Palm Leaf Manuscript Segmentation and Reading by Using Artificial Neural Networks

Thanannath Phattharachairawee[1], Montean Rattanasiriwongwut[2], and Mahasak Ketcham[3]
Faculty of Information Technology,
King Mongkut's University of Technology North Bangkok, Thailand
[1]william.thanannath@gmail.com
[2]montean@it.kmutnb.ac.th
[3]mahasak.k@it.kmutnb.ac.th

Abstract **- Palm leaf manuscript is considered as a kind of cultural heritage and the record of local wisdom of ancestors that should be transformed into digital format for educational and research benefits of the next generation. This research presents palm leaf manuscript's Isan Dhamma characters segmentation and reading conducted by using image processing. The objective of this research is to utilize the obtained data in sentence recognition process further. The input was digital photos of a palm leaf manuscript written with Isan Dhamma characters that was proposed to be adjusted on its quality by adjusting light intensity through histogram. Subsequently, its quality was improved by using median filter in order to screen data on enhancement or attenuation of some picture's properties in order to gain quality as demanded. Subsequently, characters were segmented from background (segmentation) by using Global Thresholding. In the next process, Principal Components Analysis (PCA) was used in order to match each sentence with its features. In the last procedure, obtained features were compared with database in order to find range of such sentence by using the principles of neural network before comparing with support vector machine in order to compare such sentence with database. Based on the experiment conducted with 10 images of palm leaf manuscript, it was found that neural network gave better results than support vector machine by 86%.**

Keywords **- Artificial Neural Networks, Principal Components Analysis (PCA), Threshold, Morphology, Neural Network, Support Vector Machine**

I. INTRODUCTION

Digital Age is originated from technological advancement including computer technology, communication technology, and information technology. Consequently, the growth rate of information has been changed rapidly. The format of information would be in message, image, and sound formats and such information has been collected in digital formats. Consequently, it is convenient for retrieval and application as demanded. However, there are some types of information that are collected in document format without transforming to digital format therefore it is difficult for retrieval and practical utilization. Consequently, knowledge on image processing is adjusted in order to adjust document to be digital image and operated with Optical Character Recognition (OCR) for transforming digital image to be digital document. To transform digital images to be in electronic document format, it is necessary to utilize knowledge on image processing and Optical Character Recognition (OCR).

Thai character recognition can be applied to several businesses, for example, news reporters are able to record all interviews on paper or students are able to summarize their study on notebooks. Subsequently, they input such

paper into OPCR system and the system will transform characters to be in the form of character process program. Consequently, it is not necessary for news reporters and students to type all characters repetitively. This helps to reduce and accelerate work process. Thai handwriting character recognition program can be applied to many devices, for example, electronic whiteboard or PDA for online operation, for example, when writing on electronic whiteboard, program will be able to be transformed in the form of character processing program leading to higher operational efficiency.

Generally, the principles of Optical Character Recognition are different upon the target languages because each language has distinctive format. For recognition technology, Dhamma characters have been developed to support recognition of various sizes and fonts of characters in order to improve accuracy on recognition from Thai language directly, especially in the form of similar Dhamma characters. Researches related to Dhamma character recognition proposed the learning methods based on various characters, for example, Back-Propagation Neuron Network (BPN), Time-Delay Neuron Network (TDNN), or Fuzzy Logic, etc.

II. LITERATURE REVIEWS

In this paper, we proposed the methods for Segmentation and Reading by Using Artificial Neural Networks.

