Image To Speech Synthesizer with Reference to Assamese Numerals

M. Rahman, P. Sarma, M.P. Bhuyan, A. Das, D. Dutta

Abstract*: In this research work we have shown the methodology for converting printed Assamese numerals to its corresponding utterance. We have implemented as an initial effort which will read only four digit numerals. We are using Image processing techniques to convert an image of Assamese numerals into textual/digital form. In the second phase the numerals will be pronounced as a number by Google speaker. In this system, images are stored in a dataset and then inputted data is compared with the dataset image using template matching technique. After recognition of the text output will be displayed as a speech waveform. This work has many applications in today's digital world.

Keywords:* OCR, *Image processing, Template matching, Synthesizer, Segmentation, Recognition.

I. INTRODUCTION

Nowadays people need not manually type important documents when entering them into electronic databases. It is OCR which helps us to input printed text into computer system. OCR extracts useful information and enters it automatically. The result is accurate and efficient and it requires less time.

Assamese is an Indo-Aryan language widely spoken in the North-Eastern state Assam of India. According to the 2011 census, there are more than 15 million native Assamese speakers found. Assamese is one of the 22 scheduled languages of the Indian Constitution [1]. The speakers of Assamese are also found in other North-Eastern states viz. Arunachal Pradesh and Nagaland. The Assamese language script is found in Nepal in Pashupati temple and also in Myanmar, this shows the expansion and influence of the Assamese language [1] Nefamese is an Assamese based language which is also known as Arunamese is spoken in Arunachal [2], similarly, Negamese is also an Assamese based language and the speakers of the Negamese are found in Nagaland [3]. These languages are the lingua franca of these regions. Assamese was evolved before the 7th century. In the 17th century the Assamese language was used as a court official language in the Ahom Kingdom. Assamese was evolved from the old Indo-Aryan dialects but the exact origin of the language is not clear yet. Madhav kandali, an Assamese poet composed the Saptakanda Ramayana in Assamese script and this is the first regional language translation of Ramayana after Valmiki's Ramayana which was written in Sanskrit. In 15th and 16th centuries Borgeet songs and Ankiya naat dramas were written by Bhattadeva which were the first prose in Assamese. Supangmung , a Ahom king (1663-1670) started use of Assamese coins in his kingdom [4]. The Assamese script has 41 consonants and 11 vowels [5, 6]. The Assamese language is also having many juktakhars (conjugate words), they are formed by combining more than one alphabet. The Assamese script is written from left to right. The Assamese language has eight grammatical cases viz. absolutive, ergative, accusative, genitive, dative, terminative, instrumental, and locative.

From the above discussion, notably, the Assamese is a widely used language with great history since the dawn of civilization. At the same time in the present scenario, most of the languages in the world are digitized and computerized. But, the Assamese language has just started its journey in the digital world. So, the development of the language is also important in the digital world. In this research work, Assamese numerals are given importance.

Table-I Assamese Numerals [7]

Hindu-Arabic Numerals	Assamese Numerals	Pronunciation in Assamese	Assamese digits in Assamese script (written form)
0	০	*xuinno*	শূণ্য
1	১	*ek*	এক
2	২	*dui*	দুই
3	৩	*tini*	তিনি
4	৪	*chari*	চাৰি
5	৫	*pas*	পাঁচপাঁচ
6	৬	*choi*	ছয়
7	৭	*xat*	সাত
8	৮	*aath*	আঠ
9	৯	*no*	ন
10	১০	*doh*	দহ্
11	১১	egharo	এঘাৰ
.....	.....	.....	.....
.....	.....	.....	.....
99	৯৯	*nirannobboi*	নিৰন্নবৈ্ব

There are ten Assamese numeral symbols like the Hindu-Arabic numerals; they are shown in Table-I from 0 to 9. There are another 90 Assamese numerals in table 1 with their pronunciations, which are used for the implementation of this work. In this research work,

Revised Manuscript Received on November 06, 2019.

