Recognition of Isolated Handwritten Characters in Gurmukhi Script

Abstract

Isolated handwritten character recognition has been the subject of intensive research during last decades because it is useful in wide range of real world problems. It also provides a solution for processing large volumes of data automatically. Work has been done in recognizing handwritten characters in many languages like Chinese, Arabic, Devnagari, Urdu and English. The work presented in this thesis, focuses on the problem of recognition of isolated handwritten characters in Gurmukhi script. The whole process consists of two stages. The first, feature extraction stage analyzes the set of isolated characters and selects a set of features that can be used to uniquely identify characters. The performance of recognition system depends heavily on what features are being used.

The selection of stable and representative set of features is the heart of
recognition system. The feature extraction method Zoning, is used for extracting features of the character under consideration in this problem. In Zoning method, the frame containing the character is divided into several overlapping or non-overlapping zones and the densities of object pixels in each zone are calculated. Densities are used to form a representation. The final, classification stage is the main decision making stage of the recognition system. It uses features extracted in the feature extraction stage to identify the character. K-Nearest Neighbor and Support Vector Machine are the two classifiers used for identifying the character in the problem. In k-nearest classification method, the Euclidean distance between the test point and all the reference points is calculated in order to find K nearest neighbors, and then the obtained distances are ranked in ascending order and the reference points corresponding to the k smallest Euclidean distances are taken. The Support Vector Machine (SVM) is a learning machine with very good generalization ability. SVM implements the Structural Risk Minimization Principal which seeks to minimize an upper bound of the generalization error. An SVM classifier discriminates two classes of feature vectors by generating hyper-surfaces in the feature space, which are "optimal" in a specific sense that is the hyper-surface obtained by the SVM optimization is guaranteed to have the maximum distance to the "nearest" support vectors. SVM operate on kernel evaluations of the feature vectors. An annotated sample image database of isolated handwritten characters in Gurmukhi script has been prepared which has been used for training and testing of the system.

Reference

- Beatrice Lazzerini, Francesco Marcelloni, "Fuzzy classification of handwritten characters", 18th International Conference of the North American, pp. 566 – 570, 1999
- Nozomu Araki, Miyuki Okuzaki, Yasuo Konishi and Hiroyuki Ishigaki, "A Statistical Approach for Handwritten Character Recognition Using Bayesian Filter", 3rd International
- Jianming Hu; Hong Yan, "Structural decomposition and description of printed and handwritten characters", Proceedings of 13th International Conference on Pattern Recognition, pp.230-234, 1996
- Christopher J.C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition”, pp.1-43

Index Terms
Computer Science Pattern Recognition

Key words
Isolated Handwritten Gurmukhi Script method Zoning