The major procedure related to character recognition is object segmentation. According to literature survey, it was found that there was a number of methodology used for segmentation. Khankasikam et al, [1-2] proposed histogram technique in order to sharpen pixels. Wantanable et al, [3-4] proposed different histogram technique by selecting different numbers of grey scale. This methodology utilized data related to neighboring pixels or the edge of the existing image in order to adjust for the benefits of Thresholding and another method was histogram administration. Di Gesu [5], [3] proposed the use of considered light intensity and spatial data for data input in order to improve non-linear operational efficiency [6-7]. Moreover, there was some effort to use fuzzy logic [8] and artificial neural networks [9] for character recognition. W. et al, [10] mentioned Automatic License Plate Recognition (ALPR), i.e., extraction of data on vehicles from images or image orders of extracted data can be used with or without database on inspection of venous system for traffic monitoring. ALPR uses either black and white or infared camera for using the image. Ntirogiannis et al, [11] studied that Binarization of document image is very important in analyzing and recognizing document image because it affects to the following procedures of recognition process. Evaluation of Binarization also helps to study on its algorithm behavior as same as to inspect its efficiency by providing qualitative and quantitative identification of performance. This paper focused on pixels evaluated by using Binarization for handwriting / Historical Document Printer in the form of evaluation proposing retrieval and accuracy. Evaluation measures were amended correctly by using lower weighing that may lead to bias against evaluation. Chattopadhyay et al [12] conducted a complex operation in video. OCR system was proposed to be operated on embedded platform. The novelty of proposed method was the use of processing circuit and low memory unit to provide correct recognition by 84.23% that was higher than OCR video. The correctness was normally accepted. Malakar et al. [13] explained that extraction of message line is one of the important procedures of accepted Optical Character (OCR) system. In the event of document image written by handwriting, appearance of line twisting or overlapping (s) making this process to be challenging process of current research techniques extracted from 87.09% and 89.35%, respectively. Sankaran et al [14] proposed a project that was accepted on India's Devanagiri script. Correctness of Devanagiri script recognition is unable to be compared with Roman script due to complexity of script writing style. We solved this problem by using well-known artificial neural networks as the Bi-directional Long Short-Term Memory (BLSTM). Sankaran et al [14] reported that error rate was reduced over than 20% and

characteristic of error rate was lower than 9% compared with the best OCR system.

This paper is divided into 5 parts. The second part is explanation on standards and procedures / methods used on paper for enhancing efficiency and segmenting characters while the third part is explanation on preparation of data presentation. The forth part is explanation on recognition methods while the fifth part is explanation on results and conclusion.

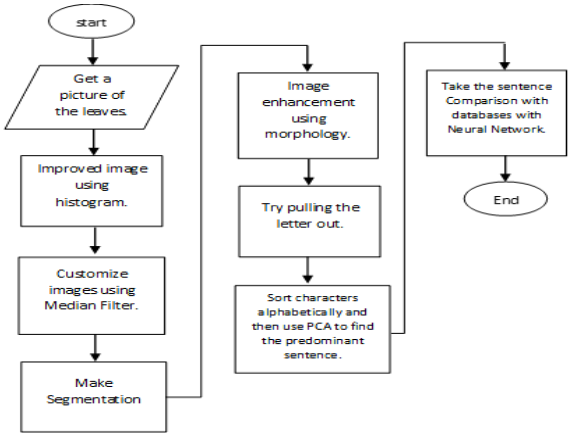

Fig. 1 The Proposed Method

III. LEADERSHIP STYLES

Segmentation and Reading by Using Artificial Neural Networks.

A. Preparation of Image Data

This research presented the method for collecting data obtained from stored documents in the form of image file of 50 images of the characters used as the sentences and its file name extension was Joint Photographic Experts Group (JPEG) with the sizes ranged from 320x240 to 2048x1536 pixels with white characters and black background.

Fig. 2 Sample Images Used in Testing

B. Median Filter [15-16]

Firstly, images are enhanced by using median filter through spatial convolution with the weight of 1 thoroughly. After conducting spatial convolution, the outcome is averaged by selecting mean as shown in the equation below:

$$g(i,j) = med[x(i+r, j+s) \in A(i,j) \in Z^2] \quad (1)$$

Whereas, median filter is a kind of non-linear filtering and the advantage of image enhancement through median filter is image blurring giving the same tone of pixel's intensity.

