



Efficient zone identification approach for the recognition of online handwritten Gurmukhi script

Harjeet Singh¹ · R. K. Sharma¹ · V. P. Singh¹

Received: 13 January 2017 / Accepted: 30 December 2017
© The Natural Computing Applications Forum 2018

Abstract

In this paper, a real-time recognition approach for online handwritten Gurmukhi character combinations with matras (Gurmukhi *Vowels*) has been addressed. Moreover, various challenges in the recognition of online handwritten Gurmukhi script have also been discussed. The strokes for writing Gurmukhi words can be drawn in one of the three horizontal zones, namely upper, middle and lower. With the huge variation in handwriting styles of writers, zone identification of online capturing strokes has become one of the major issues in Gurmukhi script recognition. In this connection, a robust zone identification algorithm has been proposed in this paper. We have considered 93 stroke classes (12 for upper zone and 81 for lower zone) to implement the proposed zone identification algorithm. The statistical tool, support vector machine, has been employed as the classifier for stroke classification. A total of 52,500 word samples were collected from 175 writers in order to train the classifier. The proposed zone identification algorithm yielded an accuracy of 99.75% when tested on a data set of 21,500 character combinations with matras, written by 10 new writers.

Keywords Online handwritten Gurmukhi script recognition · Zone identification · Support vector machine

1 Introduction

In the era of touch-based and pen/stylus-based devices, online handwriting recognition has received much attention in the form of real-time applications, including text input, handwritten notes, signature capturing and mathematical expression recognition [3, 11]. Due to the rapid revolution in information technology-related products, including touch screen mobiles, digital tablets, notepads, the demand for online handwritten text recognition-based applications is increasing day by day. The major difficulties in recognition of Gurmukhi script are: (1) its complex structure, (2) similar shapes of stroke classes and (3) variation in handwriting styles. The present paper is concerned with the

development and implementation of an algorithm to recognize Gurmukhi script accurately.

Online handwriting recognition for any language is a difficult task due to the issues posed by variations in handwriting styles of writers, structure of scripts and different type of handwriting devices used for writing. Tapert et al. [19] have described the state of art in online handwritten character and word recognition systems. They have explored online and offline recognition, shape recognition, preprocessing techniques and post-processing techniques in their survey paper. Prasad et al. [12] have implemented a divide and conquer technique in online handwritten Kannada character recognition. The character combination is segmented into three basic units, namely middle, top and bottom in their work. These units are recognized separately to overcome the complexity of the huge number of character combinations. After that the preprocessing steps, i.e., noise removal, re-sampling and size normalization, are applied on strokes of each unit. The k -NN classifier is used for stroke classification, and their recognition system achieved 81.0% accuracy for Kannada characters.

✉ Harjeet Singh
harjeet.singh@thapar.edu

R. K. Sharma
rksharma@thapar.edu

V. P. Singh
vpsingh@thapar.edu

¹ Computer Science and Engineering Department, Thapar University, Patiala, Punjab 147004, India