m & \dots & k_m \end{pmatrix} \right)^2}$$

$$d_i = \left[\sqrt{(fd_{1i} - k_1)^2 + (fd_{2i} - k_2)^2 + \dots + (fd_{Mi} - k_M)^2} \right]$$

where i = 1, 2, 3, ..., MM.

Information:

- d_e: euclidean distance
- fd_i: weight training image
- k_j: Data weighting test image
- m: the number of training data

From the weight Euclidean (d_e) above which has the smallest weighting to be used as a fitness value of an image signature used for the introduction.

a. Problem analysis

Many of ornamental plants lovers cultivation the adenium become a commercial, To get the result of crosses with beautiful flower and great stump, we should to identify what type of Adenium is to be crossed, because more better the result of crosses more higher the prices that offered, But the limitation of the human left and right brain ability to remember names and pattern matching flower shape with the same as well as the absence of a program that helps in the identification of each type of Adenium, is a several factors that cause inaccuracies in the identification of adenium species. Therefore we need a method that can identify adenium types with high degree accuracy.

b. System planning

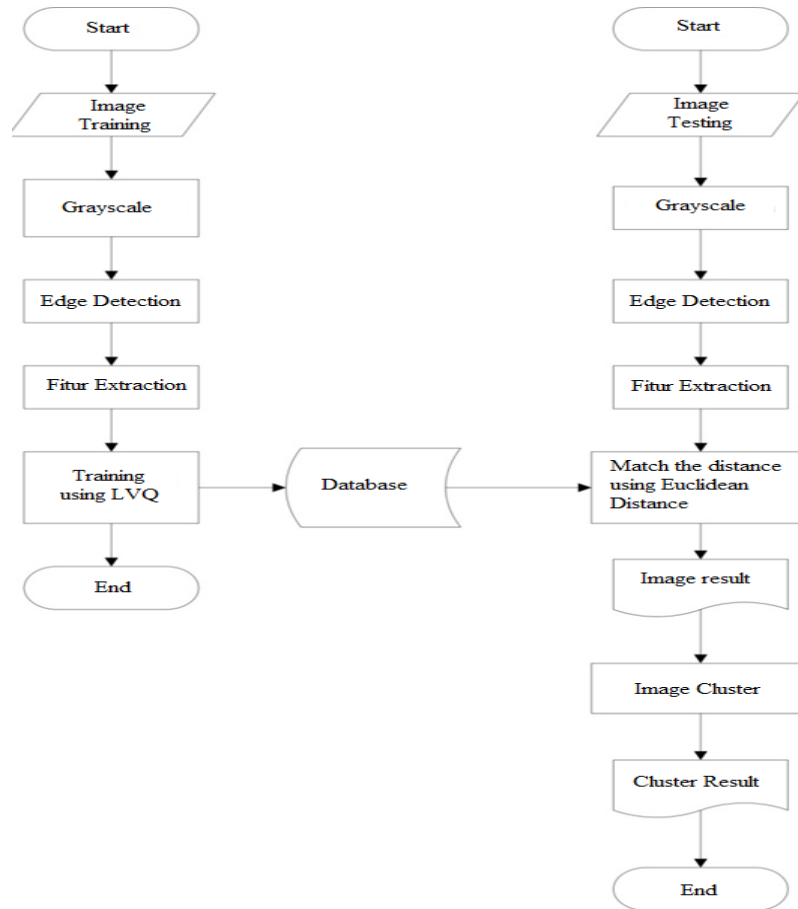


Figure 2. Flowchart System

In Figure 2 looks the flow of training data and testing, the first training input image, and then processed grayscale, after grayscale do edge detection. Edge image obtained will be characteristic extraction using LVQ. Characteristic extraction results will be stored on a database that will be used for the matching process using euclidean distance. The matching result is the adenium image identification recognized as one of Adenium.

