DEVELOPMENT OF FRAMEWORK TO RECOGNIZE AKHARA-MUNI CHARACTER USING ANN

Research Article

Development Of Framework To Recognize Akhara-Muni Character Using Ann

[1]Meenakshi, [2]Neha Goutam, [3]Raghavendra S., [4]Manjunath R., [5]Santosh Kumar J

Abstract

Pattern recognition problem always be a tough task because of the different type of shape of the characters and it became more challenging when ancient script is going to recognize. Various types of Ancient scripts are available in India and Akhara-muni is one of them scripts. To recognize the Akhara-muni character with good performance is the main task of the study. Optical character recognition technique is used to recognize the Akhara-muni characters where this study is used zoning and density method to extract the features and ANN to classify the Akhara-muni character. Achieved accuracy by the proposed system is 87.24%, which is better than the previous Akhara-muni character recognition system.

Keywords- Ancient script; Akhara-muni characters; character recognition; OCR

Introduction

Handwritten character recognition is always a popular and challenging problem of the pattern recognition and it became more challenging when ancient characters has to recognize [1]. Thus, few studies can see in the area of ancient script recognition. India has a big history and script background. Various types of scripts have been used in India where these scripts are Ancient and modern scripts. Akhara-muni scripts is also one of the Ancient scripts of India. Various type documents have been written by using Akhara-muni script and these documents were handwritten. However, only one study has been completed to recognize the Akhara-muni characters in 2015 [2]. Gautam, et al. [2] used Zoning and template matching technique to recognize the ancient characters.

Akhara-muni documents can help to understand the Indian history and culture of ancient time. However, it always be a tough task to save ancient documents. Ancient documents can be damage according to the time, whether and it will be very costly to restore the content from the ancient documents. Thus, the best way to preserve ancient documents, to convert these documents in

[1,2]Assistant Professor, Department of CSE, School of Engineering & Technology, Jain Deemed to be university, Jakkasandra Post Kanakapura Taluk Ramanagara - 562112
Meenakshib437@gmail.com and nehagautam1208@gmail.com
[3]Associate Professor, Dept. of CSE, Christ Deemed to be University, Kengeri campus, Bangalore,
raghav.trg@gmail.com
[4]Professor and Head, Dept. of CSE, R, R. Institute of Technology, Bangalore - 560090
drmanjunath.raj@gmail.com
[5]Associate Professor Department of CSE KSSEM, Bengaluru – 560109, VTU Belagavi
Santosh.kumar.j@kssem.edu.in

[1]Meenakshi, [2]Neha Goutam, [3]Raghavendra S., [4]Manjunath R., [5]Santosh Kumar J.,

digital way. Image of the document can take and store in computer of digital platform. So, documents can store for life long and will be very easy to access. However, searching and editing can be done in this image format. So, to solve these type of issue, Optical character recognition (OCR) techniques introduced. OCR system is a very popular and effective system and it has been user various type of document recognition. In OCR technique, image of any text use as input image and text will be editable in output of the OCR. Five steps have been followed to build any OCR system, input image, pre-processing, segmentation, feature extraction and classification [3].

This study is divided into 7 character. Introduction has been discussed in first chapter and details of the Akhara-muni script is presented in chapter 2. Chapter 3 presents the literature and dataset of Akhara-muni character has been discussed in chapter 4. Methodology to recognize the Akhara-muni character is mentioned in chapter 5. After that results and conclusion is presented in chapter 6 and 7, respectively.

1. Akhara-muni Script:.

Characters of the Akhara-muni script is presented in figure 1.

