Syllabus Line Segmentation from Palm Leaf Manuscripts using Vector Neural Network

Nwe Nwe Soe
University of Computer Studies (Thaton), Myanmar
nwenwesoe.mlm@gmail.com

Received: 30 September 2017
Accepted: 4 January 2018

Keywords:
Vector Neural Network (VNN), Syllabus Segmentation, Palm Leaf Manuscripts, q-valued dimensions

Abstract: This paper offers the methods of finding the syllabus line segmentations from palm leaf manuscripts. This system can be supportive for enhancing palm leaf manuscripts taken by high quality resolution digital camera. The Vector Neural Network is that object location can be determined by clustering points of interest and hierarchically forming candidate of palm leaf manuscript regions according to similarity and spatial proximity predicates. This system can be used to optimize two factors: RGB background colors and a number of vertical lines or horizontal lines to choose candidate area. Moreover, Vector Neural Network performs better accuracy and requires less calculation time than other traditional methods. This system can be applied to segmentation of candidate area which includes text. The results of the research can be used as an input image to implement an OCR system to provide information of being existence for their related fields. Neural network ensemble techniques have been shown to be very accurate classification techniques. However, in some real-life applications, a number of classifiers required to achieve a reasonable accuracy are enormously large and hence very space consuming. This paper introduces special neural method, Vector Neural Network (VNN), which has great associative memory and high performance. In this system, VNN is analyzed using various size of database that have randomly created patterns, noise levels, and fixed q-dimensions. The result shows that it has capacity much greater than conventional Neural Networks.

1. INTRODUCTION

Palm leaf manuscripts have rarely attracted research in recent times in Myanmar and received extensive attention in academic and production fields (Leavline & Sutha, 2011). It is an important area in image processing and pattern recognition. Document Image Processing is one of the key application areas of image processing. Several research works have been focusing toward evolving newer techniques and methods that would reduce the processing time while providing higher recognition accuracy. Themain cause of stains on the surface of old manuscripts is fungal growth and wound area (Forcadel, Guyader, & Gout, 2008). Most of the stains, smears and spots are bluish, greenish, blackish and brownish in color. Moreover, dust particles and insect remnants cause stains on the palm leaf manuscripts Improper and inadequate manual handling also adds to stains. Sometimes the thread used for tying the folios together and cloth used to protect palm leaves also leave stray marks that contribute to stains (Alahakoon, 2006).

U Pho Thee Library is a famous Myanmar ancient manuscripts library and situated in Thaton.

This library has the leaves over 200,000 palm leaf manuscripts, packed of over 788 Parabaik manuscripts and ancient books. These scrolls of religious books are always taken care by Thuwana Bume Buddhist Society [8]. Being Myanmar citizens, we must try our best to preserve these Myanmar valuable palm leaf manuscripts. The fungi on the surface of palm leaf manuscripts can be cleaned by mixing lemon grass oil with carbon. This step is verification and is used for eliminating unwanted things over palm leaf manuscripts. After having cleaned, palm leaf manuscripts are recorded by using a high resolution digital camera (Moni & Raju, 2011).

In the area of traditional Medical Affair, Mathematics, Literature, Myanmar culture life style etc (Miljković, 2009) and many researchers refer the manuscripts for their research works. In Myanmar, Pali-text Society at U Phoe Thee Library can be regarded as the only place where valuable palm leaf manuscripts are to be stored in the digitized version and this task is really necessary. It is, now, essential to preserve and save valuable treasures of our nation in digital format files or Web sites. Segmentation is the method of splitting the image into text lines, words and then into characters which are particularly useful for classification. Segmenting character from palm manuscript is extremely challenging, while the characters structure and content differ significantly. The correctness of the OCR system depends on the segmentation. If the characters are segmented correctly, then the recognition system gives best results. Regions or objects are divided from an image in segmentation phase. Mainly segmentation, tries to extract essential component of the script, which are certainly characters. Furthermore, some complex handwriting scripts comprise different styles for writing words (Surinta, Schomaker, & Wiering, 2013). The analysis of VNN (Soe & Htay, 2017) has shown that at the present parametric vector models of the associative memory are the best both with regard to the storage capacity and noise immunity.

2. PROBLEM ISSUES AND RELATED WORKS

Subjective domain is texture image, viewed as random plain, which characterizes some structural and brightness properties. Objective domain are image processing methods, statistical analysis of texture structural and brightness, Neural Network methods of object recognition, processing and image recognition system (Zareizadeh, Hasanzadeh, & Baghersalimi, 2013; Miljković, 2009; Widiarti, Harjoko, Marsono, & Hartati, 2014).

Research targets in this paper are texture images surface having random characteristics of structural and brightness properties. Example images are palm leaf manuscripts, taken by high resolution camera. After segmentation, each sub image has the different characteristics in the shape of objects and their color inflation. Subjected fields are image processing methods, statistical analysis of image brightness and structural characteristics of texture, object recognition methods, image processing and classification program development. An analytical system will be carried out to check the accuracy of q-dimensional Vector Neural Network (Lee, 2010). Myanmar character pattern will be introduced in this Section. Myanmar script consists of (33) basics characters and (12) basic vowels, respectively. Basic Myanmar alphabet and vowels are shown in Fig. 1.

