RESEARCH ARTICLE OPEN ACCESS

Handwriting Identification By Using Neuro Fuzzy Methods Based On Features Extraction

Aeri Rachmad
Faculty of Engineering – University of Trunojoyo Madura, Indonesia
Corresponding author: Aeri Rachmad

ABSTRACT
Handwriting recognition system is a system to recognize one's writing through paper. This technology identifies a unique and fixed piece of writing like human handwriting. The character pattern recognition on human handwriting words utilizes image processing and analysis of characters which will play an important role in the handwriting recognition process. The system will search for characters and then insert into a pre-prepared reference database through the training process. In this research it will be made recognition system by utilizing human handwriting. In the initial process it will be data retrieval in the form of words with the size of 500 x 500 pixels and segmentation of each character of 24 x 20 pixels which will then perform feature extraction using Principal Component Analysis (PCA) to determine the characteristics of the characters. After that the results will be done by recognizing Using Adaptive Neuro Fuzzy Inference System to measure the similarity between training data and test data. From the trial application using Neuro Fuzzy classification it obtained accurate recognition accuracy of 65.37%, in the scenario with 3 training data. While 7 training data, it obtained accurate recognition accuracy of 80%.

Keywords **-** Handwriting recognition, feature extraction, Principal Component Analysis, Neuro Fuzzy

I. INTRODUCTION

The introduction of handwriting pattern has been done by using artificial neural network method. This method is able to classify or select an input data into a predetermined category that has been defined by using a standard handwriting database [1]. In the handwriting data retrieval, there are the difficulty of large size, form of writing that is not standard, and inconsistent [2]. Handwriting recognition can be done in real time by considering the effectiveness and speed of image readings [3]. Some tools have been developed using handwriting detection such as digital pen, PDA, computer hardware, Smart phone. The equipment allows the user to use the hand as a stationery [4].

In this research the characters recognition pattern is a capital word that will be segmented into parts by character using PCA as extraction feature and Neuro Fuzzy for recognition. The processes undertaken in character pattern recognition are edge detection of images, Imagery image segmentation, feature extraction using PCA and identification of characters using Neuro Fuzzy.

II. RELATED WORKS

The first study was the application of neural and neuro fuzzy neural networks for fingerprint pattern recognition by Priyo Bayu Santoso. In this study it describes the imaging that recording fingerscanner in the form of digital images which will apply the process of histogram formation of the image TSB. Its application is a comparison of neural network method with neuro fuzzy method. The process of introduction of fingerprint input through ANN can be done well that is with 100% accuracy level. [5]

The second research is Multi-Face Detection on static image by using Principle Component Analysis by Hyun-Chul Cho and Se-Young Oh. In Performance for test images, it is 88% detection rate for test images. When all images are used to make eigen-face, detection rate performance reaches 97%. Time detection of less than 200 ms is used to search for full-scale, whille 90 ms for a limited-scale stops when a local minimum appears. In this case, a multi-scale algorithm and multi-face detection are recommended using PCA. This algorithm can precisely find the vertical face of the area on static images in a reasonable time, and the various sizes of face detection can be limited as needed. For invariant systems of rotation and illumination, algorithm find face rotation settings using neural networks or other methods and reduce fixed invariant lighting to work in the future. [6]

The third study was Time Delay Neural Network For Printed and Cursive Handwritten Character Recognition trliti by Guyon Isabelle Locust. In this study, it explains about handwriting recognition by using neural network. This network has been trained to recognize either a digit or a

capital character with a modified version of the backpropagation algorithm. The set training is included 12,000 examples which produced by a large number of different authors. The error rate is 3.4% of 2,500 text examples from a separated unit of authors. When it is allowed to reject 7.2%, the system makes 0.7% error replacement. Recognizer has been applied to an AT & T 6386 PC with an AT & T touch terminal device. This system has speed in reading characters up to 1.5 characters per second. Preprocessing is only 2% responsible for this time.[7]

The fourth research is Offline Handwriting Recognition using Genetic Algorithm by Rahu kala, Harsh vazirani, Anupam shukla and Ritu tiwari. In this paper the authors propose the use of genetic algorithms and graph theory to solve the problem of offline handwriting recognition. The author provides input in the form of images. The algorithm is trained on the Data training that originally existed in the database. Training data consists of at least two sets of training data per character in the language. The author uses graph theory and geometric coordinates to convert images to graphs. The author notices that this conversion changed the whole issue of handwriting recognition for graph matching problems. When a pure graph matching is done, good results are obtained. The algorithm is known to recognize the given character as input. But efficiency increases drastically when we apply the genetic algorithm. This algorithm helps in the optimization of both force and distance optimization. In style optimization, it helps us to mix two different styles to produce new ones that are in between. This is done by taking the mean coordinates of the knot of parents. We see how it helps in the identification of the character M. In this research, they got an efficiency of 98.44%, which proves that this algorithm works for most cases and correctly matches the known inputs for their characters.[8]

