

A New Research Resource for Optical Recognition of Embossed and Hand-Punched Hindi Devanagari Braille Characters: Bharati Braille Bank

Shreekanth.T

Research Scholar, JSS Research Foundation, Mysore, India.
Email: speak2shree@gmail.com

V.Udayashankara

Professor, Department of IT, SJCE, Mysore, India.
Email: v_odayashankara@yahoo.co.in

Abstract—To develop a Braille recognition system, it is required to have the stored images of Braille sheets. This paper describes a method and also the challenges of building the corpora for Hindi Devanagari Braille. A few Braille databases and commercial software's are obtainable for English and Arabic Braille languages, but none for Indian Braille which is popularly known as Bharathi Braille. However, the size and scope of the English and Arabic Braille language databases are limited. Researchers frequently develop and self-evaluate their algorithm based on the same private data set and report its behavior using ad-hoc measures of performance. There is no well-defined benchmark database for comparative performance evaluation of results obtained. The developed Braille database, *Bharati Braille-Bank*, is a large and well characterized information of Braille documents and its related data for use by the research community working in Optical Braille Recognition (OBR) for Bharati Braille. In the present form it includes databases of embossed double sided, embossed single sided, skewed, Hand punched and images with varying resolutions of Hindi Braille. The objective of this work is to stimulate current research and new investigations in the study of Hindi OBR. Without common databases such as those provided by Braille Bank it is impossible to resolve certain contradictory research results. To overcome this problem, Braille Bank provides facilities for the comparative analysis of the data and the evaluation of proposed algorithms with the standard database. In addition, it provides free access to the developed database in the form of Compact Disc-Read Only Memory (CD-ROM). This work is a step forward in the direction of development of standards for Hindi Devanagari Braille data collection for Indian languages. The mission of the resource is to accelerate the development of OBR for Bharati Braille.

Index Terms—Bharati Braille, Grade-I, Hand punched, Hindi Devanagari, Inter-point, OBR

I. INTRODUCTION

Braille is a language for the blind to read and write through the sense of touch. Braille is formatted to a standard size by Frenchman Louis Braille in 1825. Braille is a system of raised dots arranged in cells. Any combination of one to six dots may be raised within each cell and the number and position of the raised dots within a cell convey to the reader the letter, word, number, or symbol the cell exemplifies. There are 64 possible combinations of raised dots within a single cell. The locus of the dot patterns for Braille are named universally by numbering the dots by 1 through 3 from top to bottom on the left and 4 through 6 from top to bottom on the right as shown in Fig. 1[1].

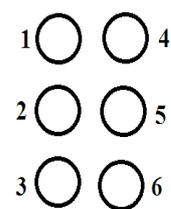


Fig. 1. Braille cell

A. Braille Dimension and Types of Braille

The library of congress have formulated the dot dimension, cell dimension, the distance between the dots within the cells, the distance between the cells in a word according to the tactile resolution of the fingertips of an individual. Dot height is approximately 0.02 inch (0.5 mm), the horizontal and vertical spacing between dot centers within a cell is approximately 0.1 inch (2.5 mm), the blank space between dots on adjacent cells is approximately 0.15 inch (3.75 mm) horizontally and 0.2 inch (5.0 mm) vertically as shown in Fig. 2(a). A standard Braille page is 11 by 11 inches and typically has

a maximum of 40 to 42 Braille cells per line and 25 lines per page [14].

There are two types of Braille documents based on embossing. The embossing of Braille dots can be on single side or on both the sides of the Braille sheet. The contents in the single sided Braille document are less, since the information present is only on one side. Thereby the volume of the Braille document will escalate. In order to reduce the volume of the Braille document, the embossing is done on both sides of the sheet. To avoid the merging of protrusions and depressions a slight diagonal offset is maintained while embossing the dots on both sides of the Braille documents depicted in Fig. 2(b) [7]. The translation of the double sided Braille document to text is a complex task, because the recognition techniques employed are based on the visual perception but not on the tactile sensing used by the visually impaired.

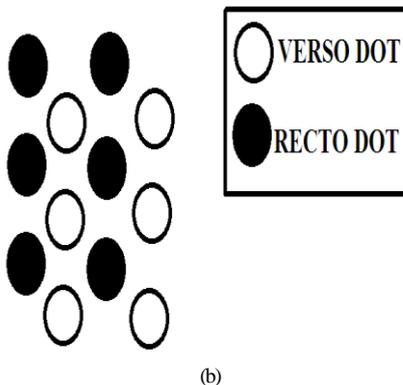
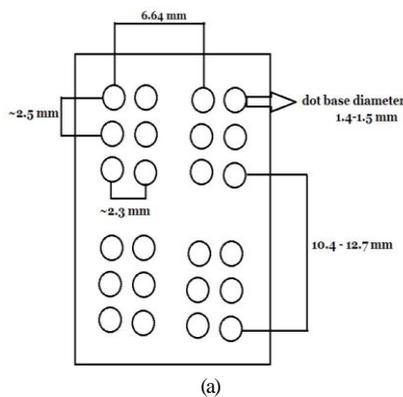


Fig. 2. (a) Braille cell dimension (b) Inter-point Braille

B. Grades of Braille

The Braille codes for Hindi and other languages use compressed codes to represent a word or a sentence instead of spelling all the letters occurring in a word. The meaning of each Braille character depends on the type of Braille encoding used. Due to the varying needs of Braille readers, there are three different grades of Braille.