C. Global Thresholding Technique [17]

In this procedure, background was segmented from foreground by Thresholding with the intensity between both thresholds of Histogram whereas the obtained Threshold was ranged between 0-255 only. Threshold was used for comparing values of each pixel. If f(x,y) is less than Threshold, such pixel will be adjusted as black or object part. If f(x,y) is higher or equal to Threshold, such pixel will be adjusted as white or background that can be written as the following equation:

$$f_{thr}(x,y) = \begin{cases} 1, f_f(x,y) < Threshold \\ 0, f_f(x,y) \geq Threshold \end{cases} \quad (2)$$

Whereas,

1 is black that is the part of object
0 is white that is the part of background

D. Morphology [18-19]

In the following procedure, noises were removed by using image corrosion technique in order to decrease the area of white color or object in the image. The processing procedures were the same as those of image expansion but there was difference based on the principles stated that all positions with factors moving would be compared with images.

$$A \Theta B = \{z \mid (B)_2 \subseteq A\} \quad (3)$$

Whereas,

A is the image requiring corrosion of image border;

B is structural elements;

Z is pixel field

Fig. 3 Image after Enhancing with Morphology

E. Extraction of Features of Each Sentence

In this paper, Principal Components Analysis (PCA) was used for extracting features of each sentence in order to be used for recognizing sentence. This will help to recognize the scope of word. To analyze on the major elements, it is necessary to prepare images for analyzing and face recognizing. The procedures are as follows:

- **Instructional Procedures**

1. Determine face of I(x,y) as Array 2 Dimension, N x N images. Moreover, it may be considered as vector of N^2 dimension.

2. Build new body with the size of 2-D image from N x N pixels, it will be 1-D vector in N^2 dimension.

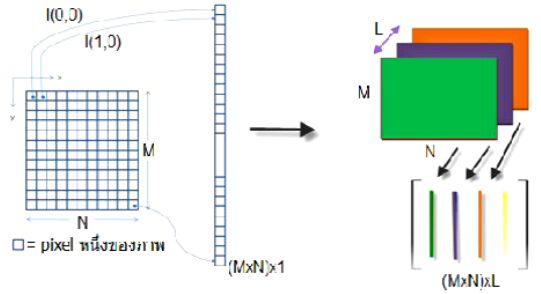

Fig. 4 Results of Vector Transformation from 2D to be 1D

3. Check vector of mean of deviation from mean and matrix of co-variance for instructional collection, especially, images in instructional images will be represented by $\{T_1, T_2, \ldots, T_M\}$ whereas each T_M is vector of N^2 dimension with M value as the number of images in instructional collection. The example of mean vector is as follows:

$$\varphi = \frac{1}{M}\sum_{n=1}^{M} T_n \tag{4}$$

4. The collection of vectors was deviated from mean $\{\Theta_1, \Theta_2, \Theta_3, \ldots, \Theta_M\}$ with difference of each individual of instructional images from each mean vector.

$$\emptyset_i = T_i - \varphi \tag{5}$$

5. Check eigenvectors of the matrix of variance of instructional collection .The matrix of co-variance will be obtained.

$$C = AA^T \tag{6}$$

Whereas,

$A = [\Theta_1, \Theta_2, \Theta_3, \ldots, , \Theta_M]$

6. Check eigenfaces eigenvectors as v_2 of such $A^T A$ whether:

$$A^T A v_i = \mu_i v_i \tag{7}$$

- **Test Procedures**

7. Transform images for testing as eigenfaces based on the aforementioned method. Subsequently, those images will be minus with the mean of images in order to find differences of images. The obtained differences are multiply with Eigenvector.

8. Calculate distance between tested faces and image groups used in learning. Subsequently, find the distance with proportion that is similar with image group used in learning by finding the shortest distance.

$$d(Y_1, Y_2) = \sum_{k=1}^{d} \left\| y_k^1 - y_k^2 \right\|^2 \tag{8}$$

Whereas,

$\|y^1_k - y^2_k\|^2$ is Euclidean distance between two feature vectors.

F. Recognition

- **Support Vector Machine [20-21]**

The method of Support Vector Machine is classification with Supervised Learning that is classified as the technique used for solving the problem on data format recognition based on the principles of calculation of coefficient of equation in order to create Optimal Separating Hyper Plane.

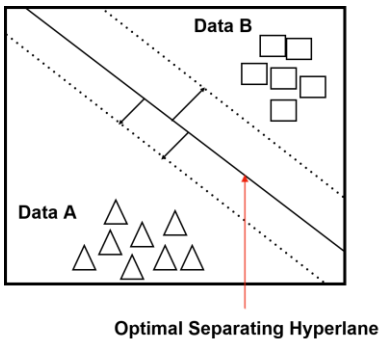

Fig. 5 Optimal Separating Hyper Plane from SVM

$$f(x) = sign(w.x + b) \quad (9)$$

Whereas,

w means weighing vector of recognition obtained from weighing process.

