

Design and Implementation of Text of Konkani to Speech Generation System using OCR

John Colaco¹, Sangam Borkar²

¹ Student M.E. (ECI), Dept. Of Electronics & Telecommunication Engineering, Goa College of Engineering, Goa, India

² Asst. Professor, Dept. Of Electronics & Telecommunication Engineering, Goa College of Engineering, Goa, India

Abstract - India is a country of having multi spoken and written languages. Different states have their own languages and many have developed their own language text to speech generation system. Konkani is the officially spoken language of Goa and requires a Text to Speech System. This paper envisages design and implementation of Konkani language Text to Speech system using image recognition technology (Optical Character Recognition). The system is also cost effective and user friendly. In this project work image is converted to text which is then converted to speech using MATLAB. This paper uses both Devanagari and Roman script. This approach will help visually impaired people for reading the Text documents and books in Konkani language.

Key Words: Optical Character Recognition, Segmentation, Feature extraction, Classification, Text Processing, Speech synthesis.

1. INTRODUCTION

Optical character recognition (OCR) is the technique of translating text images into an encoded text form. This technology allows a computer to automatically recognize characters through an optical system. OCR can identify both handwritten text as well as printed text. But depending up on quality of input documents the performance of OCR is measured. This system is designed to process images that consist almost entirely of text. This technique has four steps namely Image acquisition/scanning, preprocessing, Segmentation, Feature extraction and classification. Fig-1 shows the Block diagram of OCR. In this firstly, the preprocessing module prepares image for recognition and it involves normalization, binarization and filtering. Secondly Image is segmented to separate out the characters from each other and it involves graphics, text lines, words and characters. Then by removing special characteristics and patterns of an image in the feature extraction stage the character image is taken higher level.

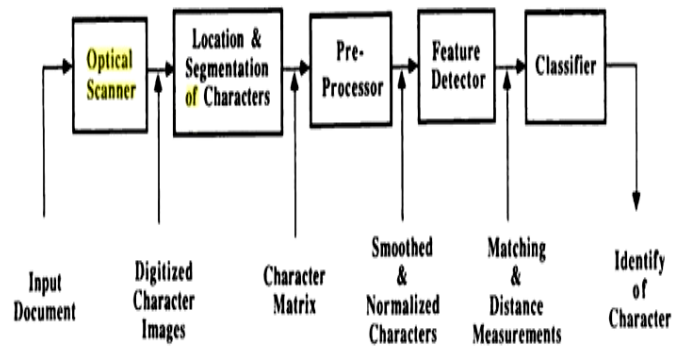


Fig -1: Block diagram of OCR

The classification stage identifies each input character image by taking into account the detected features. Various types of classifiers are in use for this purpose such as Hidden Markov Models, Bayesian theory, Template Matching Neural Networks, Syntactical Analysis, SVM etc. OCRs cover wide range of applications in government and business organizations, as well as individual companies and industries. Some of the major applications of OCR include:

- (i) Document reader systems for the visually impaired.
- (ii) Bank check and Form processing,
- (iii) Office and Library automation,

1.1 Introduction to Devanagari script

Devanagari is a Brahmic script. It is widely used script in India. Many Indian languages like Marathi, Nepali, Hindi, and Sanskrit are using it. It was also formerly used to write Gujarati. Etymologically, the word Devanagari has combination of two Sanskrit words one is 'nagara' means city 'and other is Deva' means God, Brahma or sometime the king. The Devanagari script represents the sounds which are consistent. In this script, each character has horizontal bar which is called "Shiro Rekha" at the top as shown below in Fig -2. Text word is divided into three zones. One is the upper zone which represents the part above the headline, other is the middle zone which covers the part of basic and compound characters below the headline and last is the lower zone contain where some vowel and consonant modifiers that can reside.

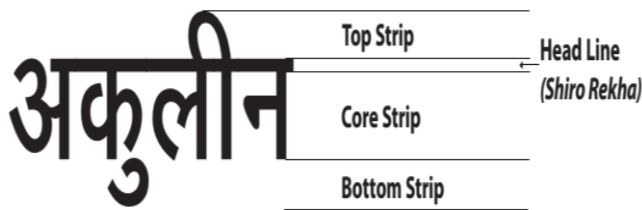


Fig -2: Three Part of Devanagari Word

The script is wrote from left side to right side. Letters hang from a head stroke is generally constant throughout the length of the word. It is very helpful to become known with the traditional order of Devanagari as shown in the Fig -3. It is read as if it were text, left to right and top to bottom.

