RESEARCHERID
THOMSON REUTERS

[ICEMESM-18]
ICTM Value: 3.00

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

CONVERSION OF HINDI BRAILLE TO TEXT IN HINDI LANGUAGE

Rasika S. Gargey, Rohini V.Amale, PravinKhawse, HarshwardhanKharpate
Cummins College of Engineering for Women, Nagpur, Maharashtra, India

DOI: 10.5281/zenodo.1229000

ABSTRACT

The Braille system has been used by the visually impaired for reading and writing. Due to limited availability of the Braille text books an efficient usage of the books becomes a necessity. This paper proposes a method to convert a scanned Braille document to text which can be read out to many through the computer. The Braille documents are pre-processed to enhance the dots and reduce the noise. The Braille cells are segmented and the dots from each cell is extracted and converted in to a number sequence. These are mapped to the appropriate alphabets of the language. The Braille cell has a standard representation but the mapping differs for each language. In this paper mapping of Hindi is considered.

Keywords: Braille Conversion, Hindi Braille conversion, Image Segmentation

I. INTRODUCTION

Visually impaired people are an integral part of the society. However, their disabilities have made them to have less access to computers and Internet than the people with clear vision. Over time Braille system has been used by them for written communication. Braille is a system of writing that uses patterns of raised dots to inscribe characters on paper [1]. This allows visually impaired people to read and write using touch instead of vision. It is the way for blind people to participate in a literate culture. First developed in the nineteenth century, Braille has become the pre-eminent tactile alphabet. Its characters are six- dot cells, two wide by three tall. Each dot may exist or may not exist giving two possibilities for each dot cell. Any of the six dots may or may not be raised; giving 64 possible characters.

Although Braille cells are used world-wide, the meaning of each cell depend on the language that they are being used to depict.Braille can be seen as the world's first binary encoding scheme for representing the characters of a writing system. However, very limited numbers of Braille books are available for usage.

Printing of Braille books is a time consuming process. The requirement of special printers and software add to their limited availability. Scanned and text converted documents can be used in the meantime to serve the needs of the blind. This paper mainly focuses on conversion of a Braille document into its corresponding alphabets in Hindi using various concepts of image processing. The presence of dots in the Braille cells has to be identified to recognize the characters. The edge detection when applied on the scanned document will not produce the dots, hence the approximate intensity range of the dots are identified from the histogram. The image is treated through a sequence of enhancement steps which increases the contrast between the dots and the background. The edge detectors are then applied and the text area is cropped excluding the borders through projection profile method. The document is then segmented into Braille cells using standard Braille measurements and projection profiles. The presence of dots in each cell is identified using a Threshold and converted to Binary sequence which is then mapped to the corresponding language alphabet.

Section 2 of this paper discusses about few commercial systems available for converting Braille to text or vice-versa. Section 3 explains about the conversion of Braille to text; section 4 covers the outcomes of the proposed method. Section 5 concludes the paper.

II. LITERATURE SURVEY

A Braille translator is a software program that translates a script into Braille cells, and sends it to a Braille embosser, which produces a hard copy in Braille script of the original text. Basically only the script is

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

transformed, not the language. One of the general purpose translators is text to Braille converter. Few other commercial translators are also available such as:

2.1 Tesseract

[2]Tesseract is an optical character recognition engine for various operating systems. It is free software, released under the Apache License, Version 2.0, and development has been sponsored by Google since 2006. In 2006 Tesseract was considered one of the most accurate open-source OCR engines then available.

2.2 Bharati Braille

[3]This converter transforms *Hindi* and *Marathi* text in the *Devanagari* script to *Bharati*Braille based on the rules outlined in *Bharati Braille Shikshak*, a publication by India's National Association for the Blind. The converter does not perform Brailletypesetting.

2.3 Other Open Source OCR's[4]

1. Google's & HP's Tesseract
2. Google's Keep
3. Microsoft Office Document Imaging (MODI)
4. Microsoft One Note
5. Microsoft Oxford Project API (This API is free until sometime)
6. FreeOCR

III. BRAILLE CONVERSION

For converting the Braille document to text, the input is taken as a scanned Braille document. The Braille character is extracted and matched with the corresponding alphabet.

3.1. Conversion of scanned Braille document

In this method the Braille document is scanned and taken as input, which by a sequence of steps is converted to appropriate text. The scanned document has to be enhanced to identify the dots clearly. The dots are extracted using horizontal and vertical profiling. The Braille cells are identified and converted to binary sequence. The binary sequence is then mapped to the corresponding alphabets or contracted words. Output will be obtained in file format. Utmost care is taken to ensure that unwanted noise or redundant information is not introduced at the time of scanning. The scanned image is then converted to grey scale image.

