CLASSIFICATION IMPLEMENTATION OF IMAGE BASED FEATURE EXTRACTION TO DETECT DENTAL ABNORMALITIES USING METHODS NAIVE BAYES AND COOCURRENCE GRAY LEVEL MATRIX (GLCM)

Luky Muhamad Harisman[1], Erick Wijaya, S.Kom., M.T.[2]
[1,2] Teknik Informatika-Universitas Komputer Indonesia
Jalan Dipatiukur No. 112-116, Coblong, Lebakgede, Bandung
E-mail : lukyharisman@gmail.com[1], erick.wijaya@email.unikom.ac.id.[2]

ABSTRACT

Dental abnormalities or so-called dental anomalies that tooth whose shape deviates from its original form. Deformities and tooth size between Gemination and Macrodontya almost the same cause dentist is difficult to distinguish. Therefore we need a utilization of computer-based image processing to assist in the analysis of the teeth by means classifies it.

One of method image processing is texture feature extraction, in this research the texture feature extraction method is the method Concurrency Matrix. Cooccurence method Matrix is a matrix whose elements are the number of pairs of pixels that have a certain level of brightness, where the pixel pair separated by a distance (d), and with an inclination angle (θ). Naïve Bayes classification is the simplest method by using the existing opportunities, where it is assumed that every variable X is free (independence).

Based on test results, obtained 66.6% accuracy value proves that Naive Bayes less well in classifying the image of dental abnormalities, because the data of texture feature extraction dental abnormalities with co-occurrence matrix method has a near distance interval between class. Naive Bayes classification so less can go well when the classification.

Keywords: Texture Citra, Feature Extraction, Co-Occurrence Matrix, classification, Naive Bayes, Teeth

1. INTRODUCTION

Teeth are structures modified epidermal and mesodermal contained in the mouth of many vertebrate [4]. Gigi is one organ that is essential for humans. Role in supporting dental health is often neglected, one of which abnormalities in the teeth. Dental abnormalities or so-called dental anomalies that tooth whose shape deviates from its original form [4]. Abnormalities or anomalies in the teeth indicate an anomaly on the body symptoms are caused by disorders of growth and development on the teeth.

Based on interviews with Dentist Arien Haryatin that abnormalities or anomalies in the teeth has eight characteristics of one of them is abnormal size and shape of the teeth. According Dentist Arien Haryatin case abnormalities size and shape of the teeth are often found there are three, namely Geminasi, Makrodontia, and Mikrodontia. To detect abnormalities of shape and size of the patient's teeth, can be done in the same way with dental disease in general, ie through the examination of the shape, texture and segmentation gear [6]. For a dentist to detect abnormalities of shape and size of the patient's teeth is done based on knowledge and experience alone therefore results obtained approximate diagnosis could be wrong because of human limitations. In addition, the deformities and tooth size between Geminasi and Makrodontia almost the same cause dentist is difficult to distinguish. In determining the deformation and tooth size should be done carefully due to abnormal size and shape of the teeth can be made so that the selection of patient diagnosis and treatment care plan that can run optimally [1]. Therefore we need a utilization of computer-based image processing to assist in the analysis of the teeth by means of classifying it appropriately.

In previous research on image processing by means of classification based on the texture has been studied by L. Roidel Simarmata, to divide the tooth region in order to view other diseases through the gears by means of classifying dental image-based texture with Gabor filter, obtained accuracy rate of 88.3 % [7] and research Shinta Puspasari, the classification of dental arch using back propagation algorithm based on digital image texture features, obtained accuracy rate of 66% [10].

Had previously been conducted research on feature extraction by Riekal Fahmi on the analysis of texture feature extraction method and the gray level difference Naive Bayes methods for identifying human tongue disease, obtained accuracy rate of 81.1% [3] and study Ariantoko Kusmian on naïve Bayes algorithm implementation for image classification based on the texture feature extraction method kookurensi matrix, obtained accuracy rate of 96% [2], which proves that the co-occurrence matrix method can extract image texture characteristics well. So in this study will use the Naïve Bayes

method to recognize by classifying abnormalities in tooth size and shape through texture to the image of the tooth. Methods Grey Level Co-occurrence Matrix is used as a feature extraction for image texture analysis on teeth.