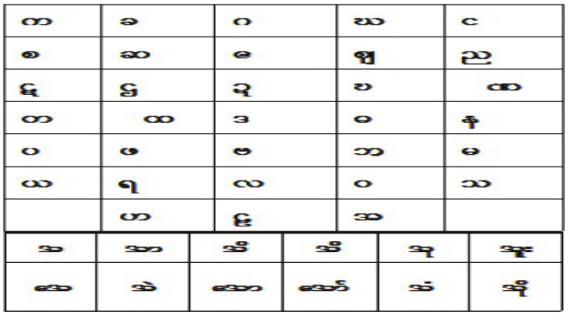


Figure 1. Basic Myanmar character and Basic Myanmar vowel

In edge detection step, the text area are detected by combining own created one-dimensional segmentation method and horizontal or vertical edge lines detection method. These unwanted text areas are cropped automatically by using vertical lines edge detections method. This system finds ROI and then automatically splits into new two palm leaf manuscript image files from original palm leaf manuscript image file by using one-dimensional segmentation methods (Project to digitize Myanmar manuscripts, 2018; Moni & Raju, 2011; Chacko & Babu, 2011; Heath, Sarkar, Sanocki, & Bowyer, 1997). The example of palm leaf manuscript images for processing are shown in Fig. 2-4. The performance of subsequent steps in document image analysis heavily depends on the performance of ROI and automatic data cropping which are the most important steps of the palm leaf manuscript segmentation. An efficient method for increasing the generalization capacity of edge lines detection is more efficient than other traditional edge methods segmentation by using one-dimensional edge lines segmentation mask. The methods are applied to palm manuscripts photos of black, green and blue background colors. As a sample, black background image of palm leaf manuscript is illustrated in Fig. 2-4.

Figure 2 Original photo of two palm leaf manuscripts (File type: jpeg, File size: 4.48MB).

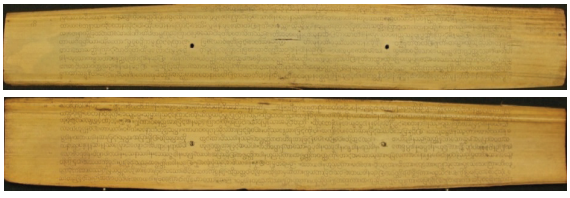


Figure 3 Finding (ROI) regions and splitting into two new better-quality palm leaf manuscript image files (File type: jpeg, File size: 270 KB).

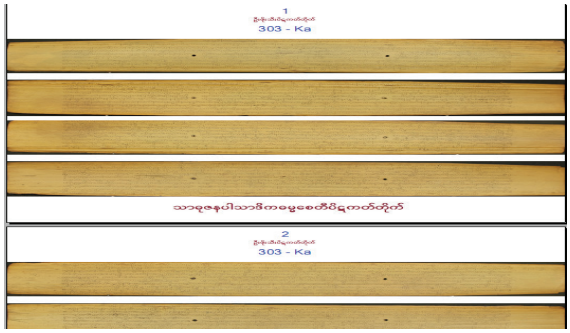


Figure 4 Pack of palm leaf manuscripts saved into PDF format files after cropping.

The objective is to improve the quality of the documents before they are used by a recognition system for subsequent information extraction. In the religious field of Myanmar country, the Buddhist laws are typed and edited to reference books. Especially, Myanmar antique Medical , Commentaries on Buddhist Pali texts, Poem of epic proportions, Prosody, Kind of four standard verse, Compilation of learned discourses, Code of Laws ,Myanmar Mathematics and Cabalistic treatises of Palm leaf manuscripts are so difficult and they need to be saved. It will be more advantageous if those palm leaf manuscripts can be converted into digital format files.

2.1 Line Segmentation

The aim of the proposed paper is to help digitization process of these valuable palm leaf manuscripts so that they could be stored in other different places, such as Myanmar Nation Library, and Yangon University Library etc. As a sample, one cropped palm leaf manuscript is illustrated in Fig. 5.

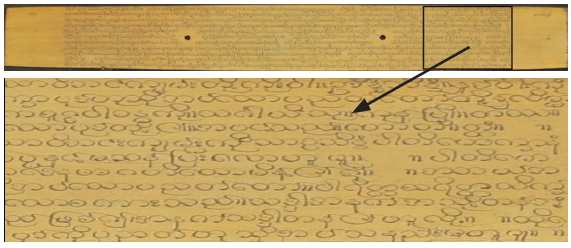


Figure 5 The cropped palm leaf manuscript image (3779 x 443) pixels.

In this paper three main tasks or experiments are considered for segmentation of palm leaf manuscript. These are:

(1) Defining which color intensity should be used for input palm leaf manuscript to get the best quality binary images.

(2) Segmentation of palm leaf manuscript to get segmented images which include hand writing character line.

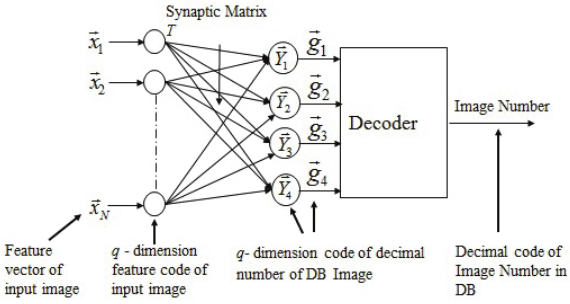
(3) Segmentation of line by line segmented images to get series of images.

