Recognition of online unconstrained handwritten Gurmukhi characters based on Finite State Automata

HARJEET SINGH*, R K SHARMA and V P SINGH

Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology (Deemed University), Patiala, India
e-mail: harjeet.singh@thapar.edu; rksharma@thapar.edu; vpsingh@thapar.edu

Abstract. This paper presents a language-based efficient post-processing algorithm for the recognition of online unconstrained handwritten Gurmukhi characters. A total of 93 stroke classes have been identified to recognize the Gurmukhi character set in this work. Support Vector Machine (SVM) classifier has been employed for stroke classification. The main objective of this paper is to improve the character level recognition accuracy using an efficient Finite State Automata (FSA)-based formation of Gurmukhi characters algorithm. A database of 21,945 online handwritten Gurmukhi words is primarily collected in this experiment. After analysing the collected database, we have observed that a character can be written using one or more strokes. Therefore, a total of 65,946 strokes have been annotated using the 93 identified stroke classes. Among these strokes, 15,069 stroke samples are considered for training the classifier. The proposed system achieved promising recognition accuracy of 97.3% for Gurmukhi characters, when tested with a new database of 8,200 characters, written by 20 different writers.

Keywords. Online handwriting recognition; Gurmukhi characters; post-processing; strokes association; Finite State Automata; SVM classifier.

1. Introduction

Study in the field of online handwriting recognition for Indian scripts has become popular nowadays. More attention is given towards online handwriting recognition because of these reasons: (i) handwriting is easier than typing, (ii) typing is not possible everywhere, (iii) non-availability of all languages keyboard and (iv) typing is difficult for varied size of characters. Moreover, due to rapid revolution in devices related to Information Technology, including mobiles based on touch screens, digital tablets, notepads, *etc.*, the demand of applications based on online handwriting recognition is increasing day by day. These devices help us in capturing information with the help of a digital-pen/stylus. This captured information is stored as *x–y* traces with progressive time. A set of such traces is referred to as a stroke in an online handwriting recognition system. A stroke is thus the set of coordinates captured between the two events, namely, pen-down and pen-up with the help of a writing device. A stroke is the building block in online handwriting recognition systems.

Online handwritten character recognition of any scripting language is a difficult task due to the problems of different handwriting styles of different individuals, complex

structure of a language and different touch-based handwriting capturing devices used by the individuals. A good amount of research work has been carried out on online handwriting recognition in the recent past for different scripts such as Chinese, Japanese, Korean and Arabic [1–4]. On the other hand, works on isolated character recognition for many Indic languages like Bangla, Hindi, Tamil and Telugu have been reported by many researchers in the past few years [5–8]. The present study deals with recognition of Gurmukhi script, the most renowned script of north India used to write the Punjabi language. A stroke is considered as the smallest unit in online Gurmukhi script recognition. Further, a Gurmukhi character can be formed using a single recognized stroke or with a combination of more than one recognized stroke. The most common difficulties that occur in the online handwritten Gurmukhi script recognition are

- complex structure of the script,
- similar shape of characters,
- huge variation in handwriting styles,
- formation of a character by combining the strokes, *etc*.

The post-processing phase of online handwriting recognition system deals with character formation, which plays a crucial role in order to improve the accuracy of the recognition system. In the literature, the researchers have

---

*For correspondence