Based on the above problems, this study will identify abnormalities of teeth based on the size and shape of the teeth that can be detected through the image texture tooth by classifying abnormalities of teeth and using different methods. This study will implement GLCM and Naive Bayes method for detecting abnormalities of teeth based on the shape and size of the texture image of the tooth. Therefore, the authors take the topic "Implementation of Image Classification Based Feature Extraction for Detecting Abnormalities Teeth Using Naïve Bayes Method And Method Cooccurence Gray Level Matrix (GLCM)".

1.1 Formulation of the problem

Based on preliminary described, this research is to formulate the problem to be discussed is:

1. How to apply the Naïve Bayes method and the method of Grey Level Co-occurrence Matrix into an application to detect abnormalities in the three classes of teeth.
2. What is the level of accuracy Naïve Bayes methods and methods of Grey Level Co-occurrence Matrix for the three classes detect abnormalities in the teeth.

1.2 Purpose And Objectives

Based on the formulation of the problem that has been studied, the purpose of this thesis is to implement image classification using Naive Bayes method and the method of Grey Level Co-occurrence Matrix into an application to detect abnormalities of teeth.

While the purpose of this research is:

1. Knowing whether Naive Bayes methods and methods of Grey Level Co-occurrence Matrix can be used and effective in detecting abnormalities in the teeth.
2. Knowing the accuracy Naïve Bayes methods and methods of Grey Level Co-occurrence Matrix for detecting abnormalities in the teeth.

2. CONTENTS OF RESEARCH

2.1 Image processing

Image processing (image processing) is processing the image, in particular by using a computer, the image of the quality to be better [1]. Digital image processing is a branch of science informatics struggling to transform an image or image into another image using certain techniques [9]. Image processing is widely used in various fields such as education, medicine, industry etc.

2.2 Texture Image

Texture is the properties or characteristics possessed by an area (in the image) is quite large so naturally to these properties can be repeated in the area. Definition of texture in this case is the regularity of certain patterns that are formed from the arrangement of pixels in a digital image [5]. To form a texture of at least two requirements that must be met include:

1. Consists of one or more pixels that form the patterns of primitive (the smallest parts). The forms of these primitive patterns can be titk, straight lines, curved lines, size and others which are the basic elements of a form.
2. The emergence of primitive patterns repeated at intervals and specific direction so that it can be predicted or discovered characteristic recurrence.

2.3 Texture analysis

Texture analysis is the basis of a wide range of applications, application of texture analysis, among others: remote sensing, medical imaging, identification of quality of a material (wood, leather, textiles, etc.), and also a wide range of other applications. In the image analysis, texture measurement is categorized into five main categories: statistical, structural, geometry, basic model, and signal processing.

Statistical approach considering that the intensity generated by a random two-dimensional field, the method is based on the frequency of the frequency space. Examples of statistical methods is the autocorrelation function, Concurrency Matrix, Fourier transformation, frequency edge, run length and other satistis methods. Structural engineering related to developing the smallest part of an image. Examples of structural methods are fractal models. Methods based on the geometry of the existing geometry on the element of texture. Examples of the method is the basic model of a random field. While the signal processing method is a method that is based on the frequency analysis such as Gabor transform and wavelet transform [1].

2.4 Cooccurrence Matrix Method

Cooccurrence Matrix Method is a matrix whose elements are the number of pairs of pixels that have a certain level of brightness, where the pixel pair separated by a distance d, and with an inclination angle θ. In other words, the matrix is the probability kookurensi graylevel i and j of the two pixels apart at a distance d and angle θ. (Ahmad U., 2005) A neighboring pixels that have the distance d between them, can be located in eight different directions, as shown in image (2).

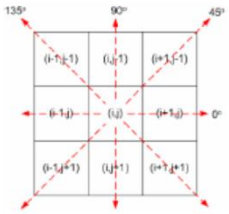


Image 2 Adjacency Relationship Between Pixels As A Function Of Spatial Orientation And Distance

In the co-occurrence matrix method Haralick et al propose various types of statistical characteristics of texture that can be extracted Some of these include among others are: Contrast (contrast), homogeneity (homogeneity), Entropy (Entropy), Energy (Energy) and Dissimilariti (dissimilarity). The equation for these features are as follows:

Contrast

Shows the size of the deployment (moment of inertia) elements of the image matrix. If located far from the main diagonal, the value of great contrast. Visually, the contrast value is a measure of the variation between the degree of gray an image area. The results contrast calculation related to the amount of gray in the image intensity diversity.

$$\textbf{\textit{Contrast}} = \sum_{i,j=0}^{n-1} P(i,j).(i-j)^2$$

Where:
i and j are gray level characteristic of two adjacent pixels resolution.
p (i, j) is Probability column (i, j).

