PENERAPAN METODE RUN-LENGTH DAN ALGORITMA SIMPLE NAIVE BAYES UNTUK IDENTIFIKASI SIDIK JARI

Heri Yulianto Sugandi[1]

[1] Teknik Informatika – Univesitas Komputer Indonesia
Jl. Dipatiukur 112 - 114 Bandung
E-mail : ulonkroses@yahoo.co.id[1]

ABSTRACT

The fingerprint is a medium that could be used to recognize a person's identity. Fingerprints have a which distinguishes characteristics of fingerprint with others. The way to differentiate these characteristics is a way to recognize the difference in the texture of the fingerprint image. Which has a texture image is the image that has a pattern, the pattern of the image occurs repeatedly meet all field image. Different image having different characteristics. The characteristics are the basis for the image classification based on texture. There are several methods to obtain the characteristics of texture in an image, one method to obtain the characteristics of the texture image is the run length matrix. The characteristics of the texture obtained from the run-length matrix method include SRE(short run emphasis), LRE(long run emphasis), GLU(grey level uniformity), RLU(run length uniformity) dan RPC(run percentage). From the results of these characteristics are then used for classification using Naïve Bayes which determines the classification results based on the value of the largest probability. The object being tested is the image of fingerprint.

From the research that has been done, it can be deduced as follows: naïve Bayes can perform image classification based on texture are extracted by the method of run length matrix. Due to the characteristics of the data extracted run length matrix is in the form of data continuously, or so-called nominal data, so that the process of data classification feature extraction results can be directly used as an input in the naïve Bayes classification.

Based on test results, obtained conclusion is naïve Bayes algorithm can classify digital fingerprint image based on the digital image extraction of run-length method and generates 95.8% accuracy rate. because of data from fingerprint feature extraction of textures with run-length matrix method has the advantage of distinguishing among smooth textures and rough textures, so naïve Bayes classification could run most leverage when performing fingerprint image classification.

Keywords: texture images, feature extraction, the run length matrix, classification, Naïve Bayes.

1. INTRODUCTION

Fingerprint (fingerprint) is a reproduction finger palm either intentionally taken, stamped with ink, as well as scars left on objects because never touched leather palms of the hands or feet. Fingerprint identification known to science that studies dactyloscopy fingerprint for recognition purposes back a person's identity by observation the lines contained in line strokes fingers and soles of the feet (Ashbaugh R, 1991). Because the texture fingerprints on every person has a characteristic that is different from one person to another person, differences in the pattern of the fingerprint is used as a means of identification.

In classifying and detecting an object-level accuracy is crucial, because to produce a classification system and object detection required a good accuracy. Previous research by Eko Sediyono (2009), to process the fingerprint image classification using wavelet feature extraction method symlet capable of producing up to 80% accuracy. Wavelet transform is used as a texture analysis which is input to the classification system. The occurrence of errors in the classification of characteristics may occur due to several reasons, among them are, thumb prints were taken using ink stamp affixed on the paper and then scanned to give effect to the sketch of the fingerprint formed, among others, the thickness of the ink stick is too thick or thin, the size of the fingerprint image that is diverse and positions fingerprints are not upright.

Based on existing phenomenon, more research is needed on the fingerprint identification to improve the level of accuracy that is better on the system. In this study, the method used as the image extraction process is run-length and naïve Bayes algorithm for image classification. Run-length is a method for obtaining the characteristics of texture image using an pixel distribution with the same intensity in sequence in one particular direction as primitive. Each primitive is defined over the length, direction, and gray levels. The characteristics of the texture or the parameter of the method of run-length is short run

emphasis (SRE), long run emphasis (LRE), gray level uniformity (GLU), run length uniformity (RLU), run percentage (RPC) (Mita I 2007). Run-length method can produce a characteristic form of diversity, density, roughness, regularity, linearity, frequency, phase, keterarahan, irregularity, smoothness, and others. Results of feature extraction is used for classification. Therefore the extraction process characteristics resulting from the run-length method will generate continuous data which will be processed on the next stage of image classification stages by using naïve Bayes methods. In the study conducted Sri Kusumadewi (2009), naïve Bayes classification process can be used for continuous data and generate a total of 93% of performance testing. Naïve Bayes classification method is one that uses the concept of probability. Methods naïve Bayes classification algorithm is a highly effective and efficient.