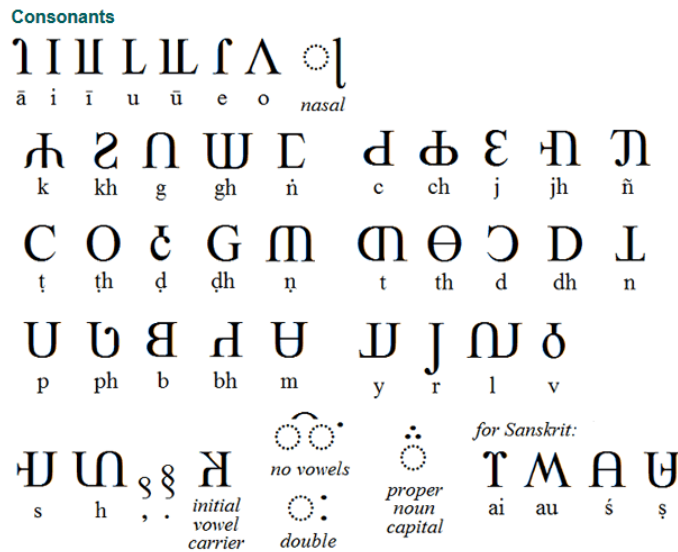Famous features of the Akhara-muni script

Fig 1: Akhara-muni character [4]

Phonemic type of alphabet was used to write Akhara-muni scripts and left to write way is followed to write this script. The most useful reason to know the Akhara-muni script is pali language. Pali language was used to read the Akhara-muni text.

2. Literature

A huge amount of work has been completed in the field of the character recognition in recent year. Such as Ptucha, et al. [5] used CNN to analysis the handwritten text, where architecture is inspired by the lexicon and RNN was used to introduce the system. Similarly, Hasan, et al. [6] was used CNN to recognize the handwritten Bangle characters. Author was used 10 layer to introduce the CNN. Cilia, et al. [7] suggested the important the features in the character recognition. Without useful feature extraction, character recognition with good accuracy will be a tough task. Feature extraction can be completed manually such as it happens in supervised learning or can be automatically like unsupervised learning. Features are the most important part of the any

recognition system. Ahmed, et al. [8] work on handwritten Urdu character recognition by using standard dataset. Urdu is a cursive script and it is always be a tough task to recognize it. 1-D BLSTM classifier was used to recognize the Urdu characters. Jayasundara, et al. [9] introduce a study where small dataset can also be used to recognize the characters because label of the big dataset is always be problematic. EMNIST (text and characters) and MNIST dataset was used to complete the study. Nongmeikapam, et al. [10] worked to recognize the Manipuri Meetei-Mayek character which was handwritten. HOG and SVM was used the extract the features of the characters and the classification is done with the help of SVM. This study achieved 96.92% accuracy, which is very good recognition rate. Chowdhury, et al. [11] worked on the handwritten Bangle characters recognition where CNN was used the recognize the Bangle characters. Bangla Lekha-Isolated dataset was used to complete this study and achieved 91.81% recognition rate, which is good recognition rate but the performance can be increased. Gogna and Majumdar [12] used Discriminative Autoencoder to extract the features of the character, However, this is used in unsupervised learning where separately feature extraction is not very important. This study is proposed a system where Discriminative Autoencoder used to extract the features and this study is behave as supervised learning. KNN was used to classify the extracted features. Aneja and Aneja [13] used various type of the pre-trained architecture of CNN such as VGG, DenseNet, Inception and Alex Net. Alex Net perform well come to other network to recognize the Devanagari characters. Kowsalya and Periasamy [14] focused on the recognition of Tamil characters where ANN was used to classify the Tamil characters. Total 5 steps were focused to suggest the Tamil characters recognition system. These steps were, input image, pre-processing, segmentation, feature extraction and classification. All work to build the system is done on MATLAB. Various recognized the script are modern script and very few is Ancient script.

Brahmi script are the popular script as Ancient script and Pali language was also used to read Brahmi script. Similarly, Ariyaka, and Akharamuni was also the script where Pali language was used to read these scripts. Brahmi script has been recognize in 1983 [15]. After that Brahmi script recognize by the zoning and template matching [3]. Again Brahmi character was recognize in 2017 with the help of Geometric feature extraction and some set of rules [16]. Apart from that, Brahmi text recognition is also completed in 2020 where various type of supervised and unsupervised techniques was used [1, 17].

Apart from that Ariyaka character recognition also be done where template matching and zoning method to recognize the Ariyaka and achieved 85.73% accuracy [18]. Similarly, Akharamuni character recognition was also has been done in 2015 [2] by using template matching and zoning method, where the performance of Akhara-muni character recognition was 85.66%. The performance of the Akharamuni character recognition system can be increased if ANN will use apart form template matching because ANN is better to the template matching to recognize the character.