In grade I Braille, each Braille cell corresponds to each character in the equivalent text. There is a one-to-one correspondence in the Braille representation and the printed text. Grade I Braille is bulkier and larger than the normal text. In grade II Braille, a cell can represent a

group of characters and it is a standardized form of contracted Braille. It is most common form of tactile writing. Grade III Braille is a complicated form of Braille and it is not standardized because each writer uses their own dialect of codes as a form of personal shorthand, which other Braille readers may not be able to read [25].

C. OBR and Need for OBR

OBR system is a computer software that automates the process of acquiring and processing the images of Braille documents. It is essential to transliterate the Braille documents to the text document in the corresponding language in order to establish a feasible bi-directional communication amid the visually impaired and the sighted community. There are a substantial number of longstanding Braille documents that need to be reproduced so that they can be preserved and accessed by more people; hence this concept of OBR has proved to be beneficial in attaining this requisite. Every person who wishes to work with blind people does not need to know how to read Braille, and these people can successfully communicate with the blind by using the OBR. The key reason for developing a system that can read Braille is to preserve and multiply large volumes of manually crafted books. It is undeniably difficult for even a skilled copyist to retype and reproduce the documents or the books on mathematics and music into Braille due to the specific rules that apply in Braille. In view of all these there is a need for OBR system for all the native languages.

D. Overview of Hindi Language

Hindi is the official language of India, it is spoken as first language by 33 percent of the Indian population, and by many more as a lingua franca. Also it is interesting to note that Hindi is the fourth-most-widely spoken language in the world.

Hindi is mainly written in Devanagari script. In the Hindi language there are 33 consonants and 13 vowels. The Hindi language has one to one correspondence with the spoken language and the written form. The phonemes are divided into two types, vowels (swara) and consonants (vyanjan).

Vowels (Swara): Vowels are the independently existing letters which are called swara.

अ आ इ ई उ ऊ ऋ ए ऐ ओ औ

Consonants (Vyanjan): Consonants are those which depend on vowels to form their independent letter.

क ख ग घ ङ,
च छ ज झ ञ,
ट ठ ड ढ ण,
त थ द ध न,
प फ ब भ म,
य र ल व श ष स ह

Therefore, we can say that the Hindi language is phonemic in nature. Amalgamation of vowels with

consonants will from a syllable and it is also called as "Baraha Khadi" [27].

E. Bharati Braille

Bharati Braille is a largely unified Braille script for writing the languages of India which was developed in April 1951 and is the collective name given to all the languages of India [26].

The Bharati Braille, the Braille scheme adopted by India and some of the Asian Countries, use phonetic equivalents from the English Braille to represent the Hindi texts. Bharati Braille follows the syllabic writing system for all the Indian languages. The Braille notations for the Hindi Braille are shown in Fig. 3. Braille notations for special symbols of Bharati Braille are shown in Fig. 4 [26].

Alphabet: Hindi		
अ a (1)	ः	घ gh (126)
आ ā (345)	ः	ङ ṅ (346)
इ i (24)	ः	च c (14)
ई ī (35)	ः	छ ch (16)
उ u (136)	ः	ज j (245)
ऊ ū (1256)	ः	झ jh (356)
ए e (26)	ः	ञ ñ (25)
ऐ ē (15)	ः	ट ṭ (23456)
ऑ ai (34)	ः	ठ ṭh (2456)
औ o (1346)	ः	ड ḍ (1246)
ओ ō (135)	ः	ढ ḍh (123456)
औ au (246)	ः	ण ṇ (3456)
ऋ ṛ (5, 1235)	ः	त t (2345)
ॠ ṝ (6, 1235)	ः	थ th (1456)
ऌ ḷ (5, 123)	ः	द d (145)
ॡ ḹ (6, 123)	ः	ध dh (2346)
क k (13)	ः	न n (1345)
ख kh (46)	ः	प p (1234)
ग g (1245)	ः	फ ph (235)
ब b (12)	ः	भ bh (46)
म m (134)	ः	य y (13456)
र r (1235)	ः	ल l (123)
ळ l (456)	ः	व v (1236)
श s (146)	ः	ष ṣ (12346)
स s (234)	ः	ह h (125)
ह h (125)	ः	क्ष ks (12345)
ज़ z (1356)	ः	ज्ञ jñ (156)
		झ ङ (12456)
		ड ढ (5, 12456)
		फ भ (124)

Diacritics		
◌̣ Virama (4)	◌̣ Visarga (6)	s Avagraha (2)
◌̣ Anusvara (56)	◌̣ Candrabindu (3)	

Numbers	
१ 1 (3456, 1)	६ 6 (3456, 124)
२ 2 (3456, 12)	७ 7 (3456, 1245)
३ 3 (3456, 14)	८ 8 (3456, 125)
४ 4 (3456, 145)	९ 9 (3456, 24)
५ 5 (3456, 15)	० 0 (3456, 245)

