Abstract

Document image segmentation is one of the critical phases in handwritten character recognition system. Correct segmentation of individual characters decides the accuracy of the recognition system. It is used to decompose the sequence of characters into individual characters to segmenting text lines and then words. Ancient Tamil scripts documents consist of vowels, consonants and various modifiers. Hence proper segmentation algorithm is required. In existing methods, segmentation of overlapping lines and characters are difficult. In order to overcome this problem, two methods are proposed one for line segmentation and another for character segmentation, first method uses projection profile and PSO for line segmentation. In second method combination of connected components along with nearest neighborhood methods are used to segment the characters. Experimental results show that these methods give better results when compared to other methods.

References

- Raghuraj Singh, S. Yadav and Prabhat Verma" Optical Character Recognition (OCR) for Printed Devnagari Script Using Artificial Neural Network"; , International Journal
- C V Lakshmi, C PAtardhan &quot;A Multi-font OCR System for printed Telugu Text. &quot; Proceeding of LECEapos;02, IEEE, 2002
- Itay Bar-Yosef et, al, "Line segmentation for degraded handwritten historical documents& quot;

**Index Terms**

Computer Science  
Pattern Recognition
Keywords
Character segmentation  Projection profile  connected components  nearest neighborhood  PSO