3. Dataset

Dataset is most important thing of any recognition system. In the precious study of the Akharamuni recognition system, a dataset was used. This study is also used same dataset to complete this study. The dataset has 4400 sample for training and 220 for testing the system. Where to train and test the vowel, 900 and 45 sample are used, respectively. Similarly, 3500 and 175 samples are used to train the test the consonants of the Akharamuni, respectively.

4. Methodology

[1]Meenakshi, [2]Neha Goutam, [3]Raghavendra S., [4]Manjunath R., [5]Santosh Kumar J.,

OCR framework is used to introduce the Akharamuni character recognition system. Input image, pre-processing, segmentation, feature extraction and classification are the step of the OCR system (figure 2).

4.1 Input Image:
input image the first step of the OCR system where image of Akharamuni character is used as input. Format of the image is uniform, and shape of the image can be anything.

4.2 Pre-processing:
this is the second step of the OCR system and the working process of this step to increase the pixel quality of the image according to the requirement. In this study, black and white is used so, all input images are converted into gray scale image by using threshold value and similarly, all gray scale images are converted into B & W images, again by using threshold value. All B & W image are in different size so, image resizing step also done under pre-processing. In this step, the image is converted into 45X27 size.

4.3 Segmentation:
Segmentation of the third step of the OCR and it is used to separate all characters from the image. Few characters are the part of an image and it is not easy to recognize all characters together so, separation of the character from the image in important and required step of OCR. All characters form the image has been separated by segmentation method and save as new image.

4.4 Feature extraction:
Feature extraction is the most important step of the any recognition system under supervised learning. Zoning method has been used to divide the characters into various type of zone and then important features is used form each zone. Binarization technique is used to extract the important features from each zone because it was used earlier to extract the features of Akharamuni characters. The size of each zone was 9x9. Means, total 15 type of features has been used to recognize a character.

4.5 Classification:
This is the decision-making step of any OCR system. All extracted features are used as input for this step where a classifier will used extracted features and take the decision. ANN is used to classify the extracted features. ANN is artificial neural network which is inspired by the human neurons. These artificial neurons help to take the decision. Where various type of parameter is also important to develop the classifier and it can help to perform well. Such as leaning rate, epoch, etc. learning rate for this study was 0.1 where epoch was 50. Validation dataset was the 20% of the training dataset.
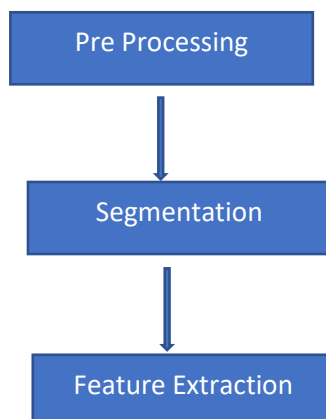
```
┌─────────────────────┐
│   Pre Processing    │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│    Segmentation     │
└─────────────────────┘
           │
           ▼
┌─────────────────────┐
│ Feature Extraction  │
└─────────────────────┘
```

Figure 2: OCR system for Akharamuni Character recognition [18]

5. Result

Methodology of the previous section is applied on the Akharamuni dataset and MATLAB is used to develop the system. Achieved accuracy of the system was 87.47%. where performance of the vowel was 88.52% and recognition rate of the consonant was 86.42%.
Comparison of this study to previous study

Table 1: Comparison the proposed study to previous study

Study	Recognition rate
Gautam, et al. [2]	85.66%
Proposed study	87.24%

As can see in table 1, proposed study has better recognition rate compare to previous study. So, it can say that proposed system is better to the previous study and ANN perform well compare to template matching to recognize the Akharamuni characters.

Conclusion

This study used OCR to recognize the Akharamuni characters where zoning method and binarization was used to extract the features of the Akharamuni characters and ANN used to classify the extracted features. Total five steps, input, pre-processing, segmentation, feature extraction and classification was used to develop the Akharamuni character recognition system. This study showed the recognition rate 87.24% which is better than the previous study where dataset of both studies was common. Although, the performance of the Akharamuni character recognition can be further increased if CNN will use.