M. Rahman, Department of Information Technology, Gauhati University Guwahati, India, Email: mirzanurrahman@gmail.com.
P. Sarma, Department of Information Technology, Gauhati University Guwahati, India, Email: parismita.sarma@gmail.com.
M.P. Bhuyan, Department of Information Technology, Gauhati University Guwahati, India, Email: mpratim250@gmail.com.
A. Das, Department of Computer Science NERIM Group of Institutions, Guwahati, India.
D. Dutta, Department of Computer Science NERIM Group of Institutions, Guwahati, India.

a model is designed which can pronounce Assamese numbers in the Assamese phoneme. For example ১০ (doh), ১১ (egharo), ৫০ (ponsas) etc. In Hindu-Arabic these are ten, eleven, one hundred, and one hundred fifty respectively.

II. RELATED WORKS

We have studied deeply a number of research papers for our research work, a few important stuff are discussed here. In [8], authors have used machine learning technique viz. Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbour (KNN) to recognize the vowels of Assamese spoken words automatically. They have used a database of 20 words spoken by 55 different speakers and each word was repeated two times by each speaker; finally, 2200 samples were obtained. They have found that the Random Forest was able to show better result than the other two. The accuracy obtained in their experiment was 86.5%.

In [9], it has been noticed that the authors used the concept of Optical Character Recognition (OCR) and Text to Speech Synthesizer to extract the text from images and then to get the speech signal of that text. The authors used two major modules. In the first stage they convert the image which was captured by Webcam into text using OCR. In the second stage they apply natural language processing and digital signal processing to get the speech of the text using Text to Speech synthesizer (TTS). It has been seen that the authors introduced an efficient and real time technique.

In [10], the authors have used various methodologies to build OCR for Bangla language. Skew correction, text graphics separation, line segmentation, zone detection, word and character segmentation using some conventional and some newly developed techniques they have used. The character unigram statistics is used to make the tree classifier efficient. Several heuristics are also used to speed up the template matching approach. A dictionary-based error-correction scheme has been used where separate dictionaries are compiled for root word and suffixes that contain morpho-syntactic information as well. The accuracy obtained by them was 95.50%.

In [11], the authors have used Tesseract to build the OCR for Bangla language. The main methodologies include training data preparation process, Tesseract integration procedure and the post-processing techniques. They have obtained accuracy ranging from 70% to 93% for different quality of images. The authors have mentioned that their technique can be used in other similar type of languages or sister languages.

In [12], the authors have used a template matching technique to find the location of an image inside a large image. The authors have mentioned that this type of template matching technique can be used to convert the image text to normal text. They have used single character and multiple character comparison using three template matching techniques like correlation method, cross correlation method and performance index method. The performance index was able to outperform than the other two.

In [13], the author has used Template Matching as the algorithm that is applied to recognize the characters, characters tested are alphabet (A – Z), grey scale images were used with times new roman font type and recognizing the alphabet by comparing between two images. The main disadvantage of their model was the model could work only the stored images templates. The model could not interpret the image which was not stored in the template dataset.

In [14] a model is designed for offline handwritten character segmentation for Bangla language. The authors have mentioned the uniqueness of Bangla script due to having the matras and also at the same time these matras bring a challenge for the people at the time of segmentation because of its wavy and discontinuous nature. The authors have applied their technique on 500 handwritten Bangla words and they have mentioned that their method was able to perform better than the earlier methods.

In [15], the authors have designed a statistical model for automatic error detection and correction of Assamese text. They have highlighted the importance of the local languages and enrich the local languages in computation field. The authors have used the character level scored based n-gram model and produce suggestion for the faulty letter or character. They have obtained accuracy ranging from 81% to 85%.

In [16], the authors have designed a model to enhance the automatic word prediction in Assamese language for the ambiguous words. They were able to improve the accuracy level by 6% using the higher order n-gram model. They have focused on the writing speed of a user along with the differently-abled people.

In [17], the authors have designed an information retrieval system for Assamese language using Assamese Wikipedia and Assamese WordNet. They have used three different techniques for their information retrieval process. The performance of the system is 60.08%.

From the above discussion it is seen that OCR is an important field and converting the image text to normal. In addition the conversion of normal text to speech is a new task in Assamese language. Also, from the above discussion it is visible that the various computational works in Assamese language is going on. So, the present work of getting the speech of Assamese numerals from image is a suitable work to enrich the field of Assamese Language Processing. In addition this research work will help the visually impaired person to recognize the Assamese numerals.

