Abstract

In this paper, database for isolated handwritten Marathi simple words (not contains compound character known as 'Jodakshare') was developed. Commonly used 50 words are chosen and total 20210 handwritten Marathi word samples database was developed. Generalized segmentation methodology is proposed here which is applicable to any simple, handwritten Marathi word containing any number of characters. The segmentation algorithm first detects header cap ('Shirorekha') that separates top modifiers and core area of the word. Statistical information and vertical projection is used for further segmentation process. The segmentation algorithm proposed here is applicable to any Marathi plain word having any number of characters also can equally applicable to many other languages like
Hindi, Sanskrit, Nepali and Konkani which are similar in structure. Using proposed algorithm maximum 95.31 percent segmentation is achieved for ‘Daar’ word and average segmentation achieved is 82.17 percent.

References

- Laurent Guichard Alejandro H. Toselli Bertrand Couasnon. Handwritten Word Verification By SVM-Based Hypotheses Re-Scoring And Multiple Thresholds Rejection. 12th International Conference On Frontiers In Handwriting Recognition 978-0-7695-4221-8/10 $26.00 IEEE 2010
- RAJIV KUMAR, AMRESH KUMAR, PERVEZ AHMED. A Benchmark Dataset For
Devnagari Document Recognition Research. Recent Advances In Telecommunications, Signals And Systems
- Veena Bansal And R. M. K. Sinha. Segmentation Of Touching And Fused Devanagari Characters.
- Veena Bansal And R. M. K. Sinha. Segmentation Of Touching Characters In Devanagari.

Index Terms
Computer Science Pattern Recognition

Keywords
Handwritten Marathi Word Segmentation Image Vertical Projection Database Development