References

[1]   N. Gautam, S. S. Chai, S. Afrin, and J. Jose, "Brahmi word recognition by supervised techniques," 2020.

[2]   N. Gautam, R. Sharma, and G. Hazrati, "Akkhara-Muni: An instance for classifying PALI characters," in *International Conference on Computational Intelligence and Communication Networks*, Jabalpur, India, 2015: IEEE, pp. 252-253.

[3]   N. Gautam, R. S. Sharma, and G. Hazrati, "Handwriting Recognition of Brahmi Script (an Artefact): Base of PALI Language," in *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems*, vol. 2: Springer, Cham, 2016, pp. 519-527.

[4]   https://omniglot.com/conscripts/akkharamuni.htm.

[1]Meenakshi, [2]Neha Goutam, [3]Raghavendra S., [4]Manjunath R., [5]Santosh Kumar J.,

[5]   R. Ptucha, F. P. Such, S. Pillai, F. Brockler, V. Singh, and P. Hutkowski, "Intelligent character recognition using fully convolutional neural networks," *Pattern Recognition,* vol. 88, pp. 604-613, 2019.

[6]   M. Hasan, F. T. Z. Asha, and T. Zubaer, "Bangla handwritten character recognition using convolutional neural network," 2019.

[7]   N. D. Cilia, C. De Stefano, F. Fontanella, and A. S. di Freca, "A ranking-based feature selection approach for handwritten character recognition," *Pattern Recognition Letters,* vol. 121, pp. 77-86, 2019.

[8]   S. B. Ahmed, S. Naz, S. Swati, and M. I. Razzak, "Handwritten Urdu character recognition using one-dimensional BLSTM classifier," *Neural Computing and Applications,* vol. 31, no. 4, pp. 1143-1151, 2019.

[9]   V. Jayasundara, S. Jayasekara, H. Jayasekara, J. Rajasegaran, S. Seneviratne, and R. Rodrigo, "Textcaps: Handwritten character recognition with very small datasets," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019: IEEE, pp. 254-262.

[10]   K. Nongmeikapam, W. K. Kumar, O. N. Meetei, and T. Tuithung, "Increasing the effectiveness of handwritten Manipuri Meetei-Mayek character recognition using multiple-HOG-feature descriptors," *Sādhanā,* vol. 44, no. 5, p. 104, 2019.

[11]   R. R. Chowdhury, M. S. Hossain, R. Ul Islam, K. Andersson, and S. Hossain, "Bangla handwritten character recognition using convolutional neural network with data augmentation," in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, 2019: IEEE, pp. 318-323.

[12]   A. Gogna and A. Majumdar, "Discriminative autoencoder for feature extraction: Application to character recognition," *Neural Processing Letters,* vol. 49, no. 3, pp. 1723-1735, 2019.

[13]   N. Aneja and S. Aneja, "Transfer Learning using CNN for Handwritten Devanagari Character Recognition," *arXiv preprint arXiv:1909.08774,* 2019.

[14]   S. Kowsalya and P. Periasamy, "Recognition of Tamil handwritten character using modified neural network with aid of elephant herding optimization," *Multimedia Tools and Applications,* vol. 78, no. 17, pp. 25043-25061, 2019.

[15]   G. Siromoney, R. Chandrasekaran, and M. Chandrasekaran, "Machine recognition of Brahmi script," *IEEE Transactions on Systems Man and Cybernetics,* vol. SMC-13, no. 4, pp. 648-654, 1983, doi: 10.1109/TSMC.1983.6313155.

[16]   N. Gautam and S. S. Chai, "Optical Character Recognition for Brahmi Script Using Geometric Method," *Journal of Telecommunication, Electronic and Computer Engineering,* vol. 9, no. 3-11, pp. 131-136, 2017.

[17]   N. Gautam, S. S. Chai, and J. Jose, "Recognition of Brahmi Words by Using Deep Convolutional Neural Network," 2020.

[18]   N. Gautam, R. Sharma, and G. Hazrati, "Ariyaka: A PALI Alphabet Recognition Script," in *International Conference on Computational Intelligence and Communication Networks*, Jabalpur, India, 2015: IEEE, pp. 293-295.