The fifth study was the Principal Component Analysis in Image Processing by M. Mudrov'a, A. Proch'azka. In this paper, it describes the properties of PCA which can be used for the determination of the selected object orientation or rotation as well. Various methods of image segmentation by object definition (such as thresholding, edge detection or other) should be used initially. Binary images contain objects or borders of black (or white) pixels on the background of the inverse results of this process. This paper presented and handled with two applications from PCA in image processing. Other applications in this area can be learned as well. The ROI will be focused on PCA using methods for processing biomedical signals and images. Further attention will be paid to the Independent Component Analysis method associated with PCA as well.[9]

III. METODOLOGY

Principal Component Analysis

The PCA method is part of a character recognition project that can be used on a dimensional X data (m * n). It is assumed that the PCA is formed from a single character, but in general it will be easier to understand if the character has been projected on a vector. PCA will calculate the main components of a collection of characters that enter in the training phase (training character). The main components which obtained PCA can also reconstruct and recognize the characters to be input. This main component is the characteristic values that produce a new model which is called the characteristics of the character (eigen). In the PCA method a character, it is also an image which can be viewed as a vector [9]. If the width and height of the image are m and n pixels, many components in the image are m * n. Each pixel is encoded by one vector component. The formation of this vector from an image is done by placing each line of the image next to another line which is commonly referred to as lexicographical ordering. The PCA algorithm is used in the process as follows which is based on the average overall object of each.

ALGORITHM[6]**:**

1. Input data vector:

$$X = [x_{11}, x_{12}, ..., x_{1n}, x_{21}, x_{22}, ..., x_{2n}, ... , x_{m1}, x_{m2}, ..., x_{mn}] \quad (1)$$

2. Calculate the average data vector (μ) based on the overall average of objects of each character.
3. The data vectors are subtracted by means of average to obtain centralized data:

$$Y = X - \mu \quad (2)$$

4. Count the covariance matrix: $\Omega = AA^T$ (3)
5. Find Eigenvalues $(V = [v_1 | v_2 | ... | v_p])$ and Eigenvectors ($(\Lambda = [\lambda_1 > \lambda_2 > ... > \lambda_p])$. (4)
6. Selection of optimal eigenvectors based on the largest eigen value.

$$Y = [y_{11}, y_{12}, ..., y_{1n}, y_{21}, y_{22}, ..., y_{2n}, ... , y_{m1}, y_{m2}, ..., y_{mn}] \quad (5)$$

Adaptive Neuro Fuzzy Inference System (ANFIZ)

According to Jang Anfis in his work using a hybrid learning algorithm, using the method of Least-Squares Estimator (LSE) is done in the 4th layer. In the 4th layer, the parameters are linear parameters to the system outputs that make up the fuzzy rule rule.[5]

In Figure 1, ANFIS architecture on one input system and one output is described as follows:

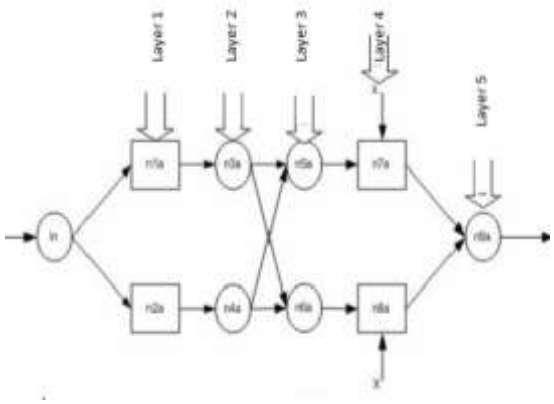

Figure 1. Architechtur ANFIZ

Layer 1, The processes in this layer run the fuzzyfication process by using the Bell membership function. Each node in this layer is an adaptive node with a node function :

n1a = Bell (x; a1, b1, c1) (6)
n2a = Bell (x; a2, b2, c2)