III. PROPOSED SYSTEM

Our proposed system deals with the problems of machine reading of Assamese printed numerals which is analogous to ability of human beings to recognize such numerals. The main aim of this research work is to:

- Extract the numerals from images.
- Convert the numerals into respective utterance.

The proposed system is built with the following tasks:

- Collection of images of Assamese numerals.

- Conversion of colored image to greyscale.
- Binarization and Noise removal
- Segmentation
- Template matching
- Recognition of the image
- Conversion of the numbers to speech

The output of this work is the numerals extracted from the image and the speech both in Assamese.

The input image is of following specification:

- Scanned or Captured Assamese numerals image in JPG format.
- Grey level, black and white or colour images as input.
- Input image contains Assamese Numerals only.
- Text image of four digits is acceptable.

A. Dataset Preparation

In Assam, Secondary Board of Assam (SEBA) makes it mandatory that all the schools under SEBA must have Assamese Language as a subject in their curriculum. All the Assamese numerals have been collected from Assamese text book of school. Images are also taken from posters, photos etc. All the images are first captured with a digital camera and then saved in a .jpeg format.

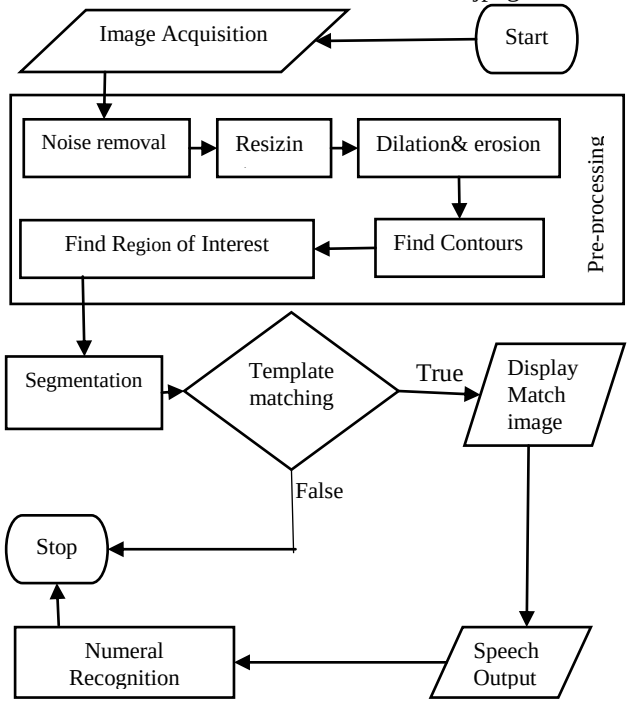

Fig. 1: Flowchart of the numeral recognition

We have taken 13 images of each of the 10 numbers of Assamese numerals and among these 7 numbers of images are selected randomly from the dataset for training purpose and remaining 6 numbers of images from the dataset are considered for testing purpose. Total 130 numbers of Assamese numerals images have been collected for the proposed system. In training phase, we have considered the images as input and then perform Binarization, dilation, erosion and then save it in a .jpeg format. In testing phase, we have taken colored image as well as black and white image as input and after pre-processing, image is matched with the template in the dataset. Fig. 1 shows the flowchart for the whole proposed system.

B. Image Input

The Image to be processed by the system should be in digital image format. We need to scan or captured the image containing the Assamese numerals into digital form.

Fig. 2: Input image

C. Pre-processing

Images are captured in colored form then it is converted into GRAY image. The purpose of this step is to make input image as noise free and ready for process. It consists the following:

- Convert the GRAY image into Binary image format.
- Apply basic operation on mathematical morphology (Dilation/Erosion) to the binary image.
- Find and sort the Contours.
- Region of interest (ROI) is found.
- Retaining the size of ROI.

1) Noise Removal

In digital images, various types of noise occur. Noise means unwanted signal. It reduces the accuracy of successive tasks of recognition. Noise may include uneven contrast, show through effects, and background spots etc. Noise is the result of errors in the image acquisition process that result in pixel values.

Fig. 3: Input image after noise removal

2) Dilation

Dilation is one of the morphological operations in digital image processing. This method adds pixels to the boundaries of objects in an image. It makes images more visible. It fills small holes in objects. Following is the syntax of this method: dilate (src, dst, kernel).