2.2 Theoretical Background of the VNN

This section provides the main characteristics of the ensembles that will be analyzed in the present work. This paper focuses on a research on hand written character recognition from Palm leaf manuscript. After segmentation and object extraction, features must be extracted from segmented object. Unfortunately, for the time of being, real features cannot still be used. To test or analyze VNN, features are generated randomly and are stored them in Pattern Database. Let us consider a recognition problem of an input feature vector of a segmented image having N features. In pattern database, there will be M patterns $\{X_\mu\}$.

$$X_\mu = \{\vec{x}_{\mu 1}, \vec{x}_{\mu 2}, ..., \vec{x}_{\mu N}\}, \quad \mu = \overline{1,\ M}$$

This q-dimensional parametric vector neural network (VNN) can be considered as a two-layer neural network. Each neuron $\vec{x}_{\mu i}$ is coded to q-dimensional space before input to the network. All the input neuron is connected to each output neuron of PVNN. N-size output vector is considered to get the exact number of pattern which is the most similar to the input. Each $\vec{x}_\mu$ must be in $Y_\mu = (\vec{y}_{\mu 1}, \vec{y}_{\mu 2}, ..., \vec{y}_{\mu n})$ N is the output neuron count and in this paper the value of n is 4 which means maximum number of pattern is 10000 (from 0 to 9999).



After that, using pattern number in DB and coded features of that image, Synaptic matrix h_i is created. Once it is created, it will be used to classify the input pattern determining which image number in DB is the most similar to that input one. Output neuron number depends on the DB size (training pattern count). In fact, output neuron count is actually digit count of the total training images. In the above Fig. 6, the functions of the Vector Neural Network on the four output neurons are demonstrated. This type of Neural Network is very fast and suitable for pattern recognition process (Kryzhanovskii & Mikaelyn, 2003). Let us consider that the $Z = (\vec{z}_1, \vec{z}_2, ..., \vec{z}_N)$ is the input feature vector to the VNN. The $(q \times q)$ matrix T_{ij} described in above equation is simply multiplied by the input vector. This matrix affects the vector $z_j \in R^q$, , converting it in a linear combination of basis vectors. This combination is an analog of the packet of quasi-monochromatic pulses that come from the j^{th} neuron to the i^{th} one after transformation in the inter connection. After decoding the h_i, a decimal number is obtained. This number is the database pattern number which is most similar to the input one (Soe & Htay, 2017).

2.3 Creation of Database for q-valued dimensions with VNN

To analyze the VNN, pattern database is created with the randomly generated features. The created database is performed as shown in Fig. 7.

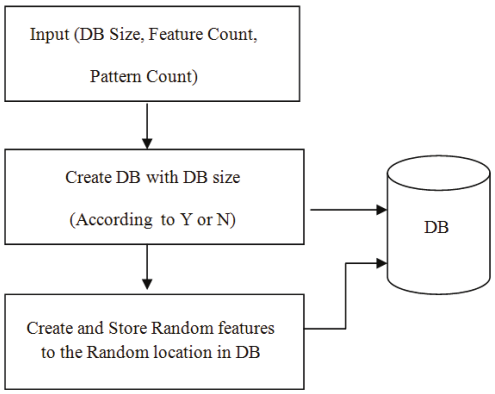


Figure 7 Creation of Database for Random features.

2.4 Performance analysis with VNN and DB

The following diagram presented in Fig. 8 shows the testing process for one selected input. In this program, the four main sub menus are included and they are Creation Pattern DB, Create T matrix, Test Each Pattern and Analysis by DB size. It is very clear that the speed of calculation (recognition) process is very high, because it needs only multiplication of two matrices and no need to consider the pattern database. The empirical results show that the VNN gets the high accuracy.

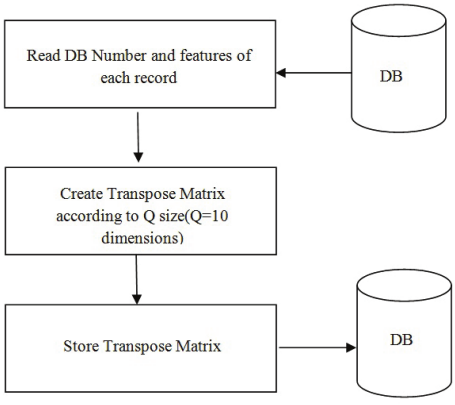


Figure 8 Creation Synaptic Matrix.

This paper is intended not to compare with other neural classifiers but to analyze its own parameters such as q-size, database size. Parameters are considered how to used, how much should be used, and to know the tolerance of noise level (Soe & Htay, 2017).

2.4 Testing Recognition power

The diagram presented in Fig. 9 shows the testing process for one selected input. In this program, the four main sub menus included and they are creation pattern DB, create T matrix, Test each pattern and analysis by DB size. It is very clear that the speed of calculation (recognition) process is very high, because it needs only multiplication of two matrices and no need to consider the pattern database. The empirical results show that the VNN gets the high accuracy.