Homogenity

Homogenity indicates a kind gray level degree of homogeneity image. Homogeneous image will have a great homogeneity prices.

$$\textbf{\textit{Homogenity}} = \sum_{i,j=0}^{n-1} P(i,j)/[1+(i-j)^2]$$

Where:
i and j are gray level characteristic of two adjacent pixels resolution.
p (i, j) is Probability column (i, j).

Entropy

Entropy can show irregularity, size, shape, if a large entropy value for the image with uneven degrees of gray transitions and image of little value if the structure is irregular (variable).

$$\textbf{Entropy} \quad \sum_{i,j=0}^{n-1} P(i,j).logP(i,j)$$

Where:
i and j are gray level characteristic of two adjacent pixels resolution.
p (i, j) is Probability column (i, j).

Energy

Energy expressed a measure of concentration gray level pair with particular intensity in the matrix, where (i, j) declared value on row i and column j in the co-occurrence matrix.

$$\textbf{\textit{Energy}} \quad \sum_{i,j=0}^{n-1} P(i,j)^2$$

Where:
i and j are gray level characteristic of two adjacent pixels resolution.
p (i, j) is Probability column (i, j).

Dissimilarity

Stated dissimilarity measure inequality gray level image so as to provide an indication of the structure in the image.

$$\textbf{\textit{Dissimilarity}} = \sum_{i,j=0}^{n-1} P(i,j).|i-j|$$

Where:
i and j are gray level characteristic of two adjacent pixels resolution
p (i, j) is the relative frequency matrix of two adjacent pixels resolution.

2.5 Classification

Classification can be defined as a job training / learning to target function f that maps each set of attributes (features) x to one of a number of class labels y available. Job training will produce a model which is then stored as a memory [8].

Classification algorithms using training data to create a model. The model has been constructed is then used to predict the class label of new data that is not yet known.

2.5.1 Classification Naïve Bayes

Naïve Bayes is a simple probabilistic-based prediction technique which is based on the application of Bayes' theorem [8].

1. The assumption of independence (independence) is strong (naive).
2. The model used is the "model of independent features".

Naïve Bayes classification is the simplest method by using the existing opportunities, where it is assumed that every variable X is free (independence). Because the assumptions are not mutually dependent variable, then obtained the following equation:

$$P(X \mid Y = y) = \prod_{i=1}^{q} P(Xi|Y = y)$$

The data used can be categorical (discrete) or continuous. Because the results of the image feature extraction is a continuous data such as figures on measurement results contrast levels, homogeneity, entropy, energy, and dissimilarity in feature extraction. Therefore For continuous data can be completed using the following steps.

Training :

1. Calculate the average (mean) of each feature in the training dataset with the following equation.

   $\mu = \frac{\sum xi}{n}$

   Where:
   μ = Mean
   n = Amount of data
   Σxi = Amount of data values
2. Then calculate the variance of the blood money of the training dataset as the following equation.

   $$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Xi - \mu)^2$$

   Where:
   σ^2= vVrians
   μ = Mean
   n = Amount of data
   Σxi = Amount of data values

Testing :

1. Calculate the probability (Prior) for each class that is by counting the amount of data each class divided by the total number of overall data.
2. Next calculate the probability density using a variance value equation. Expressing the relative probability density function. Data with mean μ and standard deviation σ, the probability density function is:

   $$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Xi - \mu)^2$$

   Dimana :
   x = Data input
   π = 3,14
   σ= Standard Deviation
   μ = Mean
3. Having obtained the probability density values, then calculate the posterior of each class using the following equation.

   $$Posterior = \frac{P(Ci)\, x\, P(X1 \ldots\ldots\ldots\ldots\ldots\ldots Xn|Ci)}{P(X1 \ldots\ldots\ldots\ldots\ldots\ldots Xn|Ci)}$$

4. The greatest posterior is the appropriate grade.

2.6 Analysis System To Be Set

Analysis of the process to be conducted in this study were divided into several stages such as: tooth image input, image processing (preprocessing), GLCM Feature extraction, testing with Naive Bayes classification method to generate the output data classification results. Here are the stages of the process analysis to be performed can be seen in image 3.

