Abstract No: SC-23   **Smart Computing**

Isolated Sinhala handwritten character recognition using part based matching technique

*C. M. Silva[1] and N. D. Jayasundere[2]

[1]Sri Lanka Institute of Information Technology, Sri Lanka
[2]Department of Electrical and Information Engineering,
University of Ruhuna, Sri Lanka
Email: *chamari.s@sliit.lk

Abstract

This paper presents a novel approach for Sinhala handwritten character recognition using a part based matching technique. The Sinhala character set consist of some common parts. Therefore, the characters can be split into its parts. Each part in turn, can be considered as an atomic element, which these characters are composed of. The proposed method splits the characters into their atomic parts and then conducts the recognition process. Template matching is used to compare the character parts and characters. To improve the recognition process, the global characteristics of the characters are used. Experimental results show that the proposed method gives an average accuracy of 46% where the maximum accuracy is 100% and the minimum accuracy is 11%.

Keywords: Part-based approach, Sinhala characters

Introduction

Character recognition is a procedure of converting images of handwritten, typewritten or printed text into machine encoded code or text (Schantz, 1982). Most of the hand written character recognition methods have been proposed for the recognition of scripts such as English and Chinese. Few attempts have been taken on the Asian languages such as Sinhala. Sinhala is the language used by Sinhala people, the major ethnic group in Sri Lanka. The characters in the Sinhala alphabet share some common parts. Hence, Sinhala characters can be split into a set of parts, which can be considered as the basic elements in which these characters are composed of. Each Sinhala character can be formed by a set of distinct parts.

Part-based approaches have been experimented for object recognition with promising results. Most of the work consider the problem of matching corresponding parts of objects across different images. The part-based methods can be applied for the experiments in character recognition as well. There is no research done in the area of evaluating the influence of decomposing Sinhala characters, recognizing them in a part wise manner and combining the results to character recognition.

Most of the related work for decomposing characters has been experimented on Chinese characters (Cao and Tan, 2000), (Lin and Tang, 2002). The reason is that the Chinese characters can be easily decomposed to a set of basic character parts. Because of the nature of using straight lines in Chinese characters some researches have used the method of identifying strokes of characters (Lin and Tang, 2002), (Su, and Wang, 2003). This method is not applicable for Sinhala characters as the straight lines are almost non-existent in the Sinhala character set. The work of Matsuo, et al. (2013) is more relevant to the work of the present research. They work on Chinese characters but they decompose the characters in to parts, which are short segments of an entire character. Then they represent each part as a segment comprised of