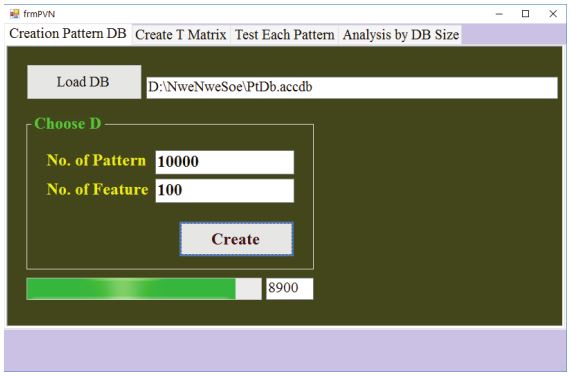


Figure 9 Creating pattern database.

To test the recognition power of the T matrix, one of the patterns in DB is used as an input one. Desired noise level is added to that input feature vector. And then it is necessary to read Synaptic matrix from the DB and just to multiply by the q-value coded input feature vector in order to get the output number. This system will easily retrieve the decimal number from DB. It can be checked whether this number is similar to input one or not and this process is illustrated in the following Fig. 10.

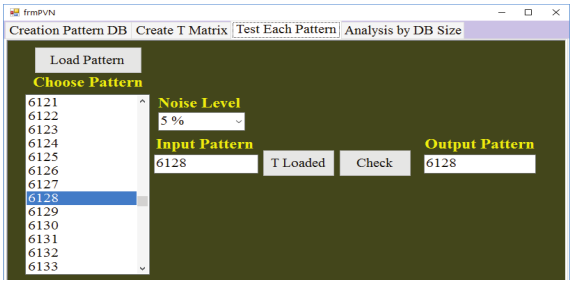


Figure 10 Testing level one by one noise level %.

2.5 Analysis by Pattern Database Size and Noise level

Accuracy mainly depends upon the noise level & database size, and pattern count. In this work, VNN classifier is analyzed in its robustness both with noise level and database size. In the first test, various noise levels (5%, 10%, 15% and 20%) are added to the input feature. The result with noise % level and Accuracy is shown in the following Fig. 11.

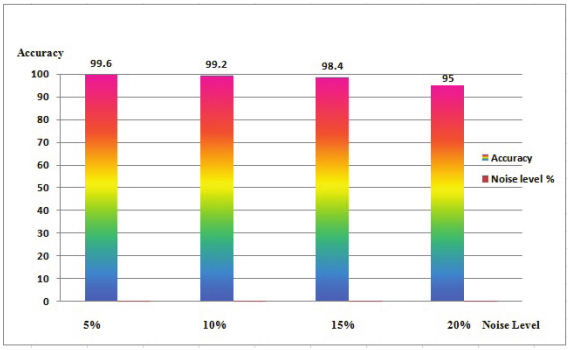


Figure 11 Comparison of VNN classifier with noise 5%, 10%, 15% and 20% and accuracies are 99.6, 99.2, and 98.4.

3. RESEARCH METHODOLOGY

Subjective domain is texture image, viewed as random plain, which characterizes some structural and brightness properties. Objective domains are image processing methods, statistical analysis of texture structural and brightness, neural network methods of object recognition, processing and image recognition system. Research target in this paper is texture images surface having random characteristics of structural and brightness properties. Example images are Palm Leaf Manuscripts, taken by using a high resolution camera. After segmentation, each sub image has the different characteristics in the shape of objects and their color inflation. Subjected fields are image processing methods, statistical analysis of image brightness and structural characteristics of texture, object recognition methods, and image processing. An analytical system will be used to check the accuracy of q-dimensional Vector Neural Network (Soe & Htay, 2017). In Table 1 lists results using both configurations of line segmentations of palm leaf manuscripts.

Table 1 Line segmentation rates of VNN

	Test set line segmentation rate %	
	All Reject patterns	**Half of Reject patterns**
Total Test Set	76.45	75.82
Reject Patterns only	78.34	75.37
Lines only	74.23	71.28

The accuracy of the methods, where the modest feature VNN classifier approach applied with corresponding savings in computation time and memory storage by using DB size and noise level %. In this paper a different type of neural classifier is introduced to enhance the performance analysis of q-dimensional VNN classifier. It just needs to create a Synaptic matrix once and after that just multiplication is needed to get result. Experimental results show that the proposed neural classifier can effectively achieve high accuracy and very fast. It can be used as the classifier for the noisy input because of good associated memory.

3.1 Instrumentation of Myanmar, Pali and Mon Languages

The Myanmar language is the official language and is more than one thousand years old. Myanmar script is considered a complex script by software developers, as it originated from India scripts like Thai or Khmer. Myanmar alphabet consists of 33 consonants, 12 vowels, 4 medals, 10 digits and a lot of Pali characters. In Myanmar (Burmese) writing system: syllabic alphabet each letter has an inherent vowel. The rounded appearance of letters is a result of the use of palm leaves as in the traditional writing

material. Straight lines would have torn the leaves. The Mon language is an Austroasiatic language spoken by the Mon people, who live in Myanmar and Thailand. Mon, like the related Khmer language but unlike most languages in mainland Southeast Asia, is not the Mon script, whose consonants belong to one of two registers: clear and breathy, each of which has different inherent vowels and pronunciations for the same set of diacritics. Whereas in Burmese, spellings with the same diacritics rhyme, in Mon, this depends on the consonant's inherent register. Mon alphabet consists of 35 characters.