Fig. 4: Input image after dilation and noise removal

3) Erosion

This operation generally uses a structuring element for probing and compressing the shapes contained in the scanned input image. The output pixel in erosion gives the minimum value

of all pixels in the neighborhood. The syntax of this method is: erode (src, dst, kernel)

Fig. 5: Input image after noise removal, dilation and erosion

4) Contours

This is useful for shape analysis and object detection. It is explained as a curve joining all the continuous points along the boundary. Once a contour is extracted its different characteristics will be examined. For better accuracy, we use binary images. In OpenCV, contour is used to find white object from black background. So object to be found should be white and background should be black. Following is the syntax of this method: findcontours (src, mode, method).

5) Region of interest (ROI)

Sometimes, we need to process a single sub region of an image, leaving other regions unchanged. It is known as region-of-interest (ROI) processing. ROI or Region of Interest is a portion of an image we want filter or perform operation on it. The pixel values that define the ROI set to 1 and all other pixels set to 0.

Fig. 6: Input image with ROI

6) Segmentation

In image segmentation, a digital image is divided into multiple segments. Image segmentation is usually used to trace objects and boundaries (lines, curves, etc.) in images. In easier language image segmentation is the process of conveying a label to each and every pixel of an image such that the pixels with the similar label share certain characteristics of the particular images that is more lucid then the entire input image.

a) Comparing

Here image is taken from list one by one and compress or expand according to segmented image of an input image. The converted image is matched with the input or scanned image with the help of template matching. If match than we generate standard output as compare to the matched dataset image. After getting the output, the output number is converted into words and then to its respective speech signal.

b) Output image

After the template is matched, the image path is known; from this we can find the image match directory. The standard output is displayed with respect to the match directory.

D. Method for conversion from image to utterance

Fig. 7 shows the flowchart for the conversion of the proposed numeral image to utterance. According to the procedure length of the input number is calculated. If it is a 1 or 2 digit numbers then the corresponding pronunciation text is taken from the Table-I. There are two algorithms given in the next section. Algorithm-1 shows the steps which extracts the individual digits of the number and stores it in an array A. Algorithm-2 generates the Assamese pronunciation for the whole input number from array A. For length 1 and 2 the pronunciation can be directly extracted from Table-I. When length is 3 then 'xo' is appended with the leftmost digit which is in the original Table-I and pronunciation for the remaining 2 digits are also taken from Table-I. For example ১২৫ (125) will be uttered as ১২৫ = ek xo posish

Once the text is obtained, it is converted in its equivalent utterance using Google speech. Similarly, if the length is 4 then *'hazar'* is append with the leftmost digit which is in the original Table-I and then the system starts working on the remaining 3 digits as explained in the above section.