1	अ	आ	इ	ई	उ	ऊ	ऋ
2	ए	ऐ	ओ	औ	अं	अः	
3	क	ख	ग	घ	ङ		
4	च	छ	ज	झ	ञ		
5	ट	ठ	ड	ढ	ण		
6	त	थ	द	ध	न		
7	प	फ	ब	भ	म		
8	य	र	ल	व			
9	श	ष	स				
10	ह						
11	क	ख	ग	ज	फ	ड	

Fig -3: The traditional order of Devanagari

Devanagari script has 18 vowels out of which 11 are mostly used as shown in Fig -5. Vowels are transliterated in two distinct forms one is the dependent form i.e. "matra" and other is the independent form. When the vowel letter emerges in single form, at the start of a word, or next another vowel letter then the independent form is used. "Matras" are used when the vowel follow a consonant. Devanagari script has 33 consonants which are arranged phonetically. The first set has 25 consonants which are occlusive consonant, as shown in Fig -6 and remaining 8 are non occlusive consonant as shown in Fig -7. The occlusive consonants have five groups: cerebral or retroflex labials, gutturals, palatals, and dentals. The last four consonants have two groups: plosive and voiced plosive and nasal consonant is the first consonant. The plosive and voiced plosive are un-aspirated and aspirated version both having one character. Vowels in combination with consonants (mātrās) always appear into view with one of the consonants. Fig-4 Shows how Devanagari characters are translated into English form. By using phonetic map, transliteration unit converts each syllabic unit in Devanagari into English. Thus Phonetic map is executed by using the translation memory.

1	अ	आ	इ	ई	उ	ऊ	ऋ
	a	ā	i	ī	u	ū	r
2	ए	ऐ	ओ	औ	अं	अः	
	e	ai	o	au	am	ah	
3	क	ख	ग	घ	ङ		
	ka	kha	ga	gha	ṅa		
4	च	छ	ज	झ	ञ		
	ca	cha	ja	jha	ña		
5	ट	ठ	ड	ढ	ण		
	ta	tha	ḍa	dha	ṇa		
6	त	थ	द	ध	न		
	ta	tha	da	dha	na		
7	प	फ	ब	भ	म		
	pa	pha	ba	bha	ma		
8	य	र	ल	व			
	ya	ra	la	va			
9	श	ष	स				
	śa	ṣa	sa				
10	ह	क्ष	त्र	ज्ञ	श्र		
	ha	kṣa	tra	jña/gya	śra		
11	क	ख	ग	ज	फ	ड	ढ
	qa	kha	ga	za	fa	ra	rha

Fig -4: Devanagari characters and their transliteration

Independent form	Modifier or Matras	Independent form	Modifier or Matras
अ	None	ए	े
आ	ा	ऐ	ै
इ	ि	औ	ौ
ई	ी	ओ	ो
उ	ु	औ	ौ
ऊ	ू	ओ	ो
ऋ	ृ	लृ	लृ
ॠ	ॠ	लृ	लृ

Fig -5: Vowels in Devanagari

	Plosive		Voiced Plosive		Nasal
	Unaspirated	Aspirated	Unaspirated	Aspirated	
Gutturals	क	ख	ग	घ	ङ
Palatals	च	छ	ज	झ	ञ
Cerebrals	ट	ठ	ड	ढ	ण
Dentals	त	थ	द	ध	न
Labials	प	फ	ब	भ	म

Fig -6: Non Occlusive Consonants

Semivowels	य	र	ल	व
Sibilants	श	ष	स	
Aspirate	ह			

Fig -7: Occlusive Consonants

1.2 Introduction to Text to Speech synthesis

Speech synthesis is the non-natural creation of human speech. For this purpose speech synthesizer has been used and this synthesizer can be put into operation in hardware as well as in software. An ordinary language text into speech has been converted by text-to-speech (TTS) system and other systems make representative linguistic presentation like phonetic notation into speech. The speech which is synthesized is made by joining together section of recorded speech which is previously stored in a database. A synthesizer joins together a depiction of the oral region and other human voice to get a "artificial" output voice.