3.2 Preprocessing for Palm Leaf Manuscripts

This Pre-processing is the name for operations on images at the lowest level of abstraction and the aim is an improvement of the image data that suppresses undesired distortions or enhances some image features important for further processing. It does not increase image information content. This method uses the considerable redundancy in images. Neighboring pixels corresponding to one object in real images have the same or similar brightness value. If a distorted pixel is removed from the image, it can be restored as an average value of neighboring pixels (Forcadel et al., 2008).

It is difficult to preprocess palm leaf manuscript system. Original condition of the manuscripts is not clean, so that images of digitations result in poor quality. Therefore, colors which can differentiate objects from backgrounds are sometimes unclear (Zareizadeh et al., 2013). Manuscript digitations process is imperfect. For example, due to low lighting, manuscript images are not obvious. In addition, manuscript positioning in digitization is not maximal because the manuscripts can't be opened widely, which makes digitations results tilted or look rolled. So, these input palm leaf manuscript images are taken from high quality digital camera.

Preprocessing step includes enhancing image quality. It has the following steps: Area Selection, Binarization, Removing Noise and Segmentation. The proposed paper detects the signal by replacing each intensity value with the average of three successive intensity values. Then, the derivative is estimated at position by averaging the two differences between the value at position and its left and right neighbors pixel RGB values. It is clear that the one-dimensional mask is illustrated in Fig. 5 and the better is the smoothing palm leaf manuscripts images. The 1st and 2nd derivative of edge detections are illustrated in Fig. 12.

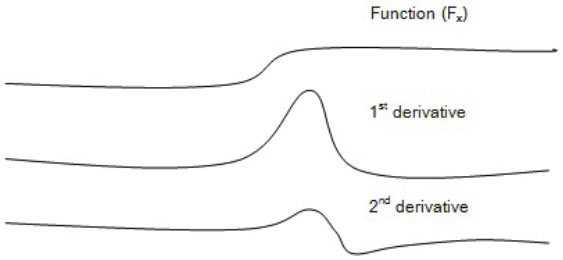


Figure 12 1st and 2nd derivative of an improved edge line detection.

4. RESULTS AND DISCUSSION

As for experiment results, palm leaf manuscript images were collected by U Pho Thee library in Thaton city, Mon state of Myanmar country. The proposed system has been implemented with Visual C# and Microsoft Access 2010 Database system. The palm leaf manuscripts image backgrounds are obtained by the Black, Green, Blue and others backgrounds in Fig. 13. Traditional edge detection methods such as Sobel, Canny and other related edge detectors needed to use more calculation steps and time. The proposed Syllabus segmentation of edge detection method is very simple and takes less time other edge segmentation methods.

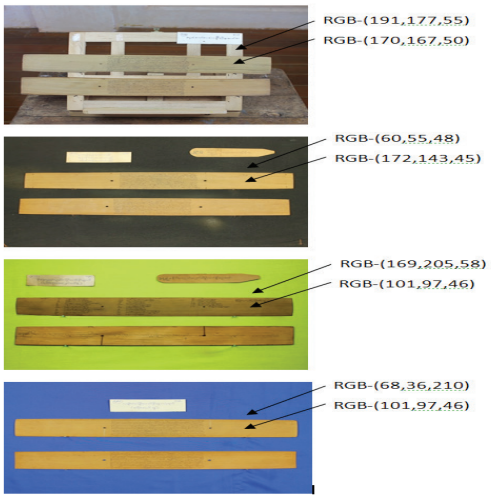


Figure 13 RGB color intensities values of normal, black, green and blue background images.

4.1 System Overview Syllabus segmentation of palm leaf manuscripts system

The system overview of the proposed character segmentation is performed as shown in Fig. 14. The user input the cropped images to the system and the system produces 4 binary images by using Otsu thresholding algorithm. The best quality binary image derived from Red intensity value array is selected to the next process; that is segmentation of line by line characters by searching optimal points which can be used to crop or extract sub images from the input one. These image arrays are the input elements to the character segmentation process which also produces image array, including line segmentations.

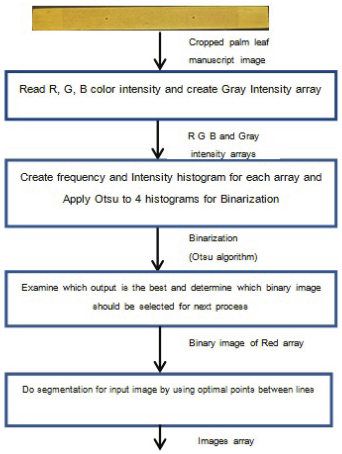


Figure 14 overview of line segmentation of palm leaf manuscripts system.

4.2 Implementation of the system

The graphical user interface design of the Syllabus segmentation of Palm Leaf Manuscript is shown in Fig. 9. The proposed system has been implemented with Visual C# and Microsoft Access 2010 Database system. Firstly, the user must load the original palm leaf manuscript image. The user needs to click the "Open" button and the following photo will appear as in Fig. 15.

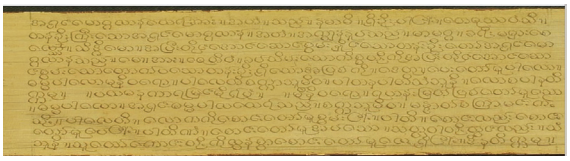


Figure 15 Original image of Palm Leaf Manuscript.