Fig. 3. Hindi Braille Notations

Punctuation: Indian Languages		
, comma	(2)	⠠
; semicolon	(23)	⠠⠠
: colon	(25)	⠠⠠
danda, . period, full stop	(256)	⠠⠠
double danda	(256, 256)	⠠⠠⠠
? question mark	(236)	⠠⠠
! exclamation	(235)	⠠⠠
' apostrophe	(2)	⠠
"..." quote	(236...356)	⠠⠠⠠⠠
[...] brackets	(2356...2356)	⠠⠠⠠⠠
- hyphen	(36)	⠠⠠
— dash	(36, 36)	⠠⠠⠠
* asterisk	(35, 35)	⠠⠠⠠
capital *	(6)	⠠
number sign	(3456)	⠠⠠⠠⠠
italics	(46)	⠠

* Capital used only when transcribing English

Fig. 4. Special symbols of Bharati Braille

F. Review on Related Works and Need For Bharati Braille Bank

The Braille code system has been widely used by the visually impaired since it is comfortable for them to use it for reading and writing. Different countries have adopted their own Braille notations to suit their native languages. To develop Braille recognition system, it is required to have the stored images of Braille sheets. Previously, such a database was developed for English and Arabic Braille. A Braille database for both single sided and double sided Braille documents has been established for developing an OBR for English language by Nestor Falcon et.al. [20]. Zainab et.al. , have developed the database of single sided Braille document of Green and Yellow color [21]. However, the size and scope of these two databases are limited.

Many research developments are lagging behind due to want of sufficient, high-quality and accurately standardized resources for various applications. Developing the optical Braille recognition system needs high-quality Braille images; this indeed requires a Braille corpus. Various authors have attempted to develop the Braille corpora but they limited their work to generate the images of Braille sheets of a few pages and access to the developed database is not provided. Hence, developing a robust system for Braille recognition is hindered by lack of various resources. Also, the processes of the evaluation of new algorithms and comparison of the research results are complicated due to lack of standardized test data and testing methods. The set of problems and challenges is reminiscent of the research in OBR. This motivated us to develop the rich Braille corpora and provide free access to the researchers working in the area of Hindi OBR.

The primary reason for selecting the Hindi Devanagari Braille is, as English Braille has been globalized, the OBR software for the same has been commercialized, but in case of other languages it is less focused. Various authors have developed the OBR systems for English and Arabic [1] - [19]. Very less effort has been put in for the development of OBR for Hindi Braille. Hindi being the national language of India, the commercial software is unavailable for Hindi Braille. A significant effort is required to develop software for Hindi Devanagari Braille.

Optical Braille recognition is not a straight forward task for Indian languages as compared to English, as the basic character set of all the Indian languages is more than 64 and hence to represent a character, two Braille cells are required. OBR for Bharati Braille is still an area that requires the contribution of many enhanced research works. For all this, it is essential to develop the Braille corpora to enable the researchers to study the properties of different Braille documents and to develop models for OBR. So far, no efforts have been made for development of corpora for Hindi Devanagari Braille. Thus, we have provided Braille corpora built out of variety of documents that help in developing new algorithms for Hindi OBR, de-skewing algorithms and adaptive algorithms for Braille documents with varied resolutions. Large database is required to achieve a robust recognition.

Hence we have the standard Hindi Braille corpus with 155 pages of Hindi Devanagari Braille document, with approximately 2603 lines, 36570 cells and 102230 dots. These documents are manually validated and edited to ensure the correctness.

This paper describes the development of Braille corpora/ database for Hindi Devanagari language. The various application specific databases have also been introduced. Section II describes the outline of the Braille corpus design. Section III depicts the structure of the developed database. Section IV gives the statistical analysis of the developed database. Section V presents the few works being carried out at the various institutions, Universities and Research laboratories. Section IV presents the conclusion and discussion regarding the future work.

II. BRAILLECORPUS DESIGN

Braille-Bank is an archive of well characterized Bharati Braille documents pertaining to Hindi Devanagari language for use by the research community. The collection of the data is done in such a way that it includes all the possible components required for the development of robust OBR system for Hindi language.

A. Braille Corpus Collection

Traditional sources of Braille data are the machine printed (embossed), and the hand punched Braille documents. A long way has been crossed to come up with a Braille corpus to give a platform for researchers to develop the optical Braille recognition system and to check for accuracy and better performance. To start with an extensive survey of the related Braille documents of the selected language - Hindi (which is the area for which OBR is less focused) was done. The primary data collection was done in the form of 6th standard Hindi Braille text Book of Government of Karnataka (GOK). Criterion for selecting this database was that it consists of Braille cells corresponding to simple characters to complex words, numerals and special symbols. Single sided and double sided embossed Braille documents were generated with the help of Government Braille printing press situated in Mysore, Karnataka. Braille documents that were taken from Braille press were made to undergo repeated verification. Hand punched documents were prepared with the help of three blind students for the same set of data as that of embossed documents and each document was verified with a trained Braille reader.

B. Braille Document Acquisition

Firstly, Braille documents were acquired through the commercially available scanner **HP scan jet 3400C** and Braille images were stored in JPEG format. This scanner was selected depending on the illumination requirement where it can provide clear dot location. Initially the acquisition of double sided Hindi Devanagari Braille in grade-I with the resolution of 1700x1502 was carried out for huge number of Braille document consisting of 95 pages. Later all the files were cropped to 1501x2121

using MATLAB programming to remove the unwanted region in the image. Like double sided Braille documents, all other documents were also cropped to the dimensions such that it excludes the unwanted regions. Each scanned Braille image consists of three types of pixels with different intensity values for background, a light area and a dark area for each recto and verso dot.

