

Global Colloquium in Recent Advancement and Effectual Researches in Engineering, Science and Technology (RAEREST 2016)

An Approach towards Malayalam Handwriting Recognition Using Dissimilar Classifiers

Meenu Alex^{a*}, Smija Das^b

Department of Computer Science & Engineering, St. Joseph's College of Engineering & Technology, Palai, Kerala, India-686579

Abstract

The handwriting recognition in Malayalam is a challenging as well as emerging area of pattern recognition. It is a tedious process mainly due to its enormous character set. Here we propose a novel method for handwriting recognition by using two dissimilar classifiers. It can also be called as an ensemble method in which multiple classifiers are combined to solve a particular problem and thereby improve the performance of the system. The experiment is conducted in 2 phases. In the first phase, 33 isolated characters in Malayalam were used. In the second phase, Malayalam sentences were used. From the preprocessed image, we were extracted two features: SURF feature and Curvature feature. These features were fed as input to a neural network and an SVM classifier. Finally, the result of both the classifiers was combined to get the final results. The system showed an accuracy of 89.2% in the first phase. An accuracy of 81.1% was exhibited in the second phase.

© 2016 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the organizing committee of RAEREST 2016

Keywords: Handwriting Recognition; Segmentation; SURF; Curvature feature; SVM; Neural Network; Ensemble learning

1. Introduction

Now is the age of digitalization. Rather than storing the data on papers, it can be safely stored and easily accessed, then digitalized. Therefore Handwriting Recognition is of prime importance. Handwriting Recognition is an emerging as well as challenging area in the fields of pattern recognition and computer vision. Various Indian offices including post offices, banks etc handle lot of handwritten documents and thus the handwriting recognition (HCR) has attained a hotspot area of research. Other applications include automatic number plate recognition, CTS scanning, preservation of degraded documents etc. The aim of a handwriting recognition system is to convert human readable characters which are present in a photographed or digitized sheet of paper and convert it into a machine editable form. A handwriting recognition system is of two types: online or offline. In an online character recognition system, the data is captured at the same time when the user writes on a digitizer. It actually does the real time conversion of characters

to their corresponding Unicode values. In an offline system, a scanner captures the data after the writing process is over.

The recognition of printed text is a comparatively simple process. The variation in between a character printed at different parts of the world is negligible. But the scenario is extremely different in the case of handwritten texts. A character being sketched in hand differs individually depends upon a number of factors. The emotion, mood, age, health condition etc of the writer along with the variation provided by the pen used. All makes numerous variations in a character being written by different individuals at different parts of the world. The complexity of our problem thus lies in the complexity of our language added up to the huge variations in penning down the characters that vary individually.

An important feature of Malayalam language is its enormous character set. Thus the identification of characters may be posing another challenge in the form of similarity between the characters. Adding up to the scene is the similarity in writing styles of different people. Character recognition has already been successful in foreign languages like English, Japanese, Chinese, Arabic etc. The universal language English presents itself simple with only 26 characters – not forgetting the divisions based on the case of letters. Several researchers have come out with notable accuracy in the handwritten recognition of English by virtue of this.

The recognition of scripts is a tedious process for South Indian languages like Malayalam, Kannada, Tamil, Telugu etc. This is mainly due to the large character set, presence of compound characters and so on. Malayalam, being the second most difficult language in the globe, has a distinguished character set varying largely in between them. So to reach 100% accuracy is very difficult job. The scope of this work lies in converting important handwritten Malayalam documents into well kept digital data.

2. Related Work

A lot of works are reported in foreign languages [2] in the domain of Handwritten Character Recognition. Among Indian languages like Devanagari, Tamil, Oriya and Bangla many works were occurred [3,4]. But in case of South Indian languages especially Malayalam, only few works were reported. A recognition system that can identify the complete character set of Malayalam is not developed till now. The main difficulty in Malayalam character recognition system is the lack of availability of a benchmarking database for comparison.

The first work in Malayalam Character Recognition was reported in 2007 by Lajish V.L [5]. It uses fuzzy-zoning method and normalized vector distance measures for the recognition of 44 Malayalam characters. Renju John et al [6] came forward with the concept of 1D Wavelet Transform of Projection Profiles for Isolated Handwritten Malayalam Character Recognition. Handwritten character recognition by applying Daubechie wavelet coefficients was proposed in [7]. M. Abdul Rahiman et al [8] proposed an HLH intensity pattern based method for recognition of Malayalam characters. A recognizing system for Malayalam characters using discrete features was introduced by Binu P. Chacko and Babu Anto in [9]. Jomy John et al [10] proposed a chain code histogram based method for recognizing vowels of Malayalam. Bindu S. Moni et al [11] invented a handwriting recognition system based on run length count (RLC). Vidya V. et al [12] proposed a method for handwritten character recognition based on Probabilistic Simplified Fuzzy ARTMAP (PSFAM). Features like Zernike moment features, cross feature, distance feature and fuzzy depth are extracted from the character. This method gained an accuracy of 79.48% for 142 Malayalam characters. In 2014, Shanjana C et al [13] proposed a method for Malayalam character segmentation. In this work, segmentation of characters is performed by combining Vertical projection profile method along with connected component analysis method.

The above mentioned HCR systems propose a variety of feature extraction methods but a good result has not yet been achieved till now. All the above mentioned works are based on a single classifier system. Even though combination of multiple classifiers (similar classifiers) has been experimented in Malayalam, but a combination of dissimilar classifiers has not been tested till now. This motivated us to examine the outcome of using dissimilar classifier combination in Malayalam.

Download English Version:

<https://daneshyari.com/en/article/4962530>

Download Persian Version:

<https://daneshyari.com/article/4962530>

[Daneshyari.com](https://daneshyari.com)