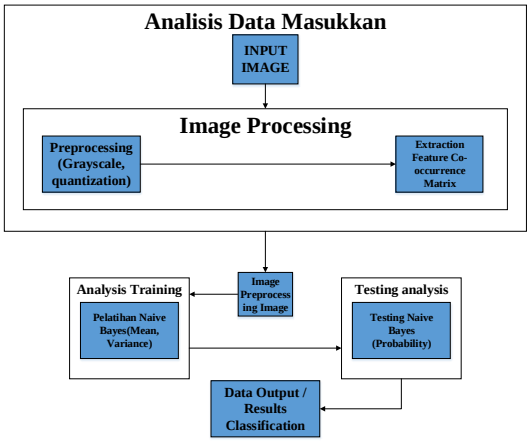


Image 3 Process Flow Diagram

2.6.1 Insert Data Analysis

Data input to the analysis of the calculations in this study of the image of the tooth with a variety of abnormalities of teeth and going through several stages of image processing, such as process, grayscaling and quantitation grayscale image and feature extraction. The following input data:



Image 4 Image Dental Practice 1 Measuring 50x25 Pixel

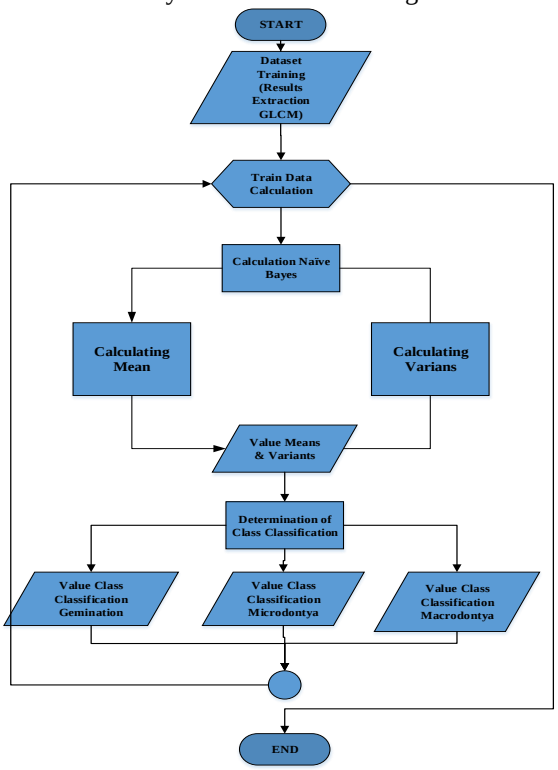
a. grayscale
   The first stage is to change the image of RGB mode, to grayscale. This process aims to simplify the pixel values in a citra.Rumus used for greyscale process is as follows:
   X = (R + G + B) / 3 (1)

b. quantitation
   Next will be the image quantization process that has been through a phase of greyscale, in this study a range of values from 0-255 converted to 0-50. The formula used for scaling the grayscale image of gray values is by the following equation:
   Xb = int (X * (2n -1) / 255

After the image converted to grayscale and will get the grayscale matrix, the matrix will be used as the image extraction process kookurensi. Gray Level Coocurrence Matrix (GLCM) is used to get the value of the content that will be used as baseline data and used as input in naïve Bayes methods. The output value to be obtained is the contrast, homogeneity, entropy, energy, and dissimilarity.

2.6.2 Analysis Training

Training in this training by using naïve bayes, naïve Bayes made to obtain training data, this process is done by calculating the mean and variance. Here is the flow of the training process with a naïve Bayes can be seen in image 5.



Gambar 5 Alur Proses Pelatihan

2.6.3 Analisis Pengujian

Pengujian adalah tahapan pada proses klasifikasi berdasarkan tekstur citra, pada proses ini citra yang dimasukkan dapat dihasilkan nilai probabilitas terbesarnya. Berikut alur pengujian dengan metode naïve bayes dapat dilihat pada gambar 6.

