Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Recognition of Handwritten Bangla Numerals using Adaptive Coefficient Matching Technique

Amitava Choudhury[a,*], Alok Negi[b] and Sanhita Das[c]

[a]*Indian Institute of Engineering Science and Technology, Shibpur, India*
[b]*IMS Unison University, Dehradun, Indian*
[c]*Cognizent Technology Services, Kolkata, India*

Abstract

This paper presents an approach to recognize handwritten bangla numerals from low pixel mobile camera based on correlation co-efficient. In this paper, an approach is identified to recognize such handwritten bangla numerals having different shape and scan through from any low pixel mobile device. It has also shown that while writing in continues sometimes two separated digit connected to each other, to distinguish such digit we used Line segmentation. High correlation coefficient will provide the successful results of recognition. A large dataset is used for training and testing purpose which gives overall 93.80% accuracy using mean of maximum correlation and 95.70% accuracy is shown when individually correlation found.

1. Introduction

Handwritten Numerical Recognition (HNR) system is aimed to identify a numeric value using computing device. HNR has always been a very challenging work in the field of pattern recognition. Using Optical Character Recognizer (OCR) we are able to edit a scan document image. Electronic translation of an image (like printed, handwritten or typewritten) into editable text is the main aim of OCR. Generally English is used in computing systems. But now days other languages are become a part of OCR. Bangla is one of the most popular languages in country like India and Bangladesh. The research of HNR is very popular for various practical applications for blind readability, bank cheque, and automatic pin code reading of postal system. Numerals in bangla are very curve in nature. Several researches are going on in the field of Bengali character recognition systems to improve its accuracy and efficiency. Still there are several challenges occur during the improvement. The writing pattern is different shape while we are talking about handwritten document. In begali number system there are ten digits, staring from 0 to 9, having different shape size and style. As hand writing is varied from person to person, to identify that particular content from the mobile captured images really a difficult work. OCR can be performed via online and offline. Offline handwritten recognition systems

*Corresponding author. Tel.: +91 8902691510.
E-mail address: a.choudhury2013@gmail.com

of finding numerals are present in digital image of handwritten document. In this paper, a template matching based recognition system is proposed. Scan document can be recognizing with the proposed method. Mobile captured images are also recognised using the described approach. The main challenge in mobile captured images is noise. Hence it should be kept on mind that noise should be eliminating properly for better result. The organization of this paper is as follows: In section 2, we discuss about the brief description of literature survey, in section 3, we describe the proposed methodology, in section 4, we describe about the experimental results and finally in section 5, conclusion is drawn.

2. Literature Survey

Bangla numeral recognition can be classified into two ways, offline and online systems. We mainly concentrate on off line HNR. The offline character recognition is presented in[3] to describe HNR using template matching technique. In Mohammed Moshiul Hoque *et al.*[2] proposed an OCR engine to recognize the Bangla numerals. The unique fuzzy base rule for applied to each numeral. The numerals features i.e. fuzzy features are mapped into some predefined linguistic variables. To recognition of a numeral, fuzzy rules are calculated, this fuzzy rule is compared with rule base and the character that contains highest percentage value is recognized character. The acceptance rate of the successful recognition is 82.099%. Vishweshwarayya C. Hallur and Ravindra S. Hegadi describe an approach based on template matching for handwritten numeral recognition of another Indian script[4]. OCR is used to text recognition. To recognize Indian handwritten script several algorithm has been proposed. K. Roy, C. Chaudhuri, U. Pal and M. Kundu describe an approach on the effect of varying training set sized on the recognition performance with handwritten bangla nurelas[5]. Amitava Choudhury and Joydeep Mukherjee also describe an approach to recognition of handwritten bangla numeral recognition using correlation coefficient method[6]. OCR can be recognizing the characters as well the numeric value also. In Ketan S. Machhale *et al.*[7] proposed an adaptive template matching and feature extraction using curvlet transform for the recognition of handwritten Bangla numerals. The feature extracted from the numeral set based on the curvelet transform morphological method. To classify the characters a KNN classifier is used. Ujjawal Bhattacharya *et al.*[8] proposed an efficient algorithm for recognition system of mixed numerals. A multistage cascaded recognition scheme using wavelet based multi resolution representations and multilayer perception of MLP classifier. The proposed method has been implemented for recognition of handwritten numerals in mixed script situation. In U. Pal, B. B. Chaudhuri[9], an automatic recognition method for unconstrained off-line Bangla handwritten numerals was described. To obtain feature from each character, they provide a concept of water overflow from the reservoir as well as topological and statistical feature and numerals. The direction of water flow, height of water level when water overflows from the reservoir, position of the reservoir with respect to the character bonding box, shape of the reservoir etc, are used in the recognition scheme. Pulak purakait and Bhabatosh Chanda describe offline recognition of handwritten Bengali numerals[10]. They proposed some novel morphological feature and *k*-curvature feature extraction technique to recognize handwritten scripts. Multi-layer perceptron (MLP) classifier used to train the feature spaces and then fuse those classifiers used as modified native bayed combination to increase accuracy of recognition. An overview of segmentation of handwritten connected digit[11] is described by R. V. Kulkarni and P. N. Vasambekar.

3. Proposed Work

Writing style of Bengali numeral is different from English font and its shown that each number has a curve on its writing skill. As the handwriting is varies from person to person so recognize that using OCR is very challenging task. In this paper, an approach is identified to recognize such handwritten Bangla numerals having different shape and scan through from any low pixel mobile device. It has also shown that while writing in continues sometimes two separated digit connected to each other, to distinguish such digit we used Line segmentation. Input image has been resized into 32×32 pixel, which implies the each image size in the dataset. The correlation coefficient is computed between the input image and the image that are stored as training data in dataset. A high correlation coefficient will provide the successful results of recognition. The numeral digits 0 to 9 as represented in Bangla language is shown in Fig. 1.