3. Scenario Test

Scenario testing applications system here is to obtain the best test results in this study. The scenarios used are as follows:

Table 1. Scenario Test Training Data

Learning rate	Epoch	Obessum	Training WhiteKnight	Explora	Total
0,01	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1,2,3,4,5,6,7,8,9,10	30
0,01	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17	51
0,01	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45	135
0,01	10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1,2,3,4,5,6,7,8,9,10	1,2,3,4,5,6,7,8,9,10	30
0,01	10	1, 2, 3, 4, 5	1,2,3,4,5	1,2,3,4,5	15

Table 1 above is a scenario tests table conducted with variation amount of diverse training data

In the first scenario training data, each type of adenium taken 10 image file names begin with "Adenium1" to "Adenium10" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In second scenario training data, each type of adenium taken 17 image file names begin with "Adenium1" to "Adenium17" with learning rate between 0.01 to 0.5 and epoch as many as 10

In third scenario the training data of each type of adenium taken 45 image file names begin with "Adenium1" to "Adenium45" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In the fourth scenario training data, each type of adenium taken 10 image file names begin with "Adenium1" to "Adenium10" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In fifth scenario training data, each type of adenium taken 5 images with file names begin "Adenium1" to "Adenium5" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

Table 2. Scenario Testing

Learning rate	Epoch	Testing			Total
		Obessum	WhiteKnight	Explora	
0,01	10	11,12,13,14,15	11,12,13,14,15	11,12,13,14,15	15
0,01	10	18, 19, 20, 21, 22, 23, 24, 25, 26, 27	18, 19, 20, 21, 22, 23, 24, 25, 26, 27	18, 19, 20, 21, 22, 23, 24, 25, 26, 27	30
0,01	10	46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60	45
0,01	10	11,12, 13, 14, 15, 16, 17, 18, 19, 20	11,12, 13, 14, 15, 16, 17, 18, 19, 20	11, 12, 13, 14, 15, 16, 17, 18, 19, 20	30
0,01	10	6, 7, 8, 9, 10, 11, 12, 13, 14	6, 7, 8, 9, 10, 11, 12, 13, 14	6,7,8,9,10,11,12,13,14	27

Table 2 above is a scenario tests table conducted with a variety amount of diverse testing data

In the first scenario testing data, each type of adenium taken 5 images with file names begin "Adenium11" to "Adenium15" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In second scenario testing data, each type of adenium taken 10 image file names begin with "Adenium18" to "Adenium27" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In third scenario testing data, each type of adenium taken 15 image file names begin with "Adenium46" to "Adenium60" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In fourth scenario testing data, each type of adenium taken 10 image file names begin with "Adenium11" to "Adenium20" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

In fifth scenario testing data , each type of adenium taken 9 images with file names begin "Adenium6" to "Adenium14" with learning rate between 0.01 to 0.5 and the epoch as many as 10.

4. Test Results

Test were conducted by using the adenium image identification application system were built using language Delphi 7 programming.

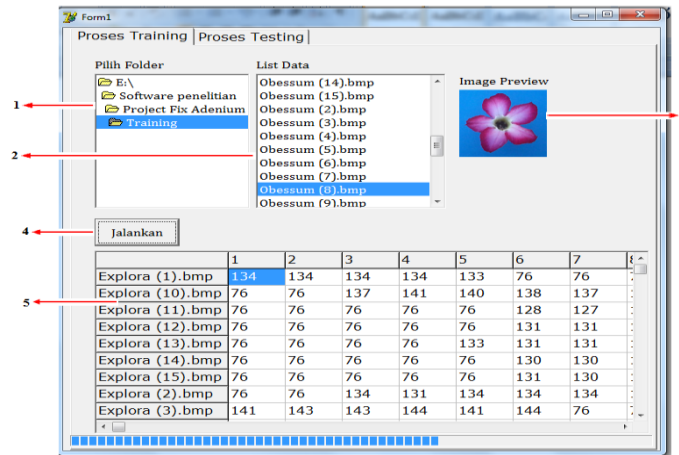


Figure 3. Interface Training

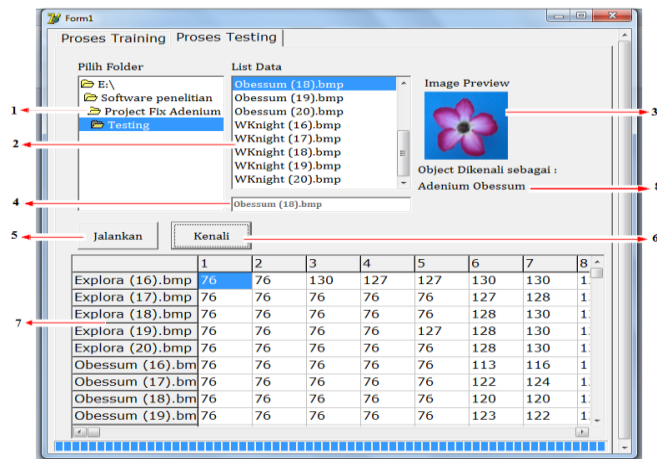


Figure 4. Interface Testing

The components used in these forms and steps on using this application are as follows: the component indicated by the number 1 is the Directory listbox used to select the folder where the dataset is, then if the destination folder has been selected so the files in that folder will be displayed File listbox component is indicated by the number 2.