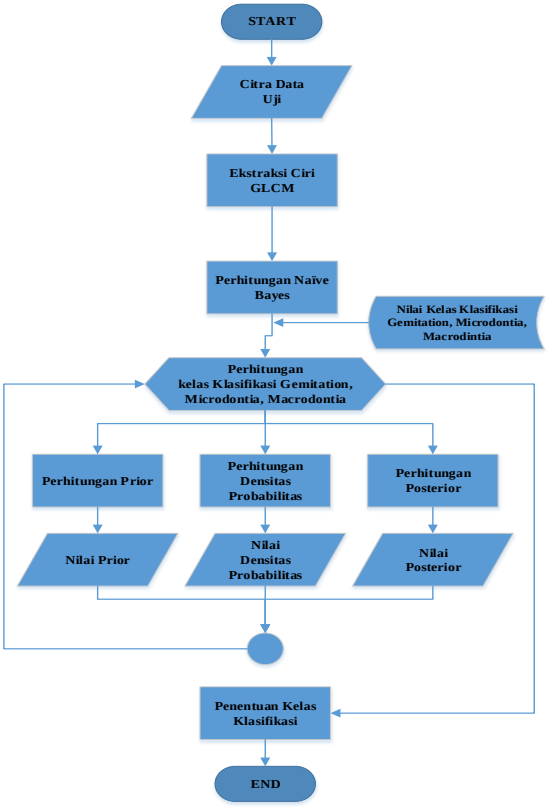


Image 6 Flow Testing Process

2.6.3 Output Data Analysis

Analysis of the output data is a search for the greatest value on the probability value obtained at the time of testing analysis. Flow process output data contained in image 7.

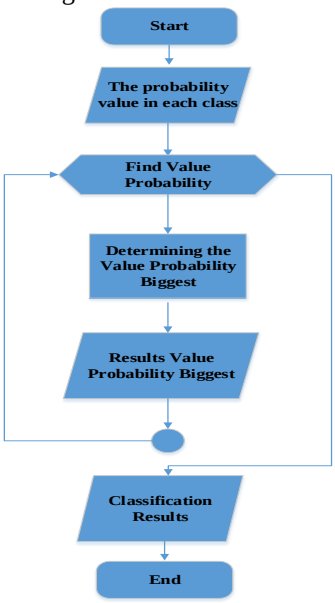


Image 8 Flow Process Output Data

2.7 Testing

In this study, testing is done using three methods. Here is the test to be performed:

1. Test the image included in the training data.
2. Test the image that are not included in the training data (test data).

3. Test the image using the k-fold cross validation, k-fold value used was 2.3.

2.7.1Testing Method 1 Test Image Data That Is Included In Train

Testing method 1 was conducted by examining images included in the training data, this test aims to determine the level of recognition of the image that has been trained, the image data that is used there are 60 pieces of imagery which consists of three classes, with each class there are 20 images.

Table 1 Confusion Matrix Testing Method 1

Class		Prediction				Accuracy
		Gemination	Macrodontya	Microdontya	Jumlah	
Target	Gemination	15	5	0	20	75%
	Macrodontya	9	10	1	20	45%
	Microdontya	10	3	7	20	50%
Average					60	56,6 %

2.7.2Testing Method 2 Testing Image Data That Is Included In Practice

Testing method 2 was conducted by examining the image that are not included in the training data, this test aims to determine the level of the test image recognition training data outside the trainer's image. The image data used to train there are 60 pieces of imagery which consists of three classes, with each class there are 20 images. And also used the test image data, there are 15 pieces of imagery which consists of three classes, with each class there are 5 images.

Table 2 Confusion Matrix Testing Method 2

Class		Prediction				Accuracy
		Gemination	Macrodontya	Microdontya	Jumlah	
Target	Gemination	4	1	0	5	80%
	Macrodontya	1	4	0	5	80%
	Microdontya	2	1	2	5	40%
Average					15	66,6 %

2.7.2Testing using the 3 Test the image using the k-fold cross validation

testing method 3 is the test image using the k-fold cross validation, k-fold value used is 2, 3.

Table 3 Confusion Matrix Test Results Average Average K-Fold 2

Class		Prediction				Accuracy
		Gemination	Macrodontya	Microdontya	Total	
Target	Gemination	18	6	1	25	72%
	Macrodontya	11	11	3	25	44%
	Microdontya	10	7	8	25	32%
Rata-rata					75	49,3 %

Tabel 4Confusion Matrix Rata Rata Hasil Pengujian K-Fold 3

Class		Prediction				Accuracy
		Gemination	Macrodontya	Microdontya	Jumlah	
Target	Gemination	18	6	1	25	72%
	Macrodontya	12	13	0	25	52%
	Microdontya	11	9	5	25	20%
Average						48%

2.8 Conclusion Testing

Based on the results of one test scenario that is testing the same test data with training data, it can be deduced that the naïve Bayes methods can be classified with an accuracy of 56.6%. Based on the test scenario 2 is test of test data that is not contained in the training data, naïve Bayes methods can be classified with an accuracy of 66.6%.