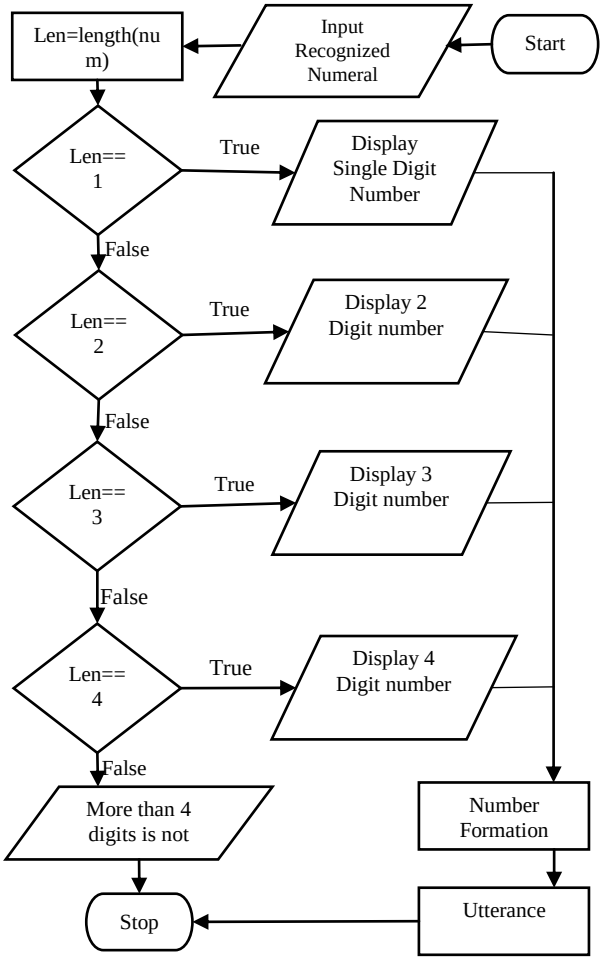

Fig. 7: Flowchart for pronunciation of numerals

Algorithm-1

1. Scan the image from left to right until a white space is encounter.
2. Set length = 0.
3. Extract images of all Assamese digits D_i from the scanned image, compare with the set of standard template images D_t and maximum matching template image D_{t_max} is considered.
4. Store the standard digit against D_{t_max} in an

array *A*.

5. length = length +1

There is a mapping table for all the Assamese numbers from 0 to 99 which is shown in Table-I.

Algorithm-2

1. if length = 1 or 2

   a. Pronounce the digit/number from Table-I in Assamese.

2. if length = 3

   a. Add *'xo'* after pronouncing the 1st element of the array A and then use Table-I to pronounce the remaining two digits.

3. if length = 4

   a. Add *' hazar'* after pronouncing the1st element of the array A and then goto Step-2.

IV. RESULT AND DISCUSSION

We have taken six (6) different images of the same digit for our experiment. Such one example is shown in Table-II.

Table-II: Experimental Result

Numbers	Img 1	Img 2	Img 3	Img 4	Img 5	Img 6	Input Accuracy (%)
১	1	1	0	0	1	1	66.66
২	0	1	1	1	0	1	66.66
১০	1	1	1	1	0	1	83.33
৯৯	1	1	0	1	1	1	83.33
১২৩	0	0	1	1	1	1	66.66
১২৫	1	0	0	1	1	1	66.66
১২০০	0	1	1	0	1	0	50.00
১১৫৬	0	1	1	1	1	1	83.33
Average							70.82

From the above Table-II, it has been observed that the average accuracy of recognition is 70.82 %. As we have taken only 7 images of each digit for training phase, the accuracy found is less. If we increase the number of images in the training dataset there is a chance of getting higher accuracy and the system will give a better result.

V. CONCLUSION AND FUTURE WORKS

This research work takes input of various Assamese numerals images of maximum length 4 using template matching technique. The text is extracted from the images and converted into speech. Template Matching is one of the simplest ways of finding the solution or comparing in the field of Image processing. Finally, after doing the experiment we were able to get a relatively good result and this work will help the visually impaired people to read the Assamese number efficiently. This work can be extended to read currency note in Assamese which will greatly visually impaired people to know the value of currency.

This research work is restricted to Printed Assamese numerals only. Later we can extend it for handwritten numerals or Assamese alphabets, word etc. We can also try to develop the following modules:

- Detection of handwritten numerals.
- Detection of symbols of Assamese language.
- Detection of Assamese modifiers.
- Synthesizer of Assamese characters.
- Segmentation of Cursive word image.