Then click the "Binarization button", and threshold value converted to output the binary image Black and White have values of 1 (white) for all pixels in the input image with Otsu threshold method. It is shown in Fig. 16.

Figure 16 Otsu threshold image of palm leaf manuscript.

4.3 Finding ROI of palm leaf manuscripts

The proposed methodology of determining a threshold in a gradient histogram is deduced rigorous analysis and it helps in achieving consistently appreciable edge detection performance in Fig. 17.

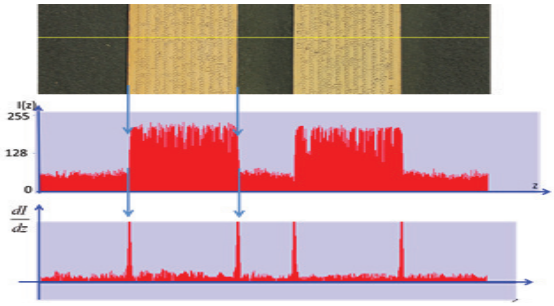


Figure 17 (a) A segment of Palm leaf image with a line (b) Intensity function along the line (c) differential value

This system, pre-processing step includes file, zoom and line segmentation. In this program, the user must open the binary image of the palm leaf manuscript. The example the source image is the left most side and in the middle Horizontal projection values of the binary image. In the right most side of the data grid view shows the pixel count and number of black pixel values of the binary image in Fig. 18.

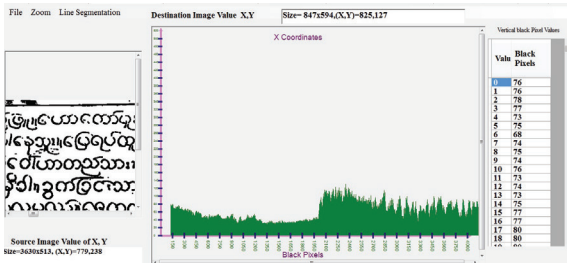


Figure 18 Vertical projection of palm leaf manuscript.

In Fig. 19 illustrates the syllabus segmentation and the vertical projection of the palm leaf manuscript of the binary image. In the data grid view shows the values of the black pixels according to horizontal projection.

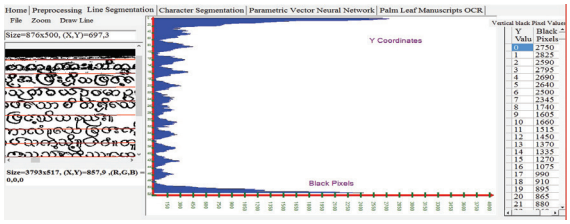


Figure 19 Horizontal projection of palm leaf manuscript.

Although this method uses information related to vowels and touching components of two consecutive lines to detect the lines, the technique does not consider the prolonged part of the characters. Incorrect estimation of vowels may occur if the lower vowel and upper vowel of two consecutive lines touches to each line. In addition, if the upper vowel of current line is closer to upper lines than the current line, or the lower vowel of current line is closer to lower line than the current line, line estimation may be defined to the wrong line as shown in Fig. 20.

Figure 20 Syllabus horizontal line segmentation of palm leaf manuscript with straight lines.

Therefore, a number of Text Line segmentation techniques have also been proposed to solve the problem. Line segmentation of Roman, Arabic, Indian, Chinese and Japanese handwritten has been published but there are few research reports on Myanmar handwritten palm leaf manuscript documents. The connected components were separated by tracing background skeleton and the correct rate of segmented characters had improved at 82%. After the combined method of segmentation for touching components from the second step was processed, the correct rate of segmented characters was 84.63%. After the contour tracing algorithm was applied and touching components from this step was separated by a trace of background skeleton and combined method of segmentation, the correct rate had increased 2.49%. This process could improve the syllabus horizontal line segmentation of palm leaf manuscript with irregular Lines the segmentation method for touching components with shown as Fig. 21.

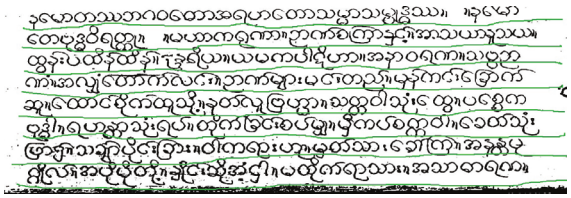


Figure 21 Syllabus horizontal line segmentation of palm leaf manuscript with irregular Lines.

4.4 Line Segmentation by using Optimal Points and Otsu Binarization

Otsu binarization of palm leaf manuscript is the important step for the next consequent pre-processes of character extraction work. In most of the image

processing system, only gray level intensity vales are used for processing images. But various type of palm leaf manuscripts are normally yellow in color and this paper work tries to do the experiment for binarization process. In Fig. 22, the four binary images are shown.

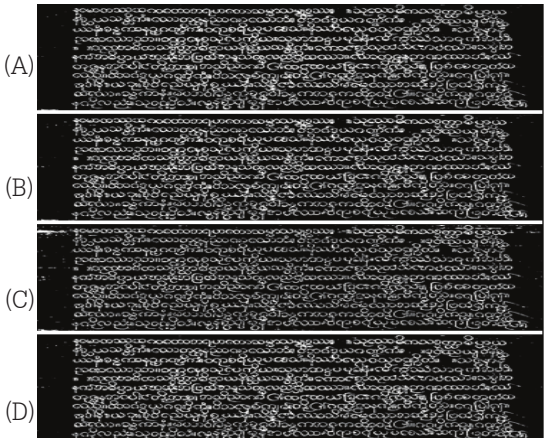


Figure 22 Binary Images: (a) from red color intensity, (b) from green color intensity, (c) from blue color intensity, (d) from gray color intensity.