Based on the test scenario 3 is test using the k-fold cross validation, naïve Bayes methods can be classified with an average accuracy of 49.3% to the value of k is 2 and 48% by value of k is 3. From the test results, the input image has an error rate is Microdontia, while the image Geminasi and Macrodontia recognizable although less well. This is because the image feature extraction microdontia has a value that is almost close to the probability of gemination and macrodontia, so making mistakes detection.

From the test results, the accuracy has bad rate, because the feature extraction of data produced by the co-occurrence matrix of the image of dental abnormalities have a great degree of similarity, so that the process of recognition not run well.

3. CLOSING

3.1 Conclusion

From the research that has been done, it can be deduced as follows:

1. Naïve Bayes can perform image classification based on the texture abnormalities of teeth extracted with co-occurrence matrix method. Due to the feature extraction data of the co-occurrence matrix is a form of data continue,

or so-called nominal data, so that the process of data classification feature extraction results can be directly used as input in naïve Bayes classification.

2. Based on test results, the conclusions obtained are naïve Bayes can classify the image of dental abnormalities, but its accuracy is not good because the data extraction dental abnormalities characteristic texture with co-occurrence matrix method has closely spaced intervals between classes. So naïve Bayes classification not run properly when the classification.

3.2 Conclusion Testing

In the making of this final project, there are still many deficiencies that can be corrected for the next development. Some advice that can be given is:

1. Trying to examine the image of a more homogeneous or try to use x-ray image, so as to improve the accuracy of naïve Bayes methods.
2. Adding some other feature extraction, feature extraction such as color, shape, and so forth.
3. To be able to compare the performance of co-occurrence matrix method as this feature extraction, texture analysis can be made by different methods.

REFERENCES

[1] Ahmad, U. 2005 . *Pengolahan Citra Digital dan Teknik Pemogramannya* .Yogyakarta: Graha Ilmu.

[2] Ariantoko Kusmian. 2014 ."*Implementasi Algoritma Naïve Bayes Untuk Klasifikasi Citra Berdasarkan Ekstraksi Ciri Tekstur dengan Metode Matriks Kookurensi*" .Tugas Akhir Teknik Informatika. Universitas Komputer Indonesia . Bandung.

[3] Fahmi Riekal. 2016 . "*Analisis Gray Level Difference Method Dan Metode Naive Bayes Mengindentifikasi Penyakit Lidah Manusia".* Tugas Akhir Teknik Informatika. Universitas Komputer Indonesia . Bandung.

[4] George H. Fried, Ph.D, George J. Hademenos, Ph.D. 2005. *Schaum's Outlines Biologi* Edisi Kedua. Erlangga.

[5] Kadir, A. dan Susanto, A. . 2012. *Teori dan Aplikasi Pengolahan Citra* . Yogyakarta: Andi Offset.

[6] Kahng Luke.2008.Understanding The Natural Tooth.Tersedia di https://www.dentalaegis.com/id/2008/05/esthetics-recreating-the-texture-shape-and-color-of-a-natural-tooth. [Diakses 26-12-2016].

[7] L, Roidel Simarmata1.2012. " *Klasifikasi Citra Gigi Berbasis Tekstur Dengan Filter Gabor*". Jurnal. Fakultas Teknologi Informasi. Institut Teknologi Sepuluh Nopember (ITS). Surabaya.Indonesia.

[8] Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta : Andi.

[9] Purnama, Agus."*Operasi Pengolahan Citra Digital*". Tersedia di http://elektronika-dasar.web.id/operasi-pengolahan-citra-digital/ [Di akses 28 10 2016].

[10] Puspasari Shinta. 2013. "*Klasifikasi Bentuk Lengkung Gigi Menggunakan Algoritma Propagasi Balik Berdasarkan Fitur Tekstur Pada Citra Digital*". Tesis. Program Studi Teknik Informatika. STMIK Global. Palembang.