Of problems and solutions that have been described, this thesis research will classify fingerprint based texture fingerprints by applying run length for the extraction process the image and methods naïve Bayes for image classification, the expected method of naïve Bayes can classify fingerprints by texture and measure the level of accuracy classification.

1.1 Formulation of the Problem

Based on the background described, this research is to formulate the problem to be discussed is how to implement the run length method for extracting images and naïve Bayes algorithm for image classification.

1.2 Purpose and Objectives

The point of the study of this thesis is for implementing naïve Bayes method to classify the fingerprint image based on the extraction of digital imagery.

The objectives to be achieved in this thesis research is

1. Can be classifying digital image of a fingerprint by fingerprint texture
2. To determine the level of accuracy.

2. CONTENTS RESEARCH

2.1 Artificial Intelligence

Artificial intelligence is a branch of science which deals with the use of machines to solve complex problems in a more humane way. This is usually done by following / copying the characteristics and analogy to think of human intelligence, and apply it as an algorithm known by the computer.

Artificial intelligence can be used to analyze the image of the scenery in the calculation of the symbols that represent the content of the scene after the image is processed to obtain characteristic. Artificial intelligence could be seen as three integrated entity that is the perception, understanding and action. Perception translate signals from the real world in the image become symbols of a more simple, the notion of manipulating the symbols was to facilitate extracting certain information, and acts to translate the symbols that have been manipulated into other signals that can be the end result or outcome between appropriate with the purpose of (Ahmad U, 2005).

2.2 Texture

Generally texture refers to the repetition of elements of basic textures are often called primitive or texel (texture element). A texel is composed of several pixels with the rules of periodic position, kuasiperiodik, or random (U. Ahmad, 2005).

The terms of the formation of the texture of at least two, namely:

1. The existence of primitive patterns consisting of one or more pixels. The forms of these primitive patterns can be a point, a straight line, curved line, area and others which are the basic elements of a form.
2. Primitive patterns had appeared repeatedly at intervals of a certain distance and direction so it can be predicted or found characteristics of repetition.



Figure 1 Example texture of VisTex Database

An image of an interpretation different texture when viewed at different distances and angles. Humans look at the texture based on the description that is random, such as smooth, rough, regular, irregular, and so on. This is a description of improper and non-quantitative, so it needed a quantitative description (mathematical) to facilitate analysis (U. Ahmad, 2005).

2.3 Tekstur Analisys

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 materials (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 random two-dimensional field, the method is based on the frequency of the frequency space. Examples of statistical methods is the autocorrelation function, cooccurence Matrix, Fourier transform, frequency edge, run length and others. Structural engineering

related to developing the smallest part of an image. Examples of structural method is the fractal models. The method is based on the geometry of existing geometry in the texture elements. 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 analysis of the frequency of such transformation, and Gabor wavelet transform (U. Ahmad, 2005).

2.3.1 Method Run-lenth

Gray level run length matrix commonly abbreviated to GLRLM is one popular method to extract the texture in order to obtain statistical characteristics or attributes contained in texture to estimate the pixels that have the same degree of gray. Extraction texture with run-length method is done by making a series of value pairs (i, j) in each row of pixels. Keep in mind the purpose of the run-length itself is the number of pixels in sequence in a particular direction which has a degree of gray / value of the same intensity. If it is known a run-length matrix with matrix elements q (i, j | θ) where i is the degree of gray at each pixel, j is the value run-length, and θ is the orientation towards certain shifts are expressed in degrees. Orientation formed with a four-way shift at intervals of 450, 00, 450, 900, and 1350.

Based on research conducted by Galloway (1975), there are several types of textural characteristics that can be extracted from the run-length matrix. Here are the variables contained in the extraction of the image by using statistical methods Grey Level Run Length Matrix:

i = the value of degrees of gray
j = successive pixels (*run*)
M = The number of degrees of gray in an image
N = The number of pixels in an image sequence
r(j) = The number of pixels in sequence by many order (*run length)*
g(i) = The number of pixels in sequence based on the degree of grayed.
s = The amount of the total value of the resulting run in a certain direction
p(i,j) = The set of matrices i and j
n = The number of rows * number of columns.