C. Braille Document Annotation and Equivalent Text Generation

The visual contents of the Braille documents i.e. the dots can be annotated in various possible ways to favour the researchers. The annotation of Braille document in our case was done using three experienced blind Braille readers. The information regarding the number of dots, number of lines, number of characters and number of words for each document is provided. To understand the contents in the Braille corpora and to ease the analysis of the work, it is required to provide the corresponding text of the Braille language. The Braille documents were read with the help of 3 blind students who are trained to read Hindi Braille language and equivalent text was generated using the "*Baraha software*".

III. STRUCTURE OF THE DATABASE

Initially the database has been validated in order to check the corpus before acquisition and to make sure that it would correspond to the specifications and include the versatility of the Braille text data. A huge database has been created in order to develop a robust OBR system for Hindi Braille with as many variants as we could. To gather the versatility of Hindi OBR six categories were identified, namely:

- Double sided embossed Braille document
- Single sided embossed Braille document
- Hand punched Double sided Braille document
- Hand punched single sided Braille document
- Skewed Braille documents
- Braille documents acquired with different resolution

All the above six categories of Hindi Braille database are confined to 6th standard Hindi Braille text book of GOK. A double sided database has been developed for the entire text book whereas rest of the databases are developed on a selective basis of the Braille sheets. The database consists of healthy documents that sans inter-dot cracks that exist during the embossing process and unhealthy documents with inter-dot cracks due to the surface tension of the paper. The unhealthy documents require efficient image pre-processing techniques to eliminate the inter-dot noise components. The embossed single sided documents are free of inter-dot noise components. This is also true for the case of hand punched documents as they have been developed using the inter-dot Braille slate. We could not get much of the hand punched documents, because even if a single dot is punched wrongly, the entire documents are to be rejected and re-punched from scratch, which is a laborious and

time consuming process. The skewed database as well as the database with varying resolution have been created in view of development of robust OBR systems. For documents to be free from tilt, it has to be aligned exactly with the edge of the scanner which is not always the case and this results in skewness in the document thus posing a major limitation to the system. The researcher has to take care of this skew angle correction while developing the robust OBR model. In this direction the skewed database has been created with tilt in both clockwise and in anti-clockwise direction. Documents with varied pixel resolution have been developed. Recognizing the dot components of these databases requires an efficient thresholding technique.

The developed database can be made use of by those researchers who are working on various areas as specified in Fig. 5. For double sided Braille the database consists of images of both sides of the Braille sheet. A sample database is as shown in Fig. 6, and its equivalent text is provided in Fig. 7, so that the researcher can verify the trueness of the system that he has developed. To clearly define the accuracy of the dot recognition this database provides the complete details of each Braille document. As an example, Table 1, depicts the detailed information of page number 9 and page number 10 of the double sided Braille document. All the documents are of size 28.5cm (horizontal) x21cm (vertical) and every document has been scanned with 200dpi and bit-depth of 24 bit.

The information on each and every document regarding the dot counts, cell counts, line counts, character counts, and information regarding the document type, size of the document, resolution of the acquired image database, pixel resolution and image format are summarized and depicted in Table 2 to Table 7, for each category of the developed databases.

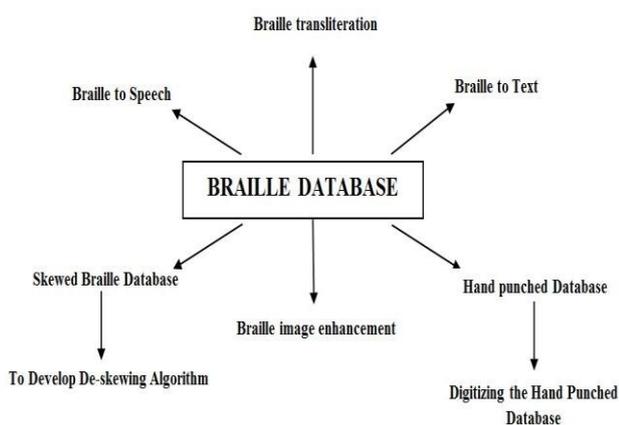
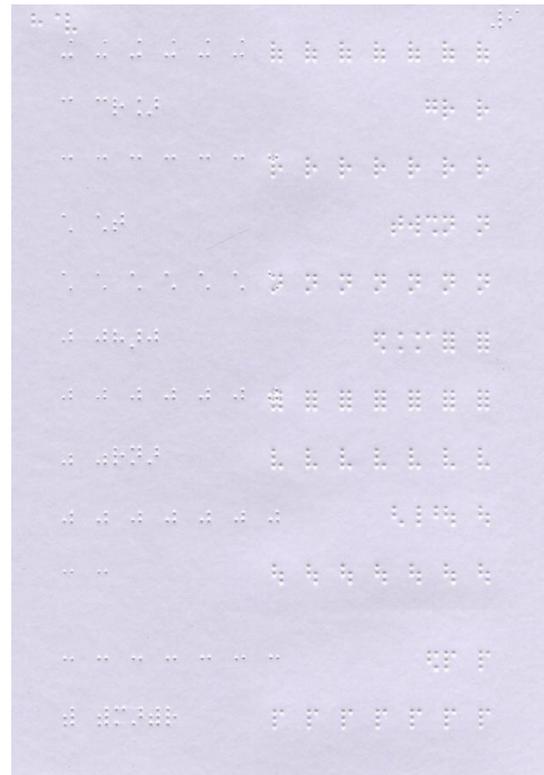
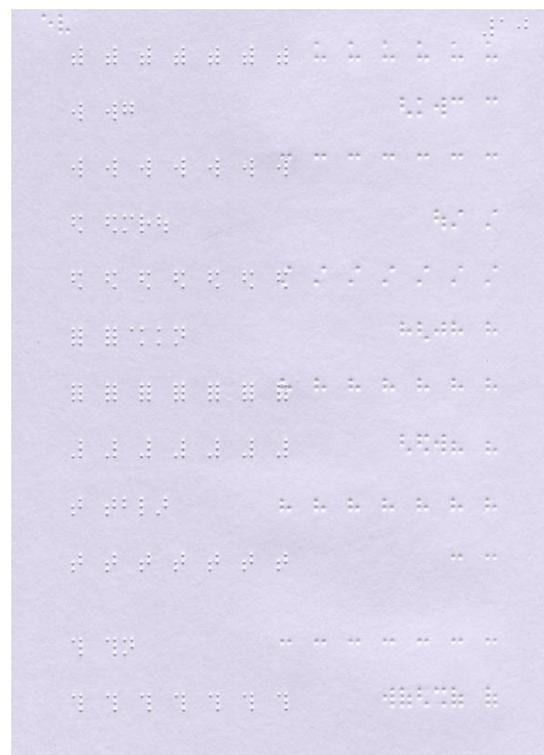