The function x is the input for n1a and n2a nodes, whereas a1, b1, c1, a2, b2, c2 are bell membership function parameters. The bell function used is expressed by the following equation:

$$\mu A\,(x) = \frac{1}{1+\left|\frac{x-c_i}{a_i}\right|^{2b}} \quad (7)$$

The parameters in this layer are called the premise parameters when ai, bi, ci are the set of parameters. In layer 2, each node in this layer is labeled with n3a and n4a which are nonadaptive (fixed parameters) that forward the result of layer 1. Since the system used is only one input, there is no AND inference mechanism. Thus the output of the 2nd layer is :

$$n3a = n1a \quad (8)$$
$$n4a = n2a$$

Each node output states the degree of activation of the fuzzy rule. In general some T-norm operators that can reveal AND fuzzy logic which can be used as node functions in this layer.
In layer 3, each node in this layer is labeled with n5a and n6a which are also non-adaptive. Each vertex displays the degree of activation which is normalized by shape.

$$n5a = \frac{n3a}{n3a+n4a} \quad (9)$$
$$n6a = \frac{n4a}{n3a+n4a}$$

In layer 4, each node in this layer is an adaptive node, and this layer obtained the matrix A, as follows:

$$A = \begin{bmatrix} (n5a)x_{(i)} & (n5a) & (n6a)x_{(i)} & (n6a) \\ \vdots & \vdots & \vdots & \vdots \\ (n5a)x_{(n)} & (n5a) & (n6a)x_{(n} & (n6a) \end{bmatrix} \quad (10)$$

The number of rows of the matrix A is the sum of the input data x. This layer sought consequential parameter value by using LSE method.
The equation for LSE method is stated as follows:

$$\theta = inv\,(A^T A)A^T.y \quad (11)$$

y = output or desired target

$$\theta = [p1\ q1\ p2\ q2]^T \quad (12)$$

Furthermore, the following equations are used to calculate the output from the 4th layer:

$$n7a = n5a\,(p1x + q1) \quad (13)$$
$$n8a = n6a\,(p2x + q2)$$

In layer 5, the single node in this layer is labeled with n9a which calculates all outputs as the sum of all incoming signals:

$$n9a = n7a + n8a \quad (14)$$

After that the network output may result in learning output of each pixel.

IV. RESULT AND DISCUSSION

In Figure 2, the trial will be done with the process of inserting the test image in the process of grayscale. Grayscale is making the truecolor image of 24 bits to 8 bits. The second process is a binary process that makes the test image from 8 bits to 1 bit. In binary images, each point is 0 or 1 which each point is presenting a certain color. The third process is the histogram, the histogram is the thresholding process. In the statistics field, the histogram is the graphical display of the frequency tabs which is represented by graphics as the manifestation of binary data.

The fourth process is segmentation or separating the image apart that represents a particular area.

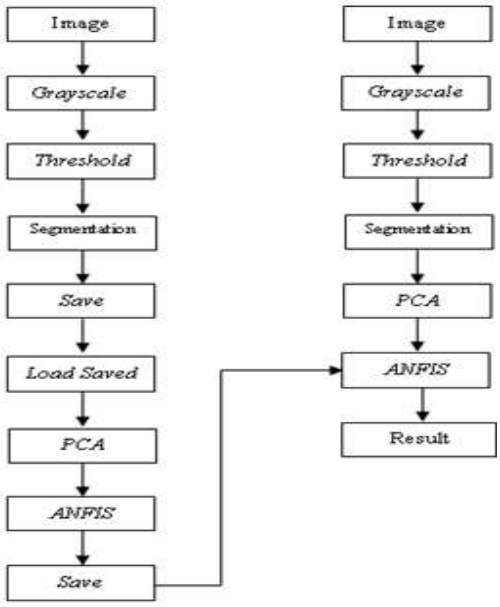

Figure 2. Design System

From this point, the calculation of the number of characters with PCA feature extraction is to set the eigen value and then it is done by image recognition with Neuro Fuzzy. The steps are done as much as 3 scenarios for the test data 1 so that the findings are obtained near the original value. These results will definitely have an error value. This error value can be known by calculating the percentage so that it can be known the level of success percentage which is shown in Figure 3.