Where the variable-the variable will be used to find the value of the texture attributes as follows:

1. *Short Run Emphasis* (SRE)

   SRE measuring the distribution of short-run. SRE is highly dependent on the number of short-run and is expected to be greater in fine texture.

$$SRE = \sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j)/s}{j2} = \sum_{j=1}^{n}\frac{r(j)/s}{j2}$$

2. *Long Run Emphasis* (LRE)

   LRE distribution measure long run. LRE is highly dependent on the number of long run and is expected to be large on a rough texture.

$$LRE = \sum_{i=1}^{M}\sum_{j=1}^{N} j2p(i,j)/s = \sum_{j=1}^{N} r(j)j2/s$$

3. *Grey Level Uniformity* (GLU)

   GLU measure the degree of gray value equation entire image and is expected to be small if the value of a similar degree of gray around the image.

$$GLU = \sum_{i=1}^{M}(\sum_{j=1}^{N} p(i,j))2/s = \sum_{i=1}^{M}\frac{g(i)2}{s}$$

4. *Run Length Uniformity* (RLU)

   RLU equation measure the length of the run throughout the image and is expected to be small if a similar run length across the image.

$$RLU = \sum_{j=1}^{N}(\sum_{I=1}^{M} p(i,j))2/s = \sum_{J=1}^{N}\frac{r(j)2}{s}$$

5. *Run Percentage* (RPC)

   RPC run measure of togetherness and distribution of an image in a particular direction. RPC-value is greatest when the run length is 1 for all degrees of gray in a particular direction.

$$RP = \sum_{i=1}^{M}\sum_{j=1}^{N} p(i,j)/n = \sum_{j=1}^{N} r(j)/n$$

2.4 Classification

Classification is a job that assessment of an object's data to fit in a particular class of a number of classes available. (Prasad, 2012).

There are two main jobs:

1. Construction of a model as a prototype to be stored as a memory.

2. Using the model to perform recognition / classification / projection on an object other data entry in which the class.

2.4.1 Classification naïve bayes

Naïve Bayes is a simple probabilistic-based prediction technique that is based on the application of Bayes theorem (Prasad, 2012)

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

Strong independence of the features is that a feature on a Data has nothing to do with the presence or absence of other features that are in the same data

Naïve Bayes is an simplest method of classification by using existing opportunities, where it is assumed that every variable X is free (independence). Because the assumptions are not mutually dependent variable, then obtained the equation (7).

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

The data used can be categorical or continuous. However, in this final project will use continuous data, as a result of the image feature extraction is a continuous data measurement results in the form of figures on the level of contrast, homogeneity, entropy, energy, and dissimilarity in feature extraction. Therefore For continuous data can be completed using the following steps.

Training :

Calculate the average (mean) of each feature in the training dataset by the equation

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

Where :

μ = mean

n = The number of data

$\sum x_i$ = The number of data Values

Then calculate the variance of the training dataset blood money as in equatio.

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

Where:

σ^2= varians

μ= mean

x_i = value data

n = The number of data

Testing :

1. Calculate the probability (Prior) for each class that is by counting the number of data for each class divided by the total number of overall data.
2. Next calculate the probability density using equation (10). Expressing the relative probability density function. Data with mean μ and standard deviation σ, the probability density function is:

$$\varphi_{\mu\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where :

x = input of data

π = 3,14

σ = standard deviation

μ = mean

3. Having obtained the probability density value, then calculate the posterior of each class using the equation

$$Posterior = \frac{P(C_i)x\,P(X1 \dots \dots \dots \dots \dots \dots \dots Xn|C_i)}{P(X1 \dots \dots \dots \dots \dots \dots \dots Xn|C_i)}$$

or can be write

$$posterior = \frac{prior\ x\ likehood}{evidnce}$$

4. The greatest posterior is the corresponding class.

2.5 Method of testing accuracy

In this study, plan testing is done with a few scenarios, here is a scenario that has been prepared :

1. Test image used as training data.
2. Examine the image that does not include training data.
3. Test the effect of the amount of training data on the accuracy and time.

1. Test image used as training data.

This test method uses the data already in training, in other words, the training data and test data is the same data.

2. Supplied set test

This test method uses different data, in other words, the training data different from the data that will be tested

3. The influence of the amount of training data for accuracy

The 3 scenario testing conducted to test the effect of the amount of training data for accuracy, this test is done by changing the ratio between training data and test data.