3.1.1. Image enhancement

Due to scanning, the dots in the Braille document cannot be distinguished clearly from the background. Hence various pre-processing techniques are applied on the scanned image in order to enhance the dots and to suppress the noise. The dots appear as a darker shade of the back ground colour and hence these intensity ranges are identified from the Histogram and enhanced in order to identify the dots. Piece wise enhancement techniques such as contrast stretching, intensity stretching were used for enhancing the dots. The limits over which image intensity values will be extended are decided from the histogram of the input image. Another level enhancement is done to the dots using Intensity adjustment. This is an image enhancement technique that maps the intensity values of an image to a new range.

3.1.2. Image filtering

To remove the unwanted noisy dots present in the scanned documents, the image is smoothened using Gaussian filter. The edges mostly correspond to the dots of the Braille cells. The border of scanned document and the stapler pin information if any present in the document are removed through image cropping.

3.1.3. Segmenting the Braille Cells

In order to simplify the process of Braille character extraction, the image is first segmented into lines and then into Braille cells. Each cell is further partitioned into binary dot patterns. These are achieved through Projection profiles and standard Braille measurements. Horizontal profiling is performed on edge detected image and zero profile indicates the absence of dots. Among many such lines, the first line from the top that is closer to the dots is taken as reference. The standard vertical distance between two Braille cells is used to draw the remaining lines where the X projection is zero. This procedure is repeated till the end of the document. After extracting horizontal lines of the Braille cell, a vertical profiling is performed. Among many such lines the leftmost one that is closer to the dots is taken as reference. The standard horizontal distance between two Braille cells is used

RESEARCHERID
THOMSON REUTERS
[ICEMESM-18]
ICTM Value: 3.00
ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

to draw the remaining lines where the Y projection is zero. This procedure is repeated till the end of the document. This segments the edge image in to Braille cells. Each segmented cell is divided into 3 x 2 grids using the standard Braille distance between two dots in a cell[1].

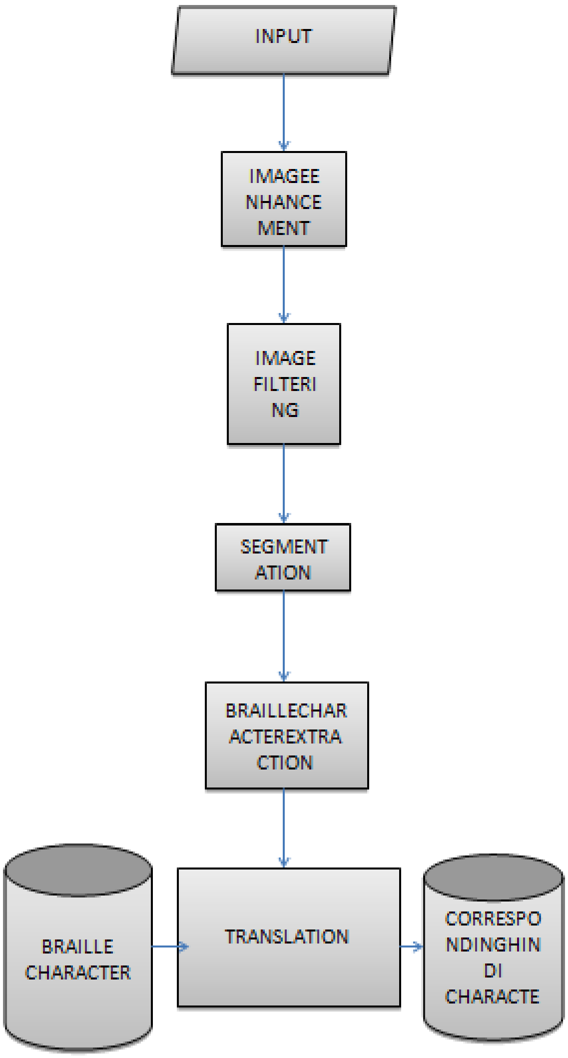

Fig1 : Block Diagram of Proposed Technique

3.1.4. Extraction of Text from pattern vector

A Binary pattern vector for each Braille cell is generated. A vector has a length of 6 each correspond to a dot in the Braille cell. The presence of dot is identified after counting the number of white pixels in each grid of a cell and checking whether it satisfies the threshold criterion. '1' indicates that dot is present and '0' indicates that dot is absent in that particular position. This string of bits for the sequence of Braille alphabets is written into a file. A sequence of 6 bits are read from the file and converted to the number sequence and subsequently into the alphabet. If the six bits of the string are 0's, it generates a space. These alphabets are stored in a text file for further processing. For Hindi, the sequences of 6 bits are taken and the corresponding Unicodes are generated using its pre-built mapping table. These Unicodes are stored in the file. The obtained file is converted to the corresponding Hindi text.

