



International Conference on Information Engineering, Management and Security
2015 [ICIEMS 2015]

ISBN	978-81-929742-7-9
Website	www.iciems.in
Received	10 - July - 2015
Article ID	ICIEMS030

VOL	01
eMail	iciems@asdf.res.in
Accepted	31- July - 2015
eAID	ICIEMS.2015.030

Akkhara-Muni: An instance for classifying PALI characters

Neha Gautam, R. S. Sharma, Garima Hazrati
Department of Computer Science Engineering, UCE, RTU

Abstract: Handwritten Recognition and Archaeology are significant facts of antediluvian epoch and scripts which are not tranquil to learn. For example, Pali written in Sinhala, Khmer, Burmese, Devanagari, Lao and many more which are having prodigious influence in the Buddhism culture since they hoard the lessons. Akkhara-Muni which constitutes Pali alphabet recognition been presented in this paper done in a few steps trailed by OCR for classification with the results that portray need of ancient scripts to be recognized and classification accuracy leads to 85.4%.

Keywords: Handwritten Recognition, Archaeology, Pali, OCR, Akkhara-Muni

1. Introduction

Man and machine are the two wheels of chariot. Humans need to get command on the machine in order to move with the wind. Keyboard is one of the most used devices now-a-days by every sphere of people from child to grandparents, student to high officials and also from hardware to virtual type. Handwritten is that field of image processing which is growing on a large scale to accomplish the requirements of global relations. Handwritten recognition is challenging though it is still a time and cost-saving work if done in daily activities and one of the most important uses of this technique is in archaeological department to learn ancient scripts written thousands of years ago. Scripts are defined as “markers” of a civilization keeping the record of lives in pictograms to phonograms as they are hoping to preserve heritage, culture, and history of the region understanding their importance [1]. A lot of work is done on the western scripts whereas eastern scripts are not having a notified work till yet and to learn about civilization history one needs to know about early writings. Optical character recognition (OCR) - a well-known technique with some modification that is used to recognize akkhara-muni (Pali) providing help to the archaeologists so that they can understand Buddhism and their teachings written in ancient period offering new thoughts to youth as past always pace an innovative way towards development.

1.1 Pali:

Language of prehistoric period- PALI has been inscribed in many scripts and languages besides being popular in southern countries for teachings of Buddhism. Pali alphabet basically consists of 41 letters: 6 vowels, 2 diphthongs, 32 consonants and 1 accessory nasal sound called as nighaitta [2]. Consonants are divided into 25 mutes, 6 semi-vowels, 1 sibilant and 1 aspirant followed by vowels comprises of

This paper is prepared exclusively for International Conference on Information Engineering, Management and Security 2015 [ICIEMS] which is published by ASDF International, Registered in London, United Kingdom. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honoured. For all other uses, contact the owner/author(s). Copyright Holder can be reached at copy@asdf.international for distribution.

2015 © Reserved by ASDF.international

Cite this article as: Neha Gautam, R. S. Sharma, Garima Hazrati. “Akkhara-Muni: An instance for classifying PALI characters.” *International Conference on Information Engineering, Management and Security (2015): 186-188*. Print.

long and short. Ian James, developer of a modified Latin font, and forms get their existence from ancient Brahmi and Pallava (ancestors of the Indic scripts) and later on named as Akkhara Muni, (Letters of the Sage). In Sri Lanka, Pali was used not only for the writing of Buddhist scriptures, but also to record the history of the country [3,4].

2 Related Work:

Several other works were done by OCR on many other languages and scripts giving top rated accuracies setting standard for others. In 2004 U. Pal and B.B. Chaudhari [5] describes work done on 12 major Indian scripts through OCR. In 2007 V.N. Manjunath Aradhya , G. Hemantha Kumar, S. Nousath [6] show a multi-lingual OCR leading to a good accuracy. Afterwards, Apurva A. Desai [7] use OCR tech-nique for recognizing gujrati handwritten digits in 2010 resulting 82% success rate and likewise Amit Choudhary, Rahul Mishra, Savita Ahlawat [8] routine this with binarization method to judge capability of OCR for english caharctes in 2013 giving accuracy of 85.62%.

Pali characters recognition using devnagri was shown by kiran s mantri, s p ramteke, s r suralkar[7] in 2012, by comprises following features like image pre-processing, feature extraction and classification algorithms that have been traversed to design software (OCR) with high performance. The recognition rate is 100% that has been done using simple feed forward multilayer perceptions also proposes a back propagation learning algorithm that is used to guide each network with the characters in that particular group.