Fig. 5. The datasets provided is intended to enable a wide range of applications



(a)



(b)

Fig. 6. Acquired inter-point Braille image. (a) Front side and (b) Back side document

9	10
क क क क क क	ट ट ट ट ट ट ट
च चरखा	ठ ठग
च च च च च च	ठ ठ ठ ठ ठ ठ ठ
च	ड डमरू
छ छत	ड ड ड ड ड ड ड
छ छ छ छ छ छ छ	ढ ढककन
ज जहाज	ढ ढ ढ ढ ढ ढ ढ
ज ज ज ज ज ज ज	ण ण ण ण ण ण ण
झ झरना	त तबला
झ झ झ झ झ झ झ	त त त त त त त
ज ज	थ थन
ज ज ज ज ज ज ज	थ थ थ थ थ थ थ
ट टमाटर	

(a) Text database of front side (b) Text database of Back side
Fig. 7. Equivalent text database for Braille document

The tabulated information is provided in a similar fashion for all the documents for facilitating those users who are unaware of Hindi language and still working on developing the OBR system for their native languages. The users who are intended to work on developing the OBR system for Hindi Devanagari Braille can make use of the developed database to the full extent and for those users who are working on development of OBR systems for other languages can use this tabulated information to assess the performance of the system that they are intending to develop as the Braille has a standard pattern and dimensions, only the mapping differs from language to language. As the phrase says, "One picture is worth more than ten thousand words", this database can be used worldwide for developing the OBR systems as there is a lot of room for fruitful research in the area of OBR.

Table 1. Inter-point Braille document information (a) Front side Braille sheet (b) Back side Braille sheet

Page No.	Line Number	No. of dots	No. of cells	Page No.	Line Number	No. of dots	No. of cells
9	1	6	2	10	1	8	3
	2	18	6		2	35	7
	3	12	5		3	12	3
	4	14	7		4	28	7
	5	8	4		5	19	5
	6	14	7		6	28	7
	7	16	5		7	21	6
	8	28	7		8	42	7
	9	17	5		9	28	7
	10	21	7		10	16	5
	11	8	2		11	28	7
	12	14	7		12	12	3
	13	25	6		13	28	7
Total	13	201	70	Total	13	305	74

Reference databases such as *Bharati Braille Bank* are essential resources for developers and evaluators to test

the algorithms with realistic data and to perform these tests repeatedly and reproducibly as and when the algorithm refinements are proposed. By making these data accessible to the researchers, these databases will make it possible to formulate and answer the problem without the need to develop a new set of reference data at greater costs in every single case. In this regard, Bharati Braille Bank can serve as a standard and permanent repository for studies related to Braille and, in particular, for Hindi Devanagari Braille.

Table 2. Embossed Double Sided Braille Database Developed

Number of Braille document	: 95pages
Total number of characters	: 25827
Average number of characters per sheet	: 271
Total number of lines	: 18
Total number of dots	: 71706
Average number of dots per sheet	: 754
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 1501x2121
Image format	: .jpg
Image Size	: 328KB

Table 3. Embossed Single Sided Braille Database Developed

Number of Braille document	: 20pages
Total number of characters	: 3508
Average number of characters per sheet	: 175
Total number of lines	: 292
Average number of lines per sheet	: 15
Total number of dots	: 9866
Average number of dots per sheet	: 493
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 1501x2121
Image format	: .jpg
Image Size	: 296KB

Table 4. Hand punched Double-Sided Braille Database Developed

Number of Braille document	: 6pages
Total number of characters	: 727
Average number of characters per sheet	: 121
Total number of lines	: 82
Average number of lines per sheet	: 14
Total number of dots	: 2047
Average number of dots per sheet	: 341
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 1700x2341
Image format	: .jpg
Image Size	: 348KB

