Abstract

In this paper, compression scheme is presented for Indian Language handwritten text document images. Document image compression is an active area of research. Current OCR technology is not effective for handling the handwritten text images. The proposed compression scheme deals with the handwritten gray level document in Devnagri script. The method is based on the separation of foreground and background of an image and connected component labeling. Experiments are done with handwritten images in Devnagri (Hindi and Marathi). Compression schemes are available for the printed text in Indian language. But there is little work reported towards the compression standards for handwritten text image. The results of the modules are showing good compression ratio. Hence compression of handwritten text images in Indian language is important.
CONCLUSION

In this paper compression strategy of handwritten text for Indian language grey level document is presented. To the best of our knowledge, this is the first effort towards the compression of handwritten text for Indian language documents. As mentioned earlier, most of the work is done for foreign language handwritten document. The proposed methodology only focuses on the grey level handwritten images. The compression technique presented here is lossless in nature. The compression technique works well and gives results accordingly. It can be extended for the color document also. This aspect is left for future extension of current study.


- Christophe Fiorio and Jens Gusted. 1996 Two linear time union-find strategies for image processing. Theoretical computer sci.
- Benarad A. Galler and Michel Fisher. 1964 An improved equivalence algorithm. Communication on ACM.

**Index Terms**

Computer Science  
Image Processing
Keywords
Handwritten Text  Connected Component Labeling Compression  Indian Language
Devnagri Script
Gray Level Document