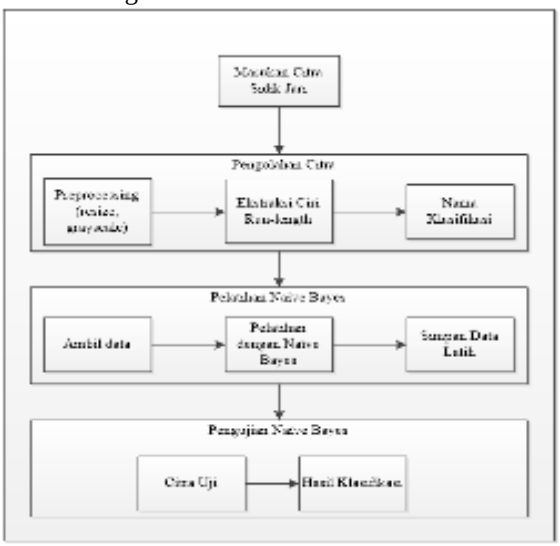
a) Comparison of test data The 3: 5
In this test done using training data comparison and test data The 3: 5 for the

data in each class, which means using The 3 images trained and 5 test images

b) Comparison of test data in 4: 5
In this test done using training data comparison and test data 4: 5 for the data in each class, which means using the image of train 4 and 5 test images.

2.6 Process Analysis

Analysis of the process to be conducted in this study is the analysis of the image based on texture classification. The stages of work processes within the classification from the input data to the output data. Here are the stages of the process of analysis to be performed and to the flow of process analysis can be seen in figure 2 :



Ficture 2 flow chart analysis process

The following by explanations of each stage is as follows :

1. Put the fingerprint image

An image data acquisition step on the media input into the system. Image input in the form of image files that contain objects fingerprints taken using the ink stamp affixed on the paper and then scanned.

2. Image processing

At this stage, the image will be resized entered into 32x32 pixels, in grayscale becomes grayish to obtain value matrix grayscale. Having obtained matrix grayscale feature extraction process is then performed run-length and give the name of the classification after the obtained value of the average feature-length run.

3. Training naïve bayes

Training on naive Bayes stage which takes the data that has been named classification and has an average value run-length feature to be calculated and the search for the mean and variance. Having obtained the mean and variance training data is saved to the database.

4. Testing naïve bayes

In the testing process naïve Bayes, the input image to be tested. Input image to be processed to find the probability density values and seek the greatest posterior value. Having obtained posterior greatest value then known classification results of the test images.

2.6.1 Analysis Data Input

In this study, the first to be done is the analysis of the input data. Analysis of the input data is done to obtain an input value that can later be used to process naïve Bayes classification methods. In this study, the data input is an image, which will be searchable content value using Run-length feature extraction matrix, the output value to be obtained is SRE, LRE, GLU, RLU, and RPC. These values will serve as baseline data which will be used as input in the naïve Bayes method. Stages that will be performed on the input data analysis is preprocessing is to resize and grayscale images. After performing preprocessing, feature extraction is carried out with a run-lrngth matrix method to get the value of the image feature.

2.6.1.1 Preprocessing

In the present study, preprocessing done to make it easier to get the value of feature extraction. Preprocessing will do is resize, grayscale and image quantization. The following by process flow of preprocessing :

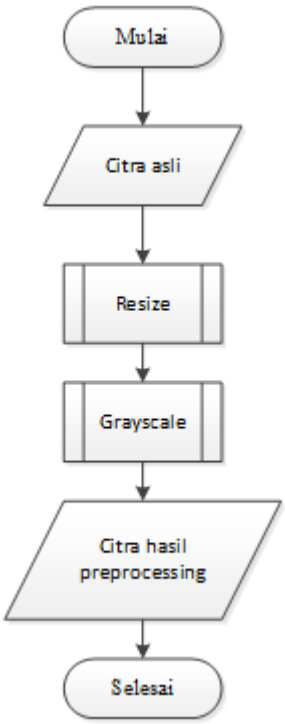


Figure 4. Alur preprocessing

2.6.1.1.1 Resize

Resize is the first stage of the preprocessing. Resize done to accelerate and facilitate the process of calculation. The following by process flow of resizeing :