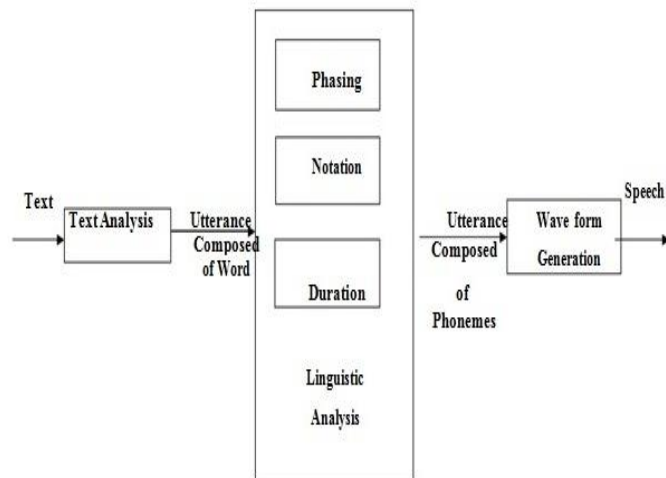


Fig -7: Processing system of Text to speech

Above fig -7 shows how text is processed into speech. A text-to-speech system is composed of two parts known as front-end and back-end. The front-end consist of two tasks. First task is that, it converts raw text having symbols like abbreviations and numbers into the equivalent of written-out words. This process is called pre-processing text normalization, or tokenization. The front-end then allots phonetic transcriptions to each word, and breaks up and marks the text into prosodic units, like sentences, clauses, and phrases. The process of assigning phonetic transcriptions to words is known as text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and Prosody information together make up the symbolic linguistic representation which is output by the front-end. The synthesizer is often known to as the back-end, and then

converts the symbolic linguistic representation into necessary sound.

2. DESIGN OF PROPOSED OCR SYSTEM

In the design of proposed OCR system following steps has been followed:

- Preprocessing;
- Segmentation;
- Feature Extraction;
- Classification.

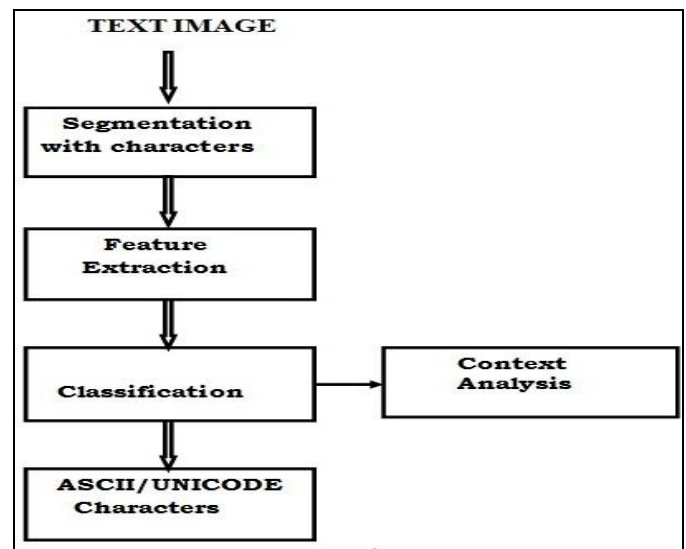


Fig -8: Stages in the design of OCR

1. Preprocessing

In this, firstly the Text image in .jpeg format is acquired and read with the help of imread command. Then digitization of text is done by using an optical scanner. Pre-processing includes Binarization, normalization and filtering. Binarization converts a gray-level scale image into binary-level image by the process called thresholding. In this thresholding process ,if the pixel is found to be larger than or equal to this threshold, it is outputted as 1 otherwise it is outputted as 0. A pixel becomes white if its gray levels is < TA pixel becomes black if its gray levels is >= T. The Image is converted into gray scale by rgb2gray command. Rgb2gray command eliminates the hue and saturation information while retaining the luminance. Then we perform image filtering for reducing salt and pepper noise by using median filter. This is done by using command imgen = medfilt2 (imagen). Normalization has been done to obtain characters of uniform size.

2. Segmentation

Segmentation is the technique where the text image is divided into its constituent objects or regions. This technique increases the performance of OCR. The segmentation methods such as line segmentation, word segmentation and character segmentation are proposed where Text lines are segmented first then words and finally characters. During segmentation of line, before a black pixel last row containing all white pixels are found first and then just after the end of black, the first row containing entire white pixel pixels are found. During Segmentation of Individual Characters, the rows having maximum number of black pixels in a word are found first, then the headlines of devanagri script are located and it is removed by converting them into white pixels. Then by applying vertical scanning, individual characters are split from each zone.

3. Feature Extraction

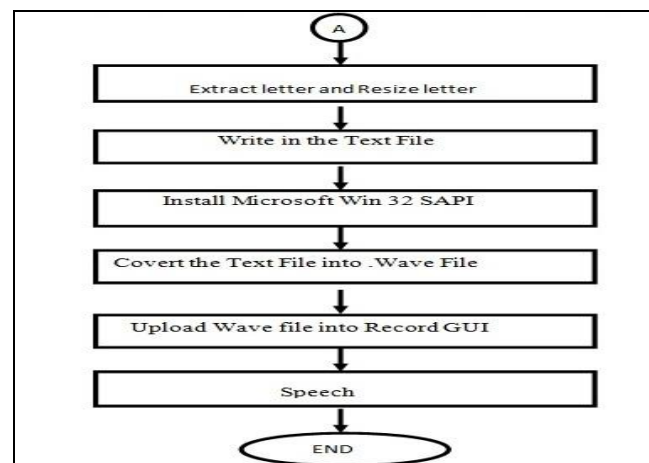
