3. Development of OHWR System for Gurmukhi

R.K. Sharma, Palki Sharma, Narpinder Sharma, Harjeet Singh, Karun Verma,
Ravinder Kumar, Rajesh Kumar,
School of Mathematics & Computer Applications,
Thapar University, Patiala, Punjab (INDIA)

Abstract

Handwritten character recognition is a complex task owing to various writing styles of different individuals. A number of authors have worked on the problem of handwritten character recognition. They have, in general, used structural and statistical features in their work. Handwritten character recognition systems have also been proposed for Gurmukhi script by some authors. This work presents a system to recognize online handwritten Gurmukhi characters, Gurmukhi numerals, Roman numerals and special characters. A sufficiently large annotated database has been created in this work for the strokes used in writing these Gurmukhi symbols. The recognition engines developed in this work are based on 100 samples (Engine 1) and 22 samples (Engines 2, 3 and 4) of each class collected from different writers. These samples have been preprocessed and annotated. LibSVM classifier has been used in this work for classification purpose. The system presented in this work could attain the highest recognition accuracy of 96.8% at 5-fold cross validation for Gurmukhi characters, of 90.0% at 4-fold cross validation for Gurmukhi numerals, of 90.2% at 5-fold cross validation for Gurmukhi numerals and Roman numerals, of 84.0% at 4-fold cross validation for special characters, Gurmukhi and Roman numerals. Postprocessing of classes has also been performed in this work in order to refine the recognition process results. In this postprocessing, the sequence of classes is analyzed; overwritten strokes are identified and resolved. A heuristic based algorithm has been developed to form Gurmukhi aksharas from the identified classes.

1. Introduction

This is a well-established fact that handwritten character recognition is a complex task. This is due to different handwriting styles of individuals, and also the cursiveness in their handwriting. Handwritten character recognition is divided into two categories: offline and online. In offline handwritten character recognition, data are scanned images taken from a prewritten text, usually on a sheet of paper. In online handwriting recognition systems, data are captured while writing with the help of a special pen and an electronic surface. These systems generally include the phases of preprocessing, features extraction and classification. Feature extraction is a very significant phase in a character recognition system, which is used to decide for the relevant shape contained in the character. The performance of a character recognition system largely depends on the features, which have been extracted.