



OUR HERITAGE

ISSN: 0474-9030 Vol-68, Special Issue-27 (Feb. 2020)
5th International Conference On "Innovations in IT and
Management"



Organised by: Sinhgad Technical Education Society's
SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION
(SIMCA), Narhe Technical Campus, Pune, Maharashtra (India) 411041.

Held on 6th & 7th February 2020

Printed Devanagari Character Standard Dataset having different font size and styles variations.

Arvind A. Tayade
G.S. Science, Arts & Commerce College, Khamgaon, Dist-Buldana-444303
Email- arvindtayade40@gmail.com

Abstract

This paper represents the new Devanagari character dataset having the various parameters. Character recognition is a huge area for research in which the several peoples are worked on the various topic and parameters. Some researcher work on handwritten character recognition with various languages. In the proposed paper, we have published the printed character dataset for Devanagari characters. In this dataset, the Total number of character images 1515 of 39 various character and recognised those images according to mid bar, end bar and no bar characters.

Keywords : *Character Recognition, Dataset, Devanagari characters.*

Introductions:

Devanagari script is the most important and widely used script in India. It is the script used by many Indian languages like Hindi, Marathi, Nepali and Sanskrit. Several other languages like Punjabi, Kashmiri use close variations of this script. It was also formerly used to write Gujrati [1]. The character recognition is the applicable in the various areas like postal address recognition, optical character recognition, character to audio translation, automatic currency recognition and several such a types of research are relates to character recognition. In which the regional language detection and recognition is also a huge area for researchers. It is a challenge to researcher how to manipulate their features due to its similarity of characters or size and shape of characters. There are around 50 basic characters in the script. The grouping of vowels and consonants is called Swaras and Vyanjanas respectively. The grouping of vowels and consonants in Devanagari is done on the basis of phonetic point of articulation [1]. Script has its own specified coin position rules for combining vowels, consonants and modifiers, some of them can be combined with their type as well [2]. It is big challenges to recognizing the language. From such types of challenges we have go for making the database having different font size and style. To improve the recognition accuracy with those challenges the slandered dataset required for research. But in the case of



OUR HERITAGE

ISSN: 0474-9030 Vol-68, Special Issue-27 (Feb. 2020)
5th International Conference On “Innovations in IT and
Management”



Organised by: Sinhgad Technical Education Society's
SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION
(SIMCA), Narhe Technical Campus, Pune, Maharashtra (India) 411041.

Held on 6th & 7th February 2020

regional languages, no any standard dataset is available. To overcome such types of limitations of recognition a characters we have need the various standard database with experimental variations. So we have created the dataset with the various variations like different font size and shape or style.

Literature survey:

The lots of researcher work on the such types of work having the different techniques from such a researcher we have analyze some highly reputed papers. According to [3] introduce a new public image dataset for Devanagari script: Devanagari Handwritten Character Dataset (DHCD). Dataset consists of 92 thousand images of 46 different classes of characters of Devanagari script segmented from handwritten documents. We also explore the challenges in recognition of Devanagari characters. Along with the dataset, we also propose a deep learning architecture for recognition of those characters. The limitation of this research it does not work on the different font size and style. These challenges covered in our dataset. The schema for description of shapes of Devanagari characters and its application in their recognition[4]. It exploits certain features of the script in both reducing the search space and creating a reference with respect to correspondence could be established, during the matching process.

Methodology:

In Devanagari character recognition, standard printed character database is not available for finding the recognition accuracy of proposed method created a typed printed character database with 10 different font style i.e. DVB-TTBhima, DVB-TTChaya, DVB-TTDhruva, DVB-TTGanesh, DVB-TTNataraj, DVB-TTRadhika, DVB-TTRaghav, DVB-TTShridhar, DVB-TTSurekh and DVB-TTVasundhana this fonts taken with the help of Gist ISM 3.0 software. Take 5 different font size of each font style for database creation. The file format of each image is monochrome bitmap image having dimension 50 px X 50 px with the help of MS-Paint Brush application. Classify database according to end bar character, mid bar character and no bar characters.

Following figure shows the database of end bar character alphabets with different font style and font size.

Devanagari Numerals Database with different font style and font size





OUR HERITAGE

ISSN: 0474-9030 Vol-68, Special Issue-27 (Feb. 2020)
5th International Conference On "Innovations in IT and
Management"



Organised by: Sinhgad Technical Education Society's
SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION
(SIMCA), Narhe Technical Campus, Pune, Maharashtra (India) 411041.

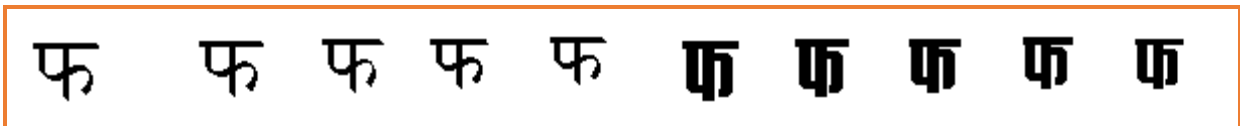
Held on 6th & 7th February 2020



Devanagari end bar Character Database with different font style and font size



Devanagari Mid bar Character Database with different font style and font size





OUR HERITAGE

ISSN: 0474-9030 Vol-68, Special Issue-27 (Feb. 2020)
5th International Conference On "Innovations in IT and



