A Novel Approach to Text Dependent Writer Identification of Kannada Handwriting

B.V.Dhandra[a], Vijayalaxmi.M.B[a]

[a] Department of P.G.Studies and Research in Computer Science
Gulbarga University, Gulbarga, India

Abstract

In this paper a text dependent writer identification method based on Kannada handwriting is proposed. The database of Kannada handwritten words collected from 25 writers is intended to provide training and testing sets for Kannada writer identification research, because there is no well-known database containing Kannada handwritten words available for writer identification problem. A feature vector consisting of directional multi-resolution spatial features based on Radon Transform and Discrete Cosine Transform and structural features such as aspect ratio and on-pixel ratio are extracted from word images. The novel approach of combining the features of two or more words is proposed and performance analysis is done through nearest neighbor classifier with modified 5-fold cross-validation. An average identification accuracy of 93.2582% is achieved by using only the single words, and accuracy of 100% is achieved using the combination of features of three or more words. The writer identification results show that feature vectors extracted from longer words, words having more structural variation and combination of features of two or more words have higher impact on writer identification.

1. Introduction

Writer identification is a promising area of research due to its very wide applications in the areas of forensic analysis, authorship determination of historical documents, providing access control to users, checking authenticity of bank cheque etc. In this paper text dependent approach of writer identification is dealt with where in every writer

has written the same text. In text independent approach the writers may write any text as a line, paragraph or document without any restriction. In text dependent approach every writer has to write the fixed text, the identification of writer by comparing the query text against the text of known authorship is performed by comparing the characteristics of individual allographs, characters, words or lines with the same content to identify the underlying writer of the text. For linking or identifying a person to handwritten document, it is not necessary to examine each and every word of the document, even the forensic examiners correlate only the characteristics of few frequently used words to identify the underlying writer, so we have used the same approach in the current study of writer identification.

2. Review of Literature

A survey on writer identification schemes is presented by Sreeraj et al. in [6]. A number of new text dependent approaches to writer identification have been proposed in the literature.

Zois et al. [1] proposed an approach based on morphological processing of horizontal projection profiles of English word "characteristic" written 45 copies by each writer. Bayesian and neural network classifiers are used for writer identification and achieved the accuracy of more than 95% for 50 writers. Somaya Al-Ma'adeed et al. [3] used edge-based directional probability distributions for the writer identification problem of 100 writers using 16 different Arabic words and phrases in text dependent approach and using KNN classifier achieved the minimum identification accuracy of 53.4% for shorter words and maximum of 93.8% for some phrases with KNN classifier. Thendral et al. [5] proposed text dependent approach of writer identification using local and global features and achieved an identification accuracy of 98.6% by decision tree classification model for 10-fold cross validation.

Saranya et al. [4] extracted features viz. edge based directional distribution and edge hinge distributions as two edge based features and the length of the word, height of the word, height from baseline to upper edge, height from baseline to lower edge, ascender and descender baseline, moment invariants for a given image as word measurement features. Additional features were the character level features like aspect ratio, loops, junctions and end points. They trained the feature vectors using support vector machine and with 10-fold crossvalidation obtained 94.27% accuracy for word level, 90.10% for character level and also found that SVM with Polynomial kernel could predict the writer of the handwritten word more accurately than the other models in their method.

Aymen Chaabouni et al. [7] presented a method for writer identification based on multifractal features for both offline and online approaches. They extracted the multi-fractal dimensions from the images of words and their on-line signals. They have combined both on-line and off-line approaches in order to enhance the performance of the writer identification system. The tests were performed on the writing of 110 writers from the ADAB database and obtained the Top-1 and Top-10 results of 70.3% and 85.9% for Offline, 73.1% and 86.5% for online and 83.8% and 91.3% for combination of offline and online. Ayumu Soma et al. [8] used character features and recognizers for writer identification of offline handwritten Kanji characters. The combination of two global features viz. the size and position of a character and the two-dimensional space spectrum, two local features viz, the weighted direction index histogram and features extracted from each stroke, and majority voting as a recognizer in their writer identification method. They evaluated their method from the standpoints: the number of character classes, votes, training samples, and writers, and obtained the identification rate of 7 points higher than the conventional method using a single feature and an identification rate over 99% by using three character classes.

Pulak et al. [9] proposed four feature set namely normalized mass of word by directional opening, by directional closing, by directional erosion and k-curvature feature of each word for writer recognition on Telugu handwritten documents. The handwritten documents from 22 writers with 5 documents per writer, 10 words were segmented from each document. Segmented words of similar content were normalized and divided into a number of cells of fixed size. All the feature sets extracted from 10 different words from each of the document in the data sets were subjected to nearest neighbour classifier for writer identification. Using the directional opening feature they could achieve the identification rate of 100% which outperformed other feature sets.

A number of approaches to text independent writer identification also exist in literature. Hertel et al. [2] used a subset of IAM database containing 250 pages written by 50 writers in English. Each writer contributed 5 pages of