Hand-Filled Form Processing System for Gurmukhi Script

Dharam Veer Sharma and Gurpreet Singh Lehal

Department of Computer Science, Punjabi University, Patiala
dveer72@hotmail.com, gslehal@gmail.com

Abstract. A form processing system improves efficiency of data entry and analyses in offices using state-of-the-art technology. It typically consists of several sequential tasks or functional components viz. form designing, form template registration, field isolation, bounding box removal or colour dropout, field-image extraction, segmentation, feature-extraction from the field-image, field-recognition. The major challenges for a form processing system are large quantity of forms and large variety of writing styles of different individuals.

Some of the Indian scripts have very complex structures e.g. Gurmukhi, Devnagri and Bengali etc. Use of head line, appearance of vowels, parts of vowel or half characters over headline and below the normal characters (in foot) and compound characters makes the segmentation and consequently recognition tasks very difficult.

The present system is a pioneering effort for developing a form processing system for any of the Indian languages. The system covers form template generation, form image scanning and digitization, pre-processing, feature extraction, classification and post-processing. Pre-processing covers form level skew detection, field data extraction by field frame boundary removal, field segmentation, word level skew correction, word segmentation, character level slant correction and size normalization. For feature extraction Zoning, DDD and Gabor filter have been use and for for classification, kNN and SVM have been put to use. A new method has been developed for post processing based on the shape similarity of handwritten characters.

The results of using kNN classifier for different values of k with all features combined are 72.64 percent for alphabets and 93.00 percent for digits. With SVM as classifier and all the features combined, the results improve marginally (73.63 percent for alphabets and 94.83 percent for digits). In this demo we shall demonstrate the working of the whole system.

Keywords: Hand-filled form processing system, form processing system, Gurmukhi script, OCR.