4. EXPERIMENTAL ANALYSIS

Currently, the work has been done to read an image containing a word in Hindi Braille Script. The image is then read and mapped accordingly to its corresponding alphabest in Hindi language. After completion of this task, we

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

will combine such images of Braille words to make a sentence and translate it. We will then move to conversion of paragraphs.

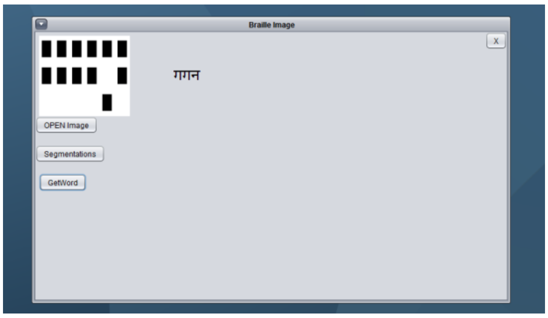

Fig2 : Conversion of braille text image to corresponding Hindi alphabet

IV. LIMITATIONS AND ADVANTAGES

Since the standard Braille dimensions are used for the segmentation of the Braille cells, the document has to be free from tilt and has to be aligned with the edge of the scanner. This poses a major limitation to the system. The presence of the unnecessary dots or noises whose size is comparable to that of the Braille dots during scanning is difficult to remove during pre-processing and hence affects the accuracy of the converted text. Since, Hindi Braille cannot depict the use of sounds (*Matras*) and half letters, conversion of this can also not depict those sounds and half words. Thus, it shows a complete letter.

It involves very less intervention of the user and helps to serve the need of large number of people using a single document. It helps resource teachers in Inclusive Education, who do not know Braille. Simplifies making of copies of old Braille books for which only one copy is available as it saves the labour of preparing the same again. Since the availability of Braille document is also limited, scanning the document also help in preserving the existing documents.

V. CONCLUSION

This paper focuses on the conversion of scanned Braille documents to corresponding text in Hindi. After identifying the start of the Braille text, the lines and subsequently the Braille cell are segmented. Grids are drawn based on the standard measurement of the Braille cells and the dots are extracted. Braille has a standard pattern of alphabets and only the mapping differs from language to language. Using appropriate mapping for each language the alphabets are identified and stored as text.

The extraction of the dots was affected when they were not confined to the standard measurement and due to the presence of noise during scanning. Mapping errors occurred when the Braille has similar representation for the alphabet and the punctuation. These are eliminated to some extent using simple rules governing the language

VI. REFERENCES

[1] S.Padmavathi,Manojna K.S.S, Sphoorthy Reddy .S and Meenakshy.D,"Conversion of braille to text in english, hindi and tamil languages".
[2] https://en.wikipedia.org/wiki/Tesseract_(software)
[3] https://bharati-braille.pareidolic.in/about.html
[4] https://www.quora.com/What-are-the-best-open-source-OCR-libraries
[5] Saad D. Al-Shamma and Sami Fathi, "Arabic Braille Recognition and Transcription into Text and Voice", 2010 5th Cairo International Biomedical Engineering Conference Cairo, Egypt, December 16-18, 2010
[6] AbdulMalik S. Al-Salman, "A Bi-directional Bi-Lingual Translation Braille-Text System", J. King Saud University, Vol. 20, Comp. & Info. Sci., Riyadh(1428H./2008).
[7] Charanya C, Kalpana S and Nithya R, "Real time Braille recognition with sonic feedback", Intel India Research Challenge 2007
[8] R.T. Ritchings, A. Antonacopoulos and D. Drakopoulos,"ANALYSIS OF SCANNED BRAILLE DOCUMENTS",In the book: Document Analysis Systems, A. Dengel and A.L. Spitz (eds.), World Scientific Publishing Co,1995

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

[9] Rawan Ismail Zaghloul,TomaderJameelBani-Ata,"Braille Recognition System – With a Case Study Arabic Braille Documents",European Journal of Scientific Research ISSN 1450-216X Vol.62 No.1 (2011)

[10] AbdulMalik,S. Al-Salman,"A Bi-directional Bi-Lingual Translation Braille-Text System", Journal of King Saud University - Computer and Information Sciences Volume 20, 2008