B means bias for recognition system.

X means Feature Vector.

Sign means +1 when value is higher than 0 and -1 when value is lower than 0.

- **Artificial Neural Networks [22-23]**

Learning of Perceptron network is in the manner of layered structure and learning data will be input in the first layer that is Input Layer. After calculating the first layer, it will be sent to the Hidden Layer whereas each unit of this layer will accept data from all units in the former layers for calculating and sending to the next layer. When data is sent to the last layer or Output Layer, the input will be given by the system. This kind of data transmission is considered as feed forward. Subsequently, the output will be checked by the system how it is incorrect based on the target. In this research, we utilized instructional algorithm for using with artificial neural networks based on Backpropagation technique. This technique is the use of layered structure of Supervised Learning with demanded target and adaptive networks were used for adjusting the weight properly as shown in Fig. 6.

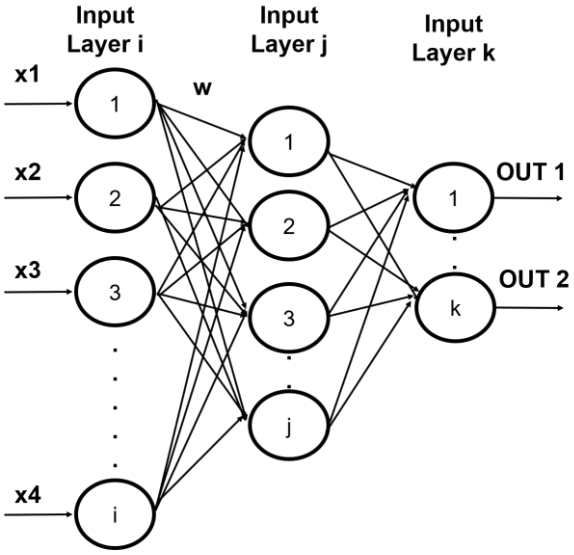

Fig. 6 Layered Artificial Neural Networks Used in Backpropagation Instruction K-Nearest Neighborhood

K-nearest neighborhood (K-NN) is considered as the good learning towards nonparametric and parametric segmentation and it is also very efficient when using with much teaching data. In this research, this appropriateness was use. Segmentation based on K-nearest neighborhood (K-NN) is considered as the good option. It is necessary for nearest neighborhood method to determine k parameter whereas k is the similar feature. There are 3 procedures of K-nearest neighborhood (K-NN) as follows:

Procedure 1: Choose K gene that is nearest to gene with Missing Values (MV) in order to estimate Missing Value X_{ij} of i^{th} gene in the j^{th} sample. Gene which is Expression Vector will be selected with expression more than that of sample.

Procedure 2: Calculate to find the distance between data of 2 Expression Vector X_i and X_j with Euclidian distance covering elements in the j^{th} sample. Euclidian distance between X_i and X_j can be calculated by:

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - x_{j,k})^2} \quad (10)$$

Whereas,

$d(X_i, X_j)$ the distance between X_i sample and X_j sample.

N is the total properties of the sample.

$X_{i,k}$ is k property of X_i sample.

$X_{i,k}$ is k property of X_j sample.

Procedure 3: Select data value with the smallest amount of dist in k amount for considering and finding the answer.

- **Naïve Bayes**

$P(x|w_{face})$ and $P(x|_{wnon\ face})$ is the class with pdfs condition of character and non-character, respectively. X Feature Vector is classified as the character format.

$$L(x) = \log(\rho(x|\omega_{face}) - \log(\rho(x|\omega_{nonface}))) \quad (11)$$

In our guidelines, PDF file with format condition will be estimated by using Naïve Bayes Model that is treated to be operated well practically. This Naïve Bayes Model estimated the statistics of independence among elements of x Feature Vector. Under this hypothesis, we obtain:

$$\rho(x|\omega_{face}) = \prod_{i=1}^{N} \rho(x|\omega_{face}) \quad (12)$$

$$\rho(x|\omega_{nonface}) = \prod_{i=1}^{N} \rho(x|\omega_{nonface})$$

$P(x|w_{face})$ and $P(x|_{Wnonface})$ calculated by using histogram technique.