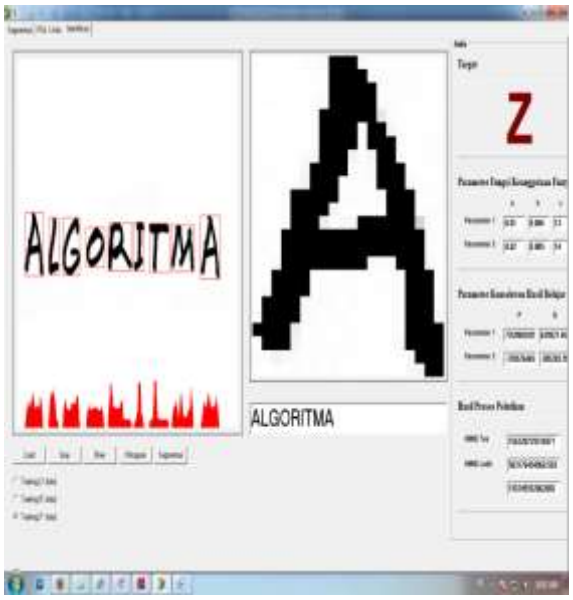

Figure 3. Simulation Using Delphi Program

For trials, there are 3 trial scenarios. The first scenario is 3 training data and 7 data testing. The second scenario is 5 training data and 5 data testing. And the third scenario is 7 and 3 data training data.

Table 1. 3 Trial Scenarios for Test Results

No	Characters	Scenario 1 (%)	Scenario 2 (%)	Scenario 3 (%)
1	A	100	100	100
2	B	42.8	80	100
3	C	42.8	60	66.7
4	D	57.2	80	100
5	E	57.2	100	100
6	F	42.8	40	66.7
7	G	100	100	100
8	H	0	40	33.3
9	I	0	20	33.3
10	J	85.7	100	100
11	K	71.4	80	100
12	L	0	100	100
13	M	100	100	100
14	N	42.8	20	66.7
15	O	85.7	100	100
16	P	85.7	100	100
17	Q	100	100	33.3
18	R	71.4	80	100
19	S	57.2	80	100
20	T	85.7	60	66.7
21	U	100	100	66.7
22	V	100	100	100
23	W	100	80	66.7
24	X	42.8	40	66.7
25	Y	85.7	100	100
26	Z	42.8	20	33.3

V. CONCLUSION

The character recognition pattern system on word handwriting which uses Neuro Fuzzy can be used to recognize image of characters with the best accuracy of 65,37% by using 3 training data with 7 data testing,; and while using 7 training data with 3 testing, it obtained best accuracy which is equal to 80% ,

REFERENCES

1. Chayaporn Kaensar. A Comparative Study on Handwriting Digit Recognition Classifier Using Neural Network, Support Vector Machine and K-Nearest Neighbor, The 9th International Conference on Computing and InformationTechnology, 2013, 155-163.
2. Guwahati, Assam, Bangla online handwriting recognition using recurrent neural network architecture**,** Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, 2016, 978-1-4503-4753-2, DOI: 10.1145/3009977.3010072

3. Rajat Aggarwal, Sirnam Swetha, Anoop M. Namboodiri, Jayanthi Sivaswamy, C. V. Jawahar, Online Handwriting Recognition using Depth Sensors, *Proceedings of the 13th IAPR International Conference on Document Analysis and Recognition, Nancy, France, 2015.*
4. Jafaar Alabodi and Xue L, An Effective Approach To Offline Arabic Handwriting Recognition, International Journal of Artificial Intelligence & Applications (IJAIA), 4(6), 2013 , DOI : 10.5121/ijaia.2013.4601
5. Bayu Santo Priyo, Neural Network Application and Neuro Fuzzy for Fingerprint Pattern Introduction. Teknik Komputer Stikom. Surabaya. 2007.
6. Cho Hyun-Chul dan Oh Se-Young, Multi-Face Detection on static image using Principle Component Analysis. Department of Electronic and Electrical Engineering, Pohang University of Science and Technology, Pohang, Korea, 2004.
7. Guyon Et.Al.. Time Delay Neural Network for printed and Cursive Handwritten Character Recognition. IEEE Trans. On Acoustics, Speech and Signal Proccessing, vol. 19, No 54, april 1992.
8. Kala Rahu, vazirani Harsh, shukla Anupam, tiwari Ritu. Offline Handwriting Recognition using Genetic Algorithm. Journal of Computer Science, 1 (March) : 1-7
9. Mudrov´a M, Proch´azka A. 2008. Principal Component Analysis in Image Processing. Institute of Chemical Technology. Prague. Czech Republic.

International Journal of Engineering Research and Applications (IJERA) is **UGC approved** Journal with Sl. No. 4525, Journal no. 47088. Indexed in Cross Ref, Index Copernicus (ICV 80.82), NASA, Ads, Researcher Id Thomson Reuters, DOAJ.

Aeri Rachmad. "Handwriting Identification By Using Neuro Fuzzy Methods Based On Features Extraction ." International Journal of Engineering Research and Applications (IJERA), vol. 7, no. 8, 2017, pp. 80–84.