REFERENCES

1. "Assamese Language," 2nd October 2019, Accessed On 3rd October 2019 Available At Https://En.Wikipedia.Org/Wiki/Assamese_Language.
2. "Nefamese," 9th September 2019, Accessed On 3rd October 2019 Available At Https://En.Wikipedia.Org/Wiki/Nefamese.
3. "Nagamese Creole," 2nd September 2019, Accessed On 3rd October 2019 Available At Https://En.Wikipedia.Org/Wiki/Nagamese_Creole.
4. "Assamese Alphabet," 24th September 2019, Accessed On 3rd October 2019 Available At Https://En.Wikipedia.Org/Wiki/Assamese_Alphabet.
5. J. Kalita, D. Talukdar, M. Bora, And M. Barua, "Assam Portal: Assamese Alphabet," 1997, Accessed On 3rd October Available At Https://Assam.Org/Node/2330.
6. M. P. Bhuyan And S.K. Sarma, "An N-Gram Based Model For Predicting Of Word-Formation In Assamese Language," Journal Of Information And Optimization Sciences, 40:2, 427-440, 2019, DOI: 10.1080/02522667.2019.1580883.
7. "Assamese Numbers (অসমীয়া সংখ্যা)," Accessed On 4th October 2019 Available At Https://Www.Omniglot.Com/Language/Numbers/Assamese.Htm
8. P. Sarma, M. Mitra, M.P. Bhuyan, V. Deka, S. Sarmah, And S.K. Sarma, "Automatic Vowel Recognition From Assamese Spoken Words," International Journal Of Innovative Technology And Exploring Engineering, Vol. 8, Issue-10, 2019, Pp. 2297-2304.
9. K. Kalaivani, R. Praveena, V. Anjalipriya, And R. Srimeena, "REAL TIME IMPLEMENTATION OF IMAGE RECOGNITION AND TEXT TO SPEECH CONVERSION,"International Journal Of Advanced Engineering Research And Technology, Vol. 2, Issue 6, 2014, Pp. 171-175.
10. B.B. Chaudhuri And U. Pal, "A COMPLETE PRINTED BANGLA OCR SYSTEM," Pattern Recognition, Vol. 31, No. 5, 1998, Pp. 531-549.
11. Md. A. Hasnat, M. Rahman Chowdhury, M. Khan, "An Open Source Tesseract Based Optical Character Recognizer For Bangla Script," 10th International Conference On Document Analysis And Recognition, 2009, Pp. 671-675.
12. S. Vijayarani And A. Sakila, "TEMPLATE MATCHING TECHNIQUE FOR SEARCHING WORDS IN DOCUMENT IMAGES," International Journal On Cybernetics & Informatics, Vol. 4, No. 6, 2015, Pp. 25-35.
13. R.V. Adhvaryu, "OPTICAL CHARACTER RECOGNITION USING TEMPLATE MATCHING (ALPHABETS & NUMBERS)," International Journal Of Computer Science Engineering And Information Technology Research, Vol. 3, Issue 4, 2013, Pp. 227-232.
14. S. Malakar, P. Ghosh, R. Sarkar, N. Das, S. Basu, And M. Nasipuri, "An Improved Offline Handwritten Character Segmentation Algorithm For Bangla Script," Proceedings Of The Fifth Indian International Conference On Artificial Intelligence, 2011, Pp.71-90.
15. M.P. Bhuyan And S.K. Sarma, "A Statistical Model For Automatic Error Detection And Correction Of Assamese Words," International Journal Of Recent Technology And Engineering, Vol. 8, Issue 2, 2019, Pp. 6111-6116.
16. M.P. Bhuyan And S.K. Sarma, "A Higher-Order N-Gram Model To Enhance Automatic Word Prediction For Assamese Sentences Containing Ambiguous Words," International Journal Of Engineering And Advanced Technology, Vol. 8, Issue 6, 2019, Pp. 2921-2926.
17. M.P. Bhuyan, R. Purkayastha, S.K. Sarma, S. Sarmah, P. Sarma, And V. Deka, "Information Retrieval In Assamese Using Wordnet & Assamese

Wikipedia," International Journal Of Scientific & Technology Research, Vol. 8, Issue 9, 2019, Pp. 1887-1891.

AUTHORS PROFILE

M. Rahman is currently working as an Assistant professor in the department of Information Technology, Gauhati University, Guwahati, India. He did BE in Computer Science & Engineering and MTech in Information Technology and PhD in Natural Language Processing. His research areas are: Natural Language Processing, Data Mining, Speech Processing etc. Email: mirzanurrahman@gmail.com

P. Sarma is currently working as an Assistant professor in the department of Information Technology, Gauhati University, Guwahati, India. She did BE in Computer Science & Engineering, MTech in Information Technology, PhD in Speech Processing. Her research areas are Speech Processing, Machine Learning, Natural Language Processing, Image Processing. Email: parismita.sarma@gmail.com

.

M. P. Bhuyan is a research scholar in the department of Information Technology, Gauhati University, Guwahati, India. He did BE in Computer Science & Engineering and MTech in Information Technology. His research areas are Artificial Intelligence, Natural Language Processing, Parallel Computing and Machine Learning. Email: mpratim250@gmail.com.

A. Das is a student of MCA in NERIM Group of institutions, Guwahati, India. Her research interest is Speech Processing and Natural Language Processing.

D. Dutta is currently working as an Assistant professor in the department of Computer Science, NERIM Group of Institutions, Guwahati, India. He is also research scholar of the department of Information Technology Gauhati University. His research areas are Speech Processing, Artificial Intelligence.