- **Bayesian Method**

In this article, we presented the measurement of similarity and probability based on the belief of Bayesian stated that differences of light intensity of images were presented by $\Delta = I_1 - I_2$ that was the characteristic of general formats in the appearance of each person. Especially, we determined two types of face format: specific format Ω_I and special individual format Ω_E. Our similarity measurement was represented in the light of probability.

$$S(I_1, I_2) = P(\Delta \epsilon \Omega_I) = P(\Omega_I|\Delta) \quad (13)$$

Whereas,

$P(\Omega_I|\Delta)$ is the following probability defined by Bayes Rule by using probability estimation $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$. These probabilities were obtained from training data by specifying efficient empty space for data with high dimension.

- **Adaboost**

The first classifier of Adaboost segmentation is consisted of week classifier e.g., linear classifier, (each classifier classified only one data dimension of input vector. As a result, it is otherwise called week classifier. The results of segmentation are as follows:

$$H(x) = sign\left\{\sum_{t=1}^{T} \beta_t h_t(x)\right\} \quad (14)$$

Whereas,

X represents Input Vector; h_t)x(, t =1, …, T, referring to the number of classification as T; β_t, t =1, …, T refers to weight of each weak classifier.

- **Template Matching Techniques**

Template matching techniques are able to be used for finding the pixel level and the scope of script in template image. There are several template matching techniques that are used in image processing for finding the location of searched images. In this paper, cross relation template matching technique was used with the stimulation of distance measurement.

$$d_{f,t}^2(u, v) = \sum_{x,y} [f(x, y) - t(x - u, y - v)]^2 \quad (15)$$

Whereas,

f is input image and it is the image of template matching.

- **Classification and Regression Tree (CART)**

CART was preferred by Breiman, Freidman, Olshen, & Stone in 1984. It is the method for rescurive partition creating classification and regression of trees for predicting continuously dependent variables (regression) and variables (classification). Generally, CART is able to be adjusted for two types of regression problems that was considered as the effort to predict continuous variables from continuous predictor or variables classification. The next problem was that classification of problems of variables depended on types of variables, for example, member group, etc., from one continuous variable or over.

Since these variables of classification were considered in this research, the research problems were classified as the problems of classification. The major objective of this analysis is to determine the condition set of logic classification that was permitted to be correct. There were 4 procedures of classification as follows:

Procedure 1: Specify criteria for accuracy of prediction;

Procedure 2: Select splits

Procedure 3: Determine the stop period of classification

Procedure 4: Select the tree with the "correct size"

IV. RESULTS

This research developed the program by using Matlab and 50 sample images of palm leaf manuscript's Isan Dhamma characters consisted of 550 characters and 120 digits therefore the total characters were 670. Table 1 represents the result on accuracy compared between SVM method and artificial neural networks.

Method	Accuracy Rate
Support vector machine	74%
Neural Network	86%
Decision Tree	65%
K-NN	70%
Naive Bayes	67%
Adaboost Cacade	60%
Bayesian	56%
Template Matching	72%

V. CONCLUSION

The proposed guidelines on palm leaf manuscript's Isan Dhamma characters segmentation and reading was started from image enhancement by using histogram and median filter. Subsequently, the obtained images were threshed by segmenting images in order to enable each image segment to use suitable threshold. In the following procedure, Principal Components Analysis (PCA) was used for finding the features of each sentence. Finally, obtained features were compared with database in order to find the range of such sentence by using neural network principles before comparing with support vector machine in order to compare such sentence with database. In addition, image vector was also replaced by image matrix in order to calculate and find intergroup distribution and internal distribution. Based on the experiment, it was found that this new method had the highest efficiency on calculation.

REFERENCES

(Arranged in the order of citation in the same fashion as the case of Footnotes.)

[1] Fu, K.S. and Mui, J.K. (1981). "A Survey on Image segmentation". Pattern Recognition, Vol. 13, p. 3-16.

