

## IMPROVED STATISTICAL FEATURES FOR CURSIVE CHARACTER RECOGNITION

TANZILA SABA, AMJAD REHMAN AND GHAZALI SULONG\*

Department of Computer Graphics and Multimedia  
Faculty of Computer Science and Information Systems  
Universiti Teknologi Malaysia  
Skudai Johor, Malaysia

stanzila2@live.utm.my; rkamjad@utm.my

\*Corresponding Author: ghazalis@utm.my

Received April 2010; revised October 2010

**ABSTRACT.** *This paper presents an improved feature extraction technique for the cursive characters recognition. This technique can be applied in the perspective of handwritten word recognition system based on segmentation. The bases of fused statistical features extraction technique are improved projection profile and transition features. To extend this principal, a technique is integrated with the projection profile information to detect shifts of background and foreground pixels in the image of a character. A classifier based on neural network is used to test the improved fused features and comparison is done with the projection profile (PP) and transition feature (TF) extraction techniques. By using standard dataset, PP and TF techniques altogether show best performance with fused features having new enhancements and the best results in the literature are compared promisingly with this technique. The characters that are taken from the CEDAR dataset show 91.38% recognition accuracy.*

**Keywords:** OCR, Pattern recognition, Computer vision, Machine learning, Feature extraction

**1. Introduction.** The key tool that permits machines to accurately identify the hand-written material emblazoned by humans is the handwriting recognition. The theory is divided into two major areas. Offline and online are two very distinct forms of handwriting recognition input. Offline handwriting recognition refers to the process of script recognition from a static surface such as paper and stored digitally in grey-scale format. In the online case, the handwriting is captured and stored in digital form via different means. Usually an electronic pen and surface (PDA) is used. Along the pen movements on the surface, the two-dimensional coordinates of successive points are represented as a function of time and stored in order [1]. Features extraction from online character patterns is simpler than offline character images. It is mainly due to reasons that most information is collected in online case such as the direction, time, speed and order of strokes of the handwriting. However, in case of offline handwriting, such information is hard to attain.

The existing main approaches for offline script recognition can be categorized into analytical (segmentation based) and holistic (segmentation free). Generally, analytical approaches employ a segmentation based strategy for recognition of segmented characters while holistic approaches treat the word image as an indivisible unit [2]. Still now, the recognition of cursive handwriting is a main issue in the research. The reasons of continued motivation are accredited due to the demanding nature of the dilemma and also its applications in various commercial fields [3]. Currently, the most particular areas having much interest in its application are postal address recognition [4], bank cheque processing