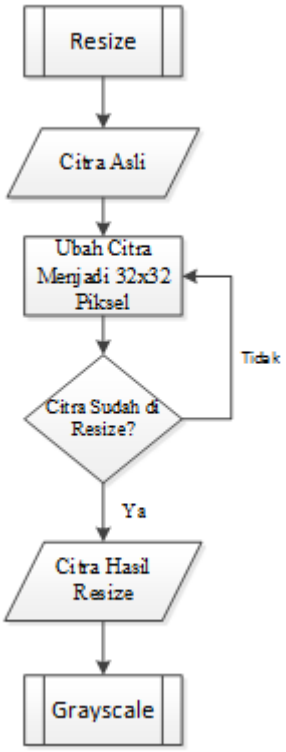


Figure 3 process flow resize

2.6.1.1.2 Grayscale

Grayscale is a process to change the color to greyish. By changing the RGB value of each pixel of the image into the same value so that each pixel has the same value for all three elements of color and grayscale matrix values obtained. The following by the flow of the process grayscale:

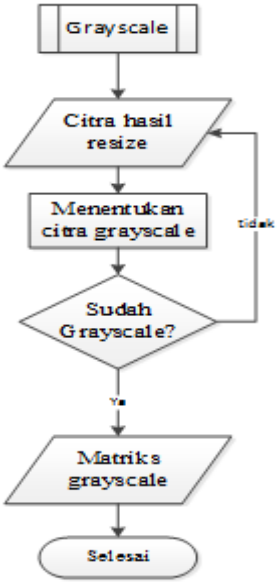


Figure 4 greyscale process flow

2.6.1.2 Feature extraction

Feature extraction is a process for mendaptkan main characteristics contained in the image, which has been in the grayscale image will produce a grayscale matrix ynag been in quantitation, matrix is exactly what will be used at this stage. This stage will calculate the 5 statistical values of run length that SRE, LRE, GLU, RLU and RPC with symmetry angles of 0, 45, 90 and 135. Having obtained all the value of the angle will be averaged. The following by process flow of extract the characteristics:

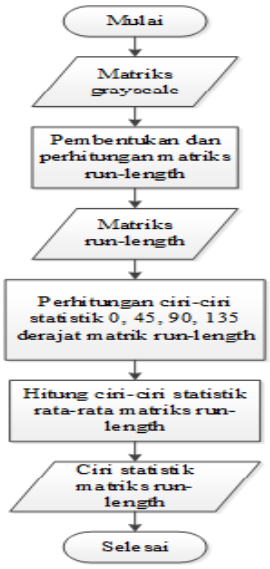


Figure 5 process flow feature extraction

2.6.1.3 Name of Classification

Name of classification used to determine the classification of names that have obtained an average statistical value from run-length feature extraction. The following by process flow of the classification name:

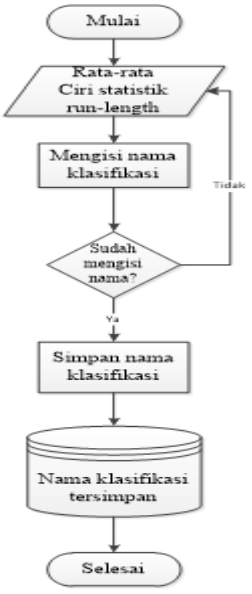


Figure 6 process flow classification name

2.6.2 Analysis of Training Naïve Bayes

Training naïve Bayes made to obtain training data, the process is carried out into 3 stages: data capture, training data and store data. These stages are described as follows:

2.6.2.1 Take Data

Retrieve data functions as retrieval of data that have been stored previously. The following by process flow of data capture:



Figure 7 process data capture

2.6.2.2 The Training naïve bayes

The training process with naïve Bayes is done by calculating the mean and variance. Here is the training process flow with naïve Bayes:



Ficture 8 flow process training

2.6.3 Pengujian Naïve Bayes



Figure 8 testing naïve bayes

2.7 The Testing

The aim of this algorithm testing to determine the performance of the run-length method and naïve Bayes for image classification. Tests on this study was to measure the level of accuracy and time dibutuhkaan for image classification process with several scenarios have been prepared, to measure the degree of accuracy using the confusion matrix.