[2] Kalas, M.S. (2010). "An Artificial Neural Network for Detection of Biological Early Brain Cancer". International Journal of Computer Applications, Vol. 1(6), p. 17-23.

[3] Storn, R. and Price, K. (1997). "Differential Evolution: A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces". Journal of Global Optimization, p. 341-359.

[4] IP, H.H.S. and Chan, S.L. (1997). "Hypertext-Assisted Video Indexing and Content-based Retrieval". ACM 0-89791-866-5, p. 232-23.3.

[5] Pal, N.R. and Pal, S.K. (1993). "A Review on Image Segmentation Techniques". Pattern Recognition, Vol. 26(9), p. 1277-1294.

[6] Weszka, J.S., Nagel, R.N., and Rosenfeld, A. (1974). "A Thresholdselection technique". IEEE Trans, Computer C-23, p. 1322-1326.

[7] Kao, Y.T. and Zahara, E. (2008). "A hybrid genetical goriqhm and particle swarm optimization for multimodal functions". Applied Soft Computing, Vol. 8(2), p. 849-857.

[8] Fonseca, J.M. (2011). "Optical character recognition using automatically generated Fuzzy classifiers". in Fuzzy Systems and Knowledge Discovery (FSKD), Vol. 1, p. 448-452.

[9] Mani, N. (1996). "An optical character recognition using artificial neural network". in IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, p. 2244-2247.

[10] Badawy, W. (2012). "Automatic License Plate Recognition (ALPR): A State of the Art Review". p. 1-1.

[11] Ntirogiannis, K., Gatos, B., and Pratikakis, I. (2013). "A Performance Evaluation Methodology for Historical Document Image Binarization". p. 1-1.

[12] Chattopadhyay, T., Ruchika, J., and Chaudhuri, B.B. (2012). "A novel low complexity TV video OCR system". Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE.

[13] Malakar, S. (2012). "Text line extraction from handwritten document pages using sparal run length smearing algorithm". Communications, Devices and Intelligent Systems (CODIS), 2012 International Conference on, IEEE.

[14] Naveen, S. and Jawahar, C.V. (2012). "Recognition of printed Devanagari text using BLSTM Neural Network". Pattern Recognition (ICPR), 2012 21st International Conference on, IEEE.

[15] Mohammed, J.R. (2008). "An Improved Median Filter Based on Efficient Noise Detection for High Quality Image Restoration". IEEE Int. Conf, p. 327-331.

[16] Gonzalez and Woods. (2002). "Digital image processing (2nd Ed.)". Prentice hall.

[17] Deravi, F. and Pal, S.K. (1983). "Gray level thresholding using second order statistics". Pattern Recognit, Zett. 1, p. 417-422.

[18] Pahsa, A. (2006). "Morphological Image Processing with Fuzzy Logic". Havacilik Ve Uzay Teknolojileri Dergisi, Ocak, Cilt 2, Sayi 3, p. 27-34.

[19] Wei-Cheng, W.L. (2000). "Mathematical Morphology and Its Applications on Image Segmentation". Dept. of Computer Science and Information Engineering, National Taiwan University, Jund 7.

[20] Taira, H. and Haruno, M. (2000). "Feature Selection in SVM Text Categorization". Transactions of Information Processing Society of Japan.

[21] Joachims. (1998). "Text categorization with support vector machines: Learning with many relevant features". Proc. of the 10th European Conf. on Machine Learning, p. 137-142.

[22] Poggio, T. and Girosi, F. (1990). "Networks for approximation and learning". Proc. IEEE, Vol. 78, p. 1481-1497.

[23] Haykin, S. (1994). "Neural Networks". A Comprehensive Foundation, New York: Macmillan.

[24] Khakham, P. and Chumuang, N. (2015). "Isan Dhamma Handwritten Characters Recognition system by using Functional Trees Classifier". In 11th IEEE International Conference on Signal-Image Technology & Internet-Based System, Mahasak Ketcham, p. 606-612.

[25] Chumuang, N. and Khakham, P. (2015). "The Intelligence Algorithm for character recognition on palm leaf manuscript". In Far East Journal of Mathematical Sciences (FJMS), Mahasak Ketcham, p. 333-345.