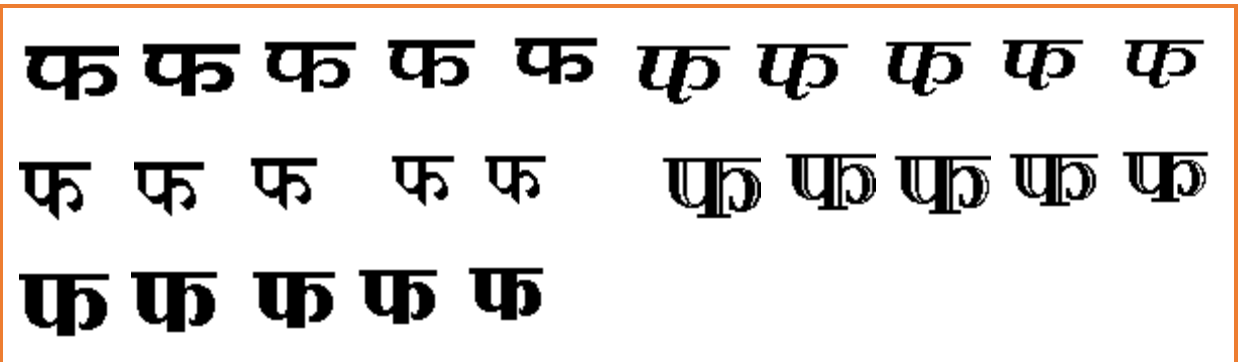
Management"

Organised by: Sinhgad Technical Education Society's

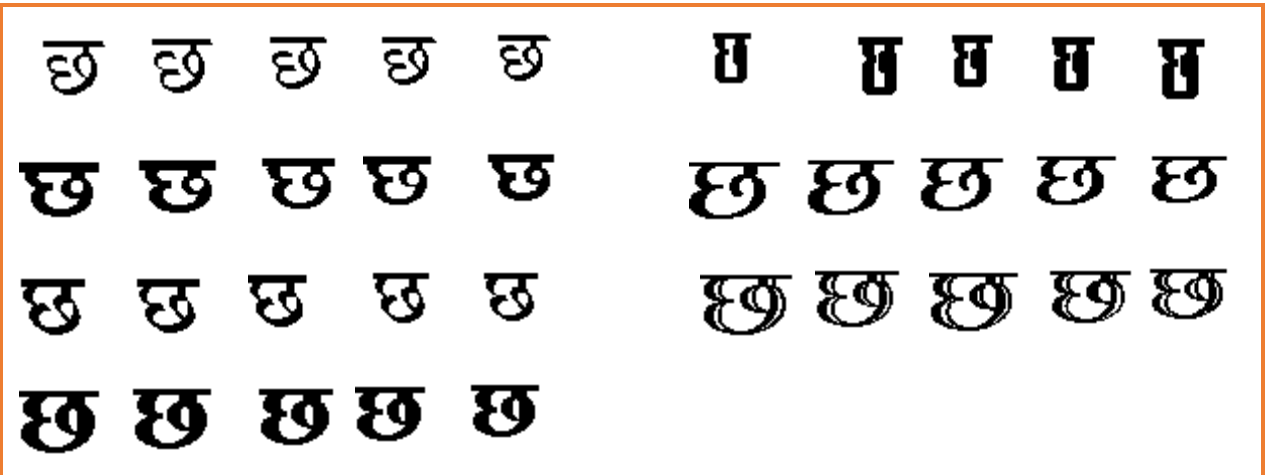
SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION

(SIMCA), Narhe Technical Campus, Pune, Maharashtra (India) 411041.

Held on 6th & 7th February 2020



Devanagari No bar Character Database with different font style and font size



Conclusion:

This database useful to analyzed the challenges of character recognition, specifically Devanagari character recognition. There is nothing any standard database to finding the character recognition accuracy. This database is useful to learn and identify automatic recognitions of character because of it is a printed character datasets. This dataset is useful to those researchers who work on the character recognition area as a standard data base. The computation is reduced because of the differentiation of no bar, end bar and mid bar characters. Due to its wideness the researcher have to take more compensations for marching that's why its help to release the limitation problem of character recognition.

References



OUR HERITAGE

ISSN: 0474-9030 Vol-68, Special Issue-27 (Feb. 2020)
5th International Conference On "Innovations in IT and
Management"



Organised by: Sinhgad Technical Education Society's
SINHGAD INSTITUTE OF MANAGEMENT AND COMPUTER APPLICATION
(SIMCA), Narhe Technical Campus, Pune, Maharashtra (India) 411041.

Held on 6th & 7th February 2020

- [1] M. Saifee, "Devanagari Font Design for Optical Character Recognition," INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, Mumbai, 2012.
- [2] R. M. K. Sinha and B. Vina, "ON DEVANAGARI DOCUMENT PROCESSING," IEEE, pp. 1621-1626, 1995 .
- [3] S. Acharya, P. A. Kumar and G. P. Kumar, "Deep Learning Based Large Scale Handwritten Devanagari Character Recognition," 9th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), 2015.
- [4] V. Bansal and S. M. K., "On How to Describe Shapes of Devanagari Characters and Use Them for Recognition," IEEE.
- [5] "Digital image processing using Matlab fundamentals," www.eie.polyu.edu, [Online]. Available: <http://www.eie.polyu.edu.hk/~enyhchan/imagef.pdf> . [Accessed 30 12 2016].