Class		Classification results		Persentase
		A	B	
Target Owner	Finger print Adit	4	01	80%
	Finger Print Ali	0	5	100%

Table 1 Confusion matrix

Use a formula to calculate accuracy :

$$\frac{5}{5} X\ 100 = 100\%$$

2.7.1 Testing Plan

In this study, plan testing is done with a few scenarios, here is a scenario that has been prepared:

1. Test image used as training data.
2. Examine the image that does not include training data.

3. Test the effect of the amount of training data on the accuracy and time.

2.7.1.1 Testing Scenario 1

Scenario 1 was conducted to determine the accuracy in identifying patterns that have trained him. If naïve Bayes can identify each image is perfect, then Naive Bayes function properly. In the test scenario 1 could be concluded that the test image classification based on the texture using image data that has been trained to have an average of 99.88% accuracy rate. Here are the results of testing scenario 1

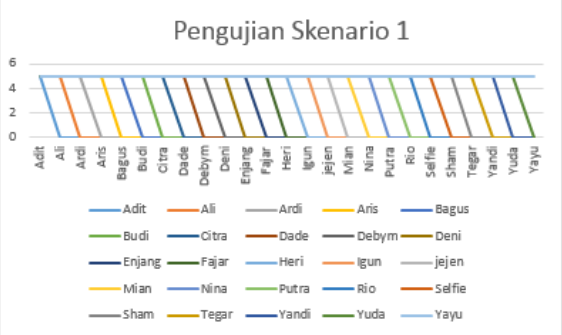


Figure 9 Testing scenario 1

2.7.1.2 Testing scenario 2

Testing of scenario 2 was conducted by examining the image that are not included in the training data, this test aims to determine level of image recognition beyond the training data. In the second scenario testing can be concluded that the image has not been trained for the average accuracy rate is 78.63%. Here are the results of testing scenario 2

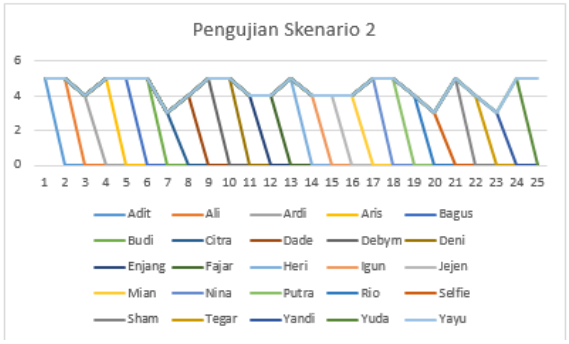


Figure 9 testing scenario 2

2.7.1.3 Testing Scenario 3

The 3 scenario testing conducted to test the effect of the amount of training data terhadapa accuracy, this test is done by changing the ratio between training data and test data. Comparison of training data and test data will be used is the ratio of 3: 5 and 4: 5.

2.7.1.3.1 Effect of test data The 3: 5

In this test done using training data comparison and test data The 3: 5 for the data in each class, which means using The 3 images trained and 5 test images. In the test scenario The 3: 5 could be concluded that the level of accuracy using The 3 training data is 57.07%, Here are the results of testing ratio of 3: 5

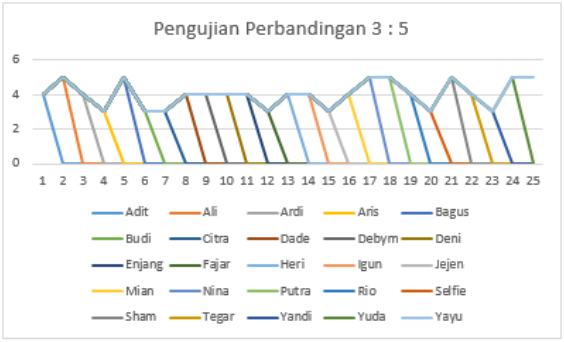


Figure 10 testing the effect of training data The 3: 5

2.7.1.3.2 Effect o the Test data The 4:5

In this test done using training data comparison and test data 4: 5 for the data in each class, which means using the image of train 4 and 5 test images. In the test scenario 4: 5 can be concluded that the level of accuracy using the 4 training data is 62.45%, Here are the test results a ratio of 4: 5