Table 5. Hand punched Single-Sided Braille Database Developed

Number of Braille document	: 4pages
Total number of characters	: 1509
Average number of characters per sheet	: 377
Total number of lines	: 64
Average number of lines per sheet	: 16
Total number of dots	: 3009
Average number of dots per sheet	: 752
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 1700x2341
Image format	: .jpg
Image Size	: 308KB

Table 6. Skewed Braille Database Developed

Number of Braille document	: 10pages
Total number of characters	: 1401
Average number of characters per sheet	: 140
Total number of lines	: 138
Average number of lines per sheet	: 14
Total number of dots	: 3985
Average number of dots per sheet	: 398
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 1501x2121
Image format	: .jpg
Image Size	: 380KB

Table 7. Braille Database Developed For Images with Different Resolution

Number of Braille document	: 10+10pages
Total number of characters	: 3598
Average number of characters per sheet	: 180
Total number of lines	: 292
Average number of lines per sheet	: 15
Total number of dots	: 10076
Average number of dots per sheet	: 504
Document size	: (28.5cmx21cm)
Digital format	: 24bit color image
Resolution	: 200dpi
Pixel resolution	: 503x685 and 985x1380
Image format	: .jpg
Image Size	: 44KB and 128KB

IV. STATISTICAL ANALYSIS AND OBSERVATION

The developed database of Hindi Braille includes all the commonly used vowels, consonants, vowel consonant combinations, special letters, punctuation marks, symbols, numerals and foreign words. In this database foreign language words of English were obtained. While designing the Braille to text conversion special rules are to be applied for recognizing the numerals, special symbols and foreign words etc. Table 8, provides the number of numerals, number of characters, number of words, number of foreign words and number of special symbols for a set of 155 documents considered.

Table 8. Bharati Braille Bank statistical details

Parameter	Quantity
Number of numerals	1120
Number of Characters	36570
Number of words	4872
Number of foreign words	29
Special symbols	2307

Potential benefits

- Stimulate firsthand investigations in the area of OBR for Hindi and other Indian languages.
- Provide greater and more comprehensive databases than any single center can accumulate.

- The combination of these six sub databases are envisioned to make the resource useful for different purposes to an extensive range of researchers.
- Without common databases such as those provided by the Braille bank, it is impossible to resolve certain contradictory research results.

A. Application Specific Braille Database

The Hand Punched Braille database has been developed for those who are intending to work on the recognition of only the Hand punched documents so that the documents of blind authors which were written in Braille format can be restored for future generation. The skewed Braille database will be useful for those who wish to work only on development of the de-skewing algorithm either for Braille documents or for any documents in general. The database of documents with varied resolution will be of help for those who are intending to develop an adaptive Braille recognition algorithm. In general these application specific databases will be helpful aids in the development of the robust OBR system.

B. Corpus Analysis

This database can act as a central for a wide variety of fields ranging from Braille image enhancement to Braille character to speech conversion. It also plays a valuable role in evaluating the performance of the algorithms for OBR. The overall analysis of the proposed database is as shown in Table 9. So we have a rich database with six categories for 155 pages of Braille document consisting of 2603 lines and 102230 dots and 36570 characters.

Table 9. Braille Corpus details

Database Type	Pages	Lines	Dots	Characters
Double sided Braille document	95	1735	71706	25827
Single sided Braille document	20	292	9866	3508
Skewed Braille document	10	138	5544	1401
Different resolution Braille document	20	292	10076	3598
Hand punched Braille document	10	146	5038	2236
Total	155	2603	102230	36570

C. Comparison With The Existing Databases

Various authors have developed databases for the Braille recognition systems. Only few have reported about the database they have used for developing their OBR system and others have not made a mention of the database they have used for their OBR systems. Here we try to compare the few we have come across. The available database information reveals that a limited set

of Braille images are acquired and the adequate information of the developed database is not provided. Also, the developed databases are not versatile. The access to those images is also not provided. Table 10, shows the various bases on which the comparative analysis of the available databases developed by different authors and the database proposed by us has been carried out.

Table 10. Comparison of the databases

Database	[20]	[21]	Bharati Braille Bank
Language	English	Arabic	Hindi Devanagari
Braille sheets	26	25 green and 25 yellow sheets	155
Total number of characters	30862	-	36570
Digital format	Gray scale	Color	Color
Resolution	100dpi	100dpi	200dpi
Image size	1360 Kbytes	274 Kbytes for yellow sheets & 336kbytes for green sheets	328 Kbytes
Image format	BMP	JPG	JPG
Braille type	Double-sided, grade-1	Single-sided, grade-1	Single and Double sided embossed, hand punched documents, skewed documents, documents with varied resolution
Document size	29.5cm (horizontal) x 30.5cm (vertical)	29.5cm (horizontal) x 30.5cm (vertical)	28.5cm (horizontal) x 21cm (vertical)

V. VARIOUS WORKS / PROJECT GOING ON FOR DEVELOPMENT OF OPTICAL RECOGNITION OF BHARATI BRAILLE

There are various research projects/ works going on in India related to Optical Braille Recognition of Bharati Braille. At SJ College of engineering, Mysore, a project for recognition of Kannada Hand punched Braille Recognition system is going on [24]. At Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, development of OBR system for Hindi and Tamil languages are being carried out [23]. At the Center for Microprocessor Application for Training Education and Research", "Project on Storage Retrieval and Understanding of Video for Multimedia" of Computer Science & Engineering Department, Jadavpur University is being carried out for development of OBR for Bengali language [22].

