Abstract

The present growth of digitization of books and manuscripts demands an immediate solution to access them electronically. This requires research in the area of document image understanding, specifically in the area of document image retrieval. There is an immense scope for such a retrieval system for a digital library of Kannada Document Images. This paper presents an efficient Content Based Document Image Retrieval system for a Kannada document image collection. A recognition-free approach is followed because recognition based approach is inefficient in terms of performance. The data is pre-processed and segmented for faster matching and retrieval. An efficient search technique - Correlation method is used to search in large collection of document images. Performance evaluation using different datasets of Kannada documents shows the effectiveness of the approach.

References

- David Shen, Zaizai Lu "Computation of Correlation Coefficient and Its Confidence Interval in SAS";
- E. Garcia "A Tutorial on Correlation Coefficients";

- Ramachandra Manthalkar and P. K. Biswas, "An Automatic Script Identification
Scheme for Indian Languages"; NCC, 2002.
- R Sanjeev Kunte, R D Sudhaker Samuel. , 2007. An OCR system for printed Kannada
International Conference on Computational Intelligence and Multimedia Applications (IEEE
Computer Society Press. ), 349-353.
- Siddhaling Urolagin, Prema K. V, V. Subba Reddy &quot;A Gabor Filters Based Method
for Segmenting Inflected Characters of Kannada Script&quot; 978-1-4244-6653-5/10/ ©2010
IEEE
- VijayaKumar B, Ramakrishnan A G 2004 &quot;Radial basis function and sub-space

- Wahl F. M., Wong, K. Y., Casey R. G.: &apos;&apos;Block Segmentation and Text
Extraction in Mixed Text/Image Documents&apos;&apos;; Computer Graphics and Image
Processing, 20 (19 82) 375- 390.

Index Terms

Computer Science Image Processing

Keywords

Content based image retrieval Correlation coefficient Median Filtering Kannada
Document Images

Segmentation.