Components are numbered 3 is image component, this component serves to display the image preview that is being processed. Next to process the data that has been selected, used the Button component, marked with No. 4. After the image processing, the characteristic extraction value will be accommodated in a string grid komponen indicated by a number 5. Then the process will be forwarded in the testing form to identify Adenium flower types.

The stages that must be done by the user when using this app is the first user will be prompted to select the directory that contains the training data and then data in the directory will be displayed on FileListBox. Image of each image will be displayed on the preview image that is beside FileListBox.

After button "Jalankan" clicked, then the training process begins, the value of the characteristic extraction from each image will be accommodated on Stringgrid then compared with data for testing there.

The components are used and step application user in the form of testing is basically as same as the form of training, but there are some additional components that need to be explained here, among other things:

TextField component indicated by the number 4 is used to accommodate the path where the image file is stored and to display the name and extension of the image. Button indicated by the number 5 will be active when the training process in the form of training has

been completed, while the button indicated by number 6 will be active when the characteristic extraction process in the form of testing has been completed. Components that indicated number 8 is the label that will display the identification result of Adenium image that will run automatically when the button number 5 in execution. But if the user requires the identification process manually, the user can select one image that displayed in the file listbox (component number 2) and then press the button number 6, and then the result will be shown on the label indicated by the number 8.

Same as processes that exist in the training data pages, the user will be prompted to choose a directory that contains the training data and then data in directory will be displayed on FileListBox. Image of each image will be displayed on the preview image that is beside FileListBox.

After button "Jalankan" is clicked, and then the feature extraction process begins. Extraction characteristics Value of each image will be accommodated on Stringgrid then compared with training data that has been accommodated in stringgrid1. When the feature extraction process is complete then the button "kenali" will be active.

When the "kenali" button is active, click on one of the existing image on file listbox and then click the button "kenali", the identification results will be shown on the label are in addition to the preview image.

In testing process of identification systems Adenium types is performed on three kinds of Adenium samples, Adenium Obesum, Adenium Eksplora, Adenium WKnight with images sample taken with different flower positions, with image size 100x100px. Determined alpha value (learning rate) 0.01. Here is a table of test results on the Adenium flowers image.

Table 3. Scenario Test Results

Learning Rate	Epoch	Konvergen		Prosentase
		True	False	
0,01	10	4	1	80%
0,01	10	7	3	70%
0,01	10	13	2	86,66%
0,01	10	6	4	60%
0,01	10	4	5	44,44%

Of the five scenarios test that have been done with the amount of training data and testing data is varied, the results obtained were also varied. So it can be concluded that more training data will create a higher level of accuracy applications.

5. Conclusions

Implementation of digital image processing for image identification adenium using Learning Vector Quantization with the approach Euclidean Distance can identify with a maximum accuracy of 86.66%. The size of the percentage of accuracy of the data depends on the amount of data Training and testing data is processed. It can be seen from the percentage of accuracy with varied results of several test scenarios.

Reference

- [1] Sugih Octa. Variasi Adenium Agar Rajin Berbunga, Jakarta: Penebar Wadaya. 2008.
- [2] Kusumadewi Sri. Membangun Jaringan Syaraf Tiruan Menggunakan Matlab & Excel Link, Yogyakarta: Graha Ilmu. 2004.
- [3] Andri. Implementasi Segmentasi Citra Algoritma Learning Vector Quantization (LVQ dalam Pengenalan Bentuk Botol. *STMIK Mikroskil*. 2012; V(13).
- [4] Uddin, Muhammed Nasir. A comprehensive Study of Digital Image Peocessing for Finding Image Quality Dependencies. *International Journal of Scientific and Research Publication*. 2012; 2(3): 1-8.
- [5] Basul, Jayanta Kumar, Debnath Bhattacharyya, et al. Use of Artificial Neural Network in Pattern Recognition. *International Journal of Software Engineering and Its Application*. 2010; 4(2): 22-34.
- [6] Saluja, Sonam, Arandhana Kumari Singh, Sonu Agrawal. A Study of Edge-Detection Methods. *International Journal of Advanced Research in Computer and Communication Engineering*. 2013; 2(1): 994- 990.
- [7] Navarette, Pablo, Javier Ruiz-del-Solar. *Self Organizing Maps- Kohonen*. 2003.