VI. DISCUSSION AND CONCLUSION

The overall goal of our work is to develop a standard database that is technically useful for building the OBR system and for evaluating the performance of the developed system. We believe that the developed database satisfies the requirements of the users as it provides a wide selection of databases for developing the useful OBR system. In this paper we have discussed the various applications of the developed Hindi Devanagari Braille database- *Bharati Braille Bank*. It consists of data pertaining to the Double sided embossed Braille document, Single sided embossed Braille document, Single sided and double sided Hand punched document, skewed documents and documents with varying resolutions that in all includes 155 pages, 2603 lines, 102230 dots and 36570 characters. This effort on database development for Hindi Devanagari Braille has been done for the first time, paving a way to develop new OBR system for Hindi language that is useful for real life application for the visually impaired. Collaborative efforts among geographically scattered researchers working to develop Hindi OBR can make use of the proposed database. This work aims at motivating researchers to develop the Braille database for all the native languages which, in turn, will help to develop robust OBR systems useful for the real life scenario. OBR for Bharathi Braille is still an area that requires the contribution from many research works.

APPENDIX A GETTING STARTED WITH BRAILLE BANK

An outsized collection of these databases is now accessible to the scientific community via CD-ROM. We provide the assistance to the researchers in making the best use of the resource. To begin using the *Braille Bank* kindly mail us on speak2shree@gmail.com and v_udayashankara@yahoo.co.in. There is no charge for using *Bharati Braille Bank* but we need the valid authentication of the researcher before sharing the data. We invite your comments and contributions.

ACKNOWLEDGEMENT

We would like to thank Council of Scientific and Industrial Research (CSIR), New Delhi, India for providing the financial assistance for this work under the research scheme No: 22(0613)/12/EMR-II and also the work is supported by JSS Research Foundation, Mysore, Karnataka. Thanks are extended to our PG scholar Sowmya.M.S, Vasumathi, Braille reader for their valuable efforts in development of this database. Our special thanks to Government Braille Printing Press, Mysore, for sharing the data with us. We acknowledge Nattarasu.V, Associate professor, Department of Electronics and communication, SJCE, Mysore for his valuable remarks and inputs during the preparation of database.

REFERENCES

- [1] J. Mennens, L. Van Tichelen, G. Francosis and J. Engelen. "Optical recognition of Braille writing", *Proceedings of*