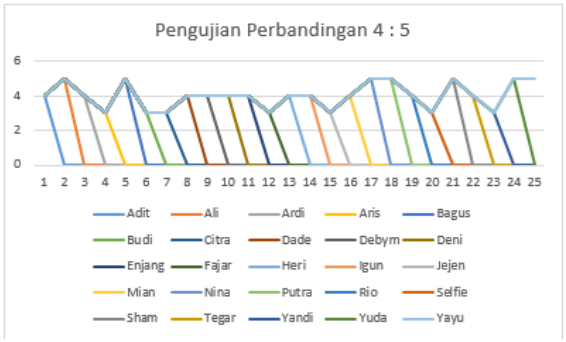


Figure 11 testing the effect of training data 4: 5

2.8 Conclusion Testing

Based From the testing that has been done so didapatkanlah conclusion that naïve Bayes method can classifying the input image with statistical data directly comparing the closest distance to the training, testing of image classification based on the texture using image data that has been trained to have an average accuracy rate of 99.88% and the image has not been trained for an average of 78.63% accuracy rate and level of accuracy using the The 3 training data is 57.07% and the use of 4 training data is 62.45%. The more training data the greater the degree of accuracy and vice versa.

3. CLOSING

3.1 Conclusion

From the research, the analysis of image processing, training and testing methods of run-length

and naïve Bayes for classification based on the texture image is obtained the following conclusion:

1. Naïve Bayes algorithm can classify digital image of a fingerprint based digital image extracted run-length method
2. Naïve Bayes algorithm generates 99.8% accuracy rate.

3.2 Suggestion

From the research, analysis, image processing, training and testing there were suggestions that it might be beneficial if there were going to conduct similar research, namely:

1. From these results, it is advisable to develop this research towards identifying the fingerprint image data belonging to the police were more complex, so it can assist law enforcement authorities in uncovering various crimes.
2. To be able to classify the input image should use a lot of training data.

DAFTAR PUSTAKA

1. Ahmad, U., 2005. *Pengolahan Citra Digital & Teknik Pemrogramanya.* Yogyakarta: Graha Ilmu.
2. Achmad, S., 2010. MySql dari Pemula Hingga Mahir. Jakarta: Achmatim.
3. Asbaugh, R., 1999. Ridgeology. *Journal of Forensic Identification*, Royal Canadian Mounted Police, Canada.
4. Eko, S., Nataliani Y., 2009. Klasifikasi Sidik Jari dengan Menggunakan Metode Wavelet Symlet. *Jurnal Informatika Volume 5 No 2*, Universitas Kristen Satya Wacana, Diponegoro.
5. Galloway, M., "Texture analysis using gray level run length", Computer Graphics Image Process., vol. 4, pp.172-179, juni 1975.
6. Hartanto, B., 2008. *Memahami Visual C#.Net Secara Mudah.* Pertama penyunt. Yogyakarta: Andi.
7. Mita, I., Santoso, I., Christyono, Y., 2007. Analisis Tekstur Menggunakan Metode Run Length. *Tugas Akhir Teknik Elektro*, Universitas Diponegoro, Semarang.
8. Munir, R., 2002. Pengolahan Citra Digital. Bandung: Informatika.
9. Nugroho, A., 2005. Pemodelan Berorientasi Objek, Bandung: Informatika.
10. Pressman, S. R., 2010. Rekayasa Perangkat Lunak. Yogyakarta: Andi.
11. Prasetyo, E., 2012. Pengenalan Pola Naive Bayes, Universitas Pembangunan Nasional, Jawa Timur.
12. Sri, K., 2009. Klasifikasi Status Gizi Menggunakan Naïve Bayesian Clasification. *Tugas Akhir Teknik Informatika*, Universitas Islam Indonesia, Yogyakarta.
13. Anik, A., 2013. Sistem Pedukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus:AMIK BSI YOGYAKARTA, Amik BSI Jakarta, Jakarta Selatan.
14. Sutopo, A. H & Masya, F., 2005. *Pemograman Berorientasi Objek dengan Java*. Yogyakarta: Graha Ilmu.
15. Wiji, L., 2014. Deteksi Sidik Jari Seseorang Dengan Menggunakan Mathematical Morphology dan Edge Detection. *Jurnal Sistem Informasi*, STMIK Duta Bangsa, Surakarta.