Another work proposes a recognition system that has taken Pali cards of Bud-dhadasa Indapanno was presented by Tanasnee Phienthrakuland and Wanwisa Chevakulmongkol [8] in 2013. Its handwritten images have been refined by contrast adjusting, grayscale converting and noise removing. Basically the features of every single character are removed by the zoning method where average of all accuracies considered in groups comes out to be approximately 81.73%.

3 Our Work:

Optical Character Recognition (OCR) attempted first in 1870, developing from era of 40's while transforming from first generation to third now in present with a wide variety of applications in numerous fields from banking to education, archaeology to space science. Entire work is done on local database of Akkhara-muni script collected online by numerous resources and is shown as:



Fig. 1. Akkhara-Muni- A PALI Alphabet (Src. [2])

OCR is combination of several components illustrated in figure below:

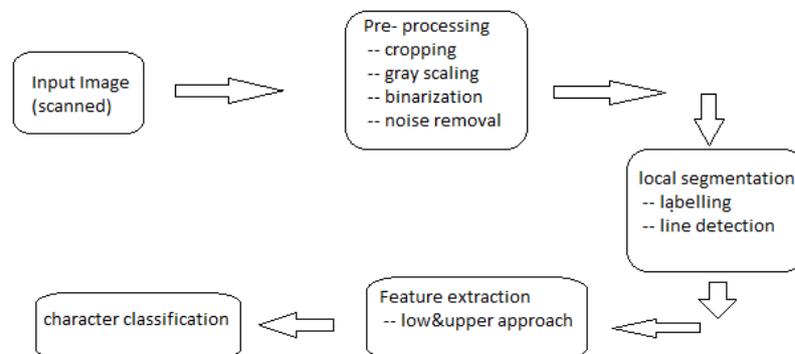


Fig. 2. Projected Accost for classification

Cite this article as: Neha Gautam, R. S. Sharma, Garima Hazrati. "Akkhara-Muni: An instance for classifying PALI characters." *International Conference on Information Engineering, Management and Security (2015)*: 186-188. Print.

Scanned input image is pre-processed by cropping it to let it fitted in proper dimensions trailed by gray scaling then binarization with noise removal by using various MATLAB functions. Subsequently local segmentation of pre-processed image is done by labeling and line detection. Later on, feature extraction is done using lower and upper approach resulting in character recognition with 85.4% success rate.

4 Results:

The overhead approach was tested on 60 sets of digits and samples used over here for experiments are collected online and trained in MATLAB. In all, we can say that this approach gave 85.4% success rate. The recognition rate for 2 vowels is high and 2 consonants are low as they are misidentified. The outcome we have obtained here is not very good if only OCR is taken but as script taken here akkhra-muni and no trained data sets available though we can't find any other work for comparison therefore we think that our result is good in this zone.

5 Conclusion:

In this approach Optical character recognition (OCR) is proposed with some other techniques for classification of Akkhara-Muni. Techniques like optical scanning, binarization, segmentation and feature extraction followed by classification of characters. The overall performance is 85.4%, but it can't be considered as a benchmark. Any model of classification is based on their feature extractions which are also need-ed to be used here for improving is performance. This attempt is unique as a whole. It offers a great value for scripts that are needed to be recognized and there is not much research done in this field of ancient scripts. We are currently working for other ancient Indic script.

References:

1. www.ancient.eu
2. "An Elementary Pali Course" book by Ven Narada, Thera pg. no. 9-12
3. www.skyknowledge.com
4. www.ancient.eu/www.ancientscripts.com
5. U. Pal, B.B. Chaudhari, "Indian scripts character recognition: A Survey" Pattern Recognition Society. Published by Elsevier Ltd 2004
6. V.N. Manjunath Aradhya , G. Hemantha Kumar, S. Nousath, "Multilingual OCR system for South Indian scripts and English documents: An approach based on Fourier transform and principal component analysis" Engineering Applications Artificial Intelligence Elsevier Ltd 2007
7. Apurva A. Desai, "Gujrati handwritten OCR through neural network" Pattern Recognition Elsevier Ltd. 2010
8. Amit Choudhary, Rahul Mishra, Savita Ahlawat, "Off- line handwritten character recognition using feature extracted from binarization technique" The Authors. Published by Elsevier B.V 2013
9. Kiran S. Mantri, R. S. Ramteke, S. R. Suralkar, "Pali Character Recognition System" at IJAIR 2012 ISSN: 2278-7844
10. Tanasanee Phienthrakuland Wanwisa Chevakulmongkol, "Handwritten Recognition on Pali Cards of Buddhadasa Indapanno" International Computer Science and Engineering Conference (ICSEC): ICSEC 2013

Cite this article as: Neha Gautam, R. S. Sharma, Garima Hazrati. "Akkhara-Muni: An instance for classifying PALI characters." *International Conference on Information Engineering, Management and Security (2015)*: 186-188. Print.