- second international conference on Document analysis and Recognition*, IEEE Oct-1993, pp. 428-431.
- [2] J. Mennens, L. Van Tichelen, G. Francosis and J. Engelen. "Optical Recognition of Braille Writing Using Standard Equipment", *IEEE transactions of rehabilitation engineering*, Vol. 2, No. 4, Dec 1994.
- [3] T. W. Hentzschel, and P. Blenkhorn. "An Optical Reading Systems for Embossed Braille Characters using a Twin Shadows Approach", *Journal of Microcomputer Applications*, 1995, pp. 341-345.
- [4] R. T. Ritchings, A. Antonacopoulos and D. Drakopoulos. "Analysis of Scanned Braille Documents", In: Dengel, A., Spitz, A.L. (eds.): Document Analysis Systems, *World Scientific Publishing Company* 1995, pp. 413-421.
- [5] Y. Oyama, T. Tajima, and H. Koga. "Character Recognition of Mixed Convex- Concave Braille Points and Legibility of Deteriorated Braille Points", *System and Computer in Japan*, Vol. 28, No. 2, 1997.
- [6] C.M. Ng, V.Ng and Y.Lau. "Regular feature extraction for recognition of Braille", *Third International conference on computational Intelligence and Multimedia Applications*, 1999, pp. 302-306.
- [7] A. Antonacopoulos and D. Bridson. "A Robust Braille Recognition System", Document Analysis Systems VI, A. Dengel and S. Marinai (Eds.), *Springer Lecture Notes in Computer Science*, LNCS 3163, 2004, pp. 533-545.
- [8] L. Wong, W. Abdulla and S. Hussmann. "A Software Algorithm Prototype for Optical Recognition of Embossed Braille", *17th Conference of the International Conference in Pattern Recognition, Cambridge, UK, IEEE-2004*, pp. 23-26.
- [9] N. Falcon, C. M. Travieso, J. B. Alonso and M. A. Ferrer. "Image Processing Techniques for Braille writing Recognitor", *EUROCAST 2005*, LNCS 3643.
- [10] A. Malik Al-Salman, Y. ALOHAI, M. Alkanhal and A. Airajith. "An Arabic Optical Braille Recognition System", *ICTA Apr 2007*, pp.12-14.
- [11] A. Malik S. Al-Salman, A. El-Zaart, Y. Al-Suhaibani, K. Al-Hokail and A. O. Al-Qabbany. "An Efficient Braille Cells Recognition", IEEE-2010.
- [12] J. Yin, L. Wang and J. Li. "The Research on Paper-mediated Braille Automatic Recognition Method", *Fifth International Conference on Frontier of Computer Science and Technology, IEEE-2010*, pp. 619-624.
- [13] J. Li and X. Yan. "Optical Braille Character Recognition with Support-Vector Machine Classifier", *International Conference on Computer Application and System Modelling*, 2010.
- [14] S. D. Al-Shamma and S. Fathi. "Arabic Braille Recognition and Transcription into Text and Voice", *5th Cairo International Biomedical Engineering Conference Cairo, Egypt*, Dec 2010.
- [15] Z. Tai, S. Cheng, P. K. Verma and Y. Zhai. "Braille document recognition using Belief Propagation", *Journal of Visual Communication and Image Representation* 21(7): 722-730 (2010).
- [16] A. Al-Saleh, A. El-Zaart and A. Malik Al-Salman. "Dot Detection of Braille Images Using A Mixture of Beta Distributions", *2011 Journal of Computer Science* ISSN 1549-3636, pp. 1749-1759.
- [17] J. Bhattacharya, S.Majumder and G.Sanyal. "Automatic Inspection of Braille character: A Vision based approach", *International Journal of computer and Organization trends* – volumel, Issue3, 2011, ISSN: 2249-2593, pp. 19-26 .
- [18] M. Wajid, M. Waris Abdullah and O. Farooq. "Imprinted Braille-Character Pattern Recognition using Image Processing Techniques", *International Conference on Image Information Processing*, 2011.
- [19] R. Ismail Zaghoul and T. Jameel Bani-Ata. "Braille Recognition System – With a Case Study Arabic Braille Documents", *European Journal of Scientific Research*, ISSN 1450-216X, Vol.62, No.1, 2011, pp. 116-122.
- [20] N. Falcon, C. M. Travieso, J. B. Alonso and M. A Ferrer, "Image Processing Techniques for Braille writing Recognitor", *EUROCAST 2005*, LNCS 3643.
- [21] Z. Authman and Z. F. Jebr, "Algorithm Prototype for Regular Feature Extraction of Arabic Braille Scripts" *International Journal of Scientific Knowledge Computing and Information Technology* ISSN 2305-1493, February 2013. Vol. 2, No.1, pp. 20-27 .
- [22] S. Halder, A. Hasnat, A. Khatun, D. Bhattacharjee and M. Nasipuri, "Development of a Bangla Character Recognition (BCR) System for Generation of Bengali Text from Braille Notation", *International Journal of Innovative Technology and Exploring Engineering* , ISSN: 2278-3075, Volume-3, Issue-1, June 2013.
- [23] S. Padmavathi, K.S.S Manojana, S. Sphoorthy Reddy and D. Meenakshy, "Conversion of Braille to Text in English, Hindi and Tamil Languages", *International Journal of Computer science, Engineering and Applications*, Vol.3, No.3, June 2013.
- [24] S. Srinath and C. N. Ravi Kumar, "Recognition of Hand Punched Kannada Braille Characters Using Knowledge Based Multi Decision Concept: Basic Symbols", *Proceedings of International Conference on Advances in Computing Advances in Intelligent Systems and Computing*, Vol. 174, 2012, pp 847-857.
- [25] "World Braille Usage". Third Edition. Perkins. International Council on English Braille. National Library Service for the Blind and Physically Handicapped.
- [26] World Braille Usage. UNESCO. National Library Service for the Blind and Physically Handicapped. Library of Congress. Washington, D.C., USA. 1990.
- [27] C. Arun Kumar, T. Shreekanth and V. Udayashankara, "Development of Speech Database for Hindi Text to Speech system Considering Syllable as a Basic Unit", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 4, 2014, pp. 531-549.

Author's Profiles



Shreekanth. T completed his Bachelor of Engineering in Electronics and Communication from PES College of Engineering, Mandya, affiliated to Visvesvaraya Technological University, Belgaum, Karnataka, and obtained his master's degree in the area of Bio-Medical Signal Processing from Sri Jayachamarajendra College of Engineering, Mysore affiliated to Visvesvaraya Technological University, Belgaum, Karnataka. Currently he is pursuing PhD in the area of Image and Speech signal processing. His areas of interest include Digital Image Processing and Speech Signal Processing. He has to his credit 15 conference and Journal papers both at national and international level. Presently he is working as an Assistant Professor in the department of Electronics and Communication at Sri Jayachamarajendra College of Engineering, Mysore, Karnataka, India.



Udayashankara. V completed his Bachelor of Engineering in Electronics and Communication from Sri Jayachamarajendra College of Engineering, Mysore Karnataka and, obtained his M.E and PhD degree from Indian Institute of Science (IISc) Bangalore. Currently he is a Professor

and Head with the department of Instrumentation Technology, at Sri Jayachamarajendra College of Engineering, Mysore, India. His research interests include Rehabilitation Engineering,

Digital Signal Processing, Speech recognition, Speech enhancement and EEG analysis. He has authored more than 100 publications in National and International Journals and Conferences in these areas. He has Authored three books, 8051 Microcontroller: Hardware, software and applications, McGraw Hill-2009, Real Time Digital Signal Processing, PHI-2010, Modern Digital Signal Processing, PHI-2012.

How to cite this paper: Shreekanth.T, V.Udayashankara,"A New Research Resource for Optical Recognition of Embossed and Hand-Punched Hindi Devanagari Braille Characters: Bharati Braille Bank", IJIGSP, vol.7, no.6, pp.19-28, 2015.DOI: 10.5815/ijigsp.2015.06.03