To extract each character image or smallest group of character image, firstly in this paper, image segmentation line by line is done by using object frequency along the horizontal lines. Object frequency histogram is used to find the best optimal points. In Fig. 23 the object (character) pixels frequency histogram is shown.

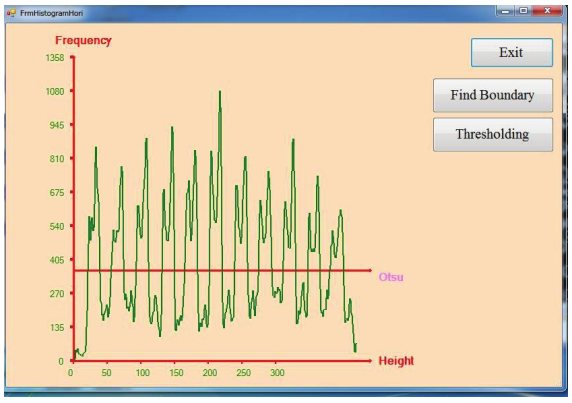


Figure 23 Histogram of object frequency along the height of image.

In Fig. 24 illustrated after thresholding the Red intensity binary image.

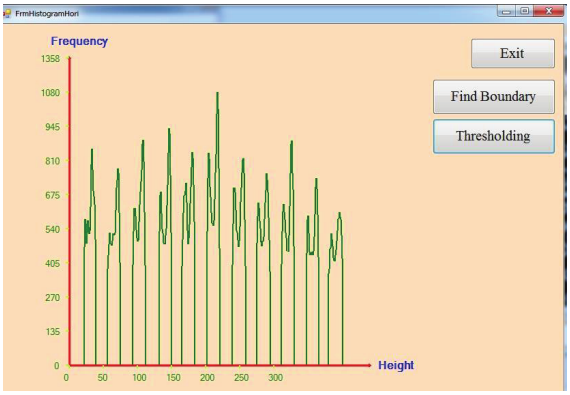


Figure 24 Histogram after thresholding frequency.

After thresholding, the frequency array using Otsu threshold value, possible candidate space regions become as clear as shown in in Fig. 25 separated with the red color vertical lines.

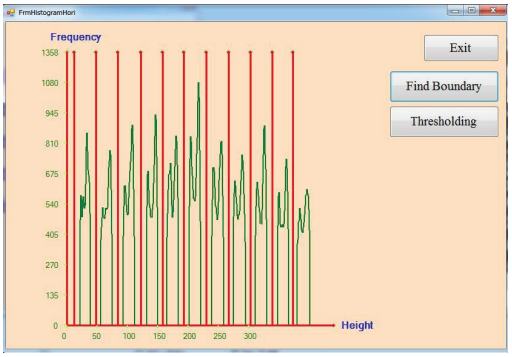


Figure 25 Searching the optimal points along the height of the image.

4.5 Experimental and results

Line segmentations by the object frequency histogram are the second and third experiment of this paper. In this experiment, palm leaf manuscript images were collected and took photos by the palm leaf manuscript preservation project in the Thaton city of Myanmar, conducted at Oo Pho Thee library. To do the necessary experiments, a small testing program is created and implemented using C# programming language. The resolution of the input images is 3863 x 498 (nearly RGB pixels) dpi in Jpeg format photos. The input images were converted to gray-scale images and then noise is reduced by using filtering techniques such as Mean, Median filters etc. Then

the filtered image is transformed to binary image by automatically selecting the Otsu binarization algorithm. After this, line and character segmentation were applied. In this study, many palm leaf manuscripts were considered for text line segmentation and multiple sub images from text line segmentation were investigated for character segmentation using Otsu thresholding algorithm.

Choosing the middle point of possible space between two character lines, sub images are segmented. In Fig. 26, the input image is demonstrated for the shown binary image, histograms and subsequent segmented images.

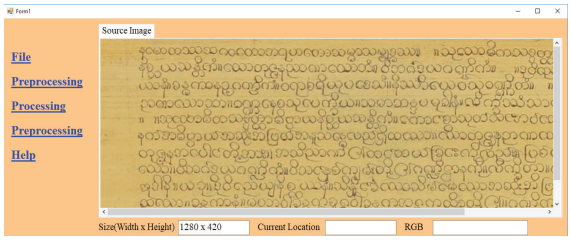


Figure 26. Input image of palm leaf manuscript.

After binarization technique and finding the optimal points for the input palm leaf manuscript, binary output image is segmented line by line using the optimal points along vertical direction as shown in Fig. 27.

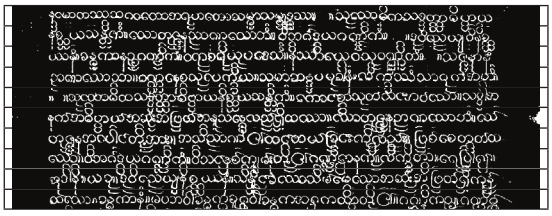


Figure 27 Image segmentation line by line from palm leaf manuscript with red intensity binary image.

In this experimental results, Table 2 presents that Overall Segmentation Accuracy for Syllabus Line segmentation. Table 2 shows the segmentation accuracy according to text line horizontal projection of palm leaf manuscripts. The accuracy of “Handwritten Isolated Syllabus Line Segmentation” is 79% with touching character Error rate of 11 % and Reject rate of 7% respectively.

Table 2 Overall Line Segmentation Accuracy

Number of Train Data	Number of correctly segmented	Segmentation Accuracy	Error Rate	Reject Rate
3005	2608	79%	11%	7%

5. CONCLUSIONS

In this paper we modeled sequence prediction as a classification problem and developed Neural Network based approaches to it. The performance analysis of VNN, an analysis tool system is created using C# programming environment. The user can save the Syllabus Line segmentations in the database. After storing all the patterns in DB, Synaptic matrix is calculated according to the DB size, features, patterns count and dimension size. Taking advantage of continuous word vectors, original discrete symbol sequences are transformed to continuous inputs for Neural Networks. The future work will include the methods of VNN classifier in distributed environments, where this step may be necessary in order to increase system throughput. Even though our work to date is concentrated on handwritten palm leaf manuscripts recognition, the methods presented can be easily generalized to more difficult problems, such as handwritten palm leaf manuscripts recognition. Otsu binarization method applied to the palm leaf manuscript should be done choosing only red color Intensity array. The output result of character images will be readily to use and the input images for the text extraction .This segmentation process is based on the new method: the red color intensity array which is the best for the palm leaf manuscript line segmentations. In the second step, the dissertations of the blue color vertical lines in the histogram are considered to be the locations between the text lines and these are the important facts of this research paper. According

to the research work and results this line segmentation system can make a conclusion as follow:

(a) Fast and robust method for ROI finding can be cropped automatically in palm leaf area of an image by using Syllabus Segmentation method.

(b) Line counts analysis is done to optimize or choose suitable amount of lines to find edges.

(c) The empirical method shows the result that in the future what background color should be used in the image acquisition process to digitize palm leaf manuscripts.

(d) The results of this research have been already using by Pali-text society to crop line segmentations from Palm leaf area automatically.

(e) The output cropped images will be the input images or palm leaf manuscripts data set for the next character segmentation text extraction work.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my most heartfelt gratitude and appreciation to my supervisor Professor Win Htay, University of Computer Studies (Thaton), Myanmar, for his support, advice, guidance and inspiration throughout the research time of my study. I am really thankful to Assistant Professor Dr. Olarik Surinta, Mahasarakham University, always encourages the development of generic skills such as intellectual understanding and judgment, problem solving skills, and critical thinking skills.

REFERENCES

Alahakoon, C. N. K. (2006). Identification of physical problems of major palm leaf manuscripts collections. Journal of the university librarians association of Sri Lanka, 10 , 54-66.

Chacko, B. P., & Babu, A. P. (2011). Online sequential extreme learning machine based handwritten character recognition. In Students' technology symposium (TechSym), IEEE (p. 142-147).

Forcadel, N., Guyader, C. L., & Gout, C. (2008). Generalized fast marching method: applications to image segmentation. Numerical Algorithms, 48 (1), 189-211.

Heath, M. D., Sarkar, S., Sanocki, T., & Bowyer, K. W. (1997). A robust visual method for assessing the relative performance of edge-detection algorithms. IEEE transactions on pattern analysis and machine intelligence, 19 (12), 1138-1359.

Kryzhanovskii, B. V., & Mikaelyn, A. L. (2003). An associative memory capable of recognizing strongly correlated patterns. Doklady Mathematics, 67 (3), 455-459.

Leavline, E. J., & Sutha, S. (2011, June). Gaussian noise removal in gray scale images using fast multiscale directional filter banks. In International conference on recent trends in information technology (ICRTIT) (p. 884-889).

Lee, T.-H. H. (2010). Edge detection analysis. Graduate institute of communication engineering, national Taiwan university.

Miljković, O. (2009). Image pre-processing tool. Kragujevac Journal of Mathematics, 97-107.

Moni, B. S., & Raju, G. (2011). Modified quadratic classifier and directional features for handwritten Malayalam character recognition. International journal of computer applications (1), 30-34.

Pali Text Society. (2018). Retrieved from http://www.palitext.com/subpages/thaton.htm

Soe, N. N., & Htay, W. (2017). Performance analysis of q-valued dimensions parametric vector neural network classifier. In 15thinternational conference on computer applications (ICCA) (p. 288-292).

Surinta, O., Schomaker, L., & Wiering, M. (2013). A comparison of feature and pixel-based methods for recognizing handwritten Bangla digits. In 12th international conference on document analysis and recognition (ICDAR) (p. 165-169).

Widiarti, A. R., Harjoko, A., Marsono, & Hartati, S. (2014). Preprocessing model of manuscripts in Javanese characters. Journal of signal and in formation processing, 5 (4), 112-122.

Zareizadeh, Z., Hasanzadeh, R. P., & Baghersalimi, G. (2013). A recursive color image edge detection method using Green's function ap proach. Optik - International Journal for Light and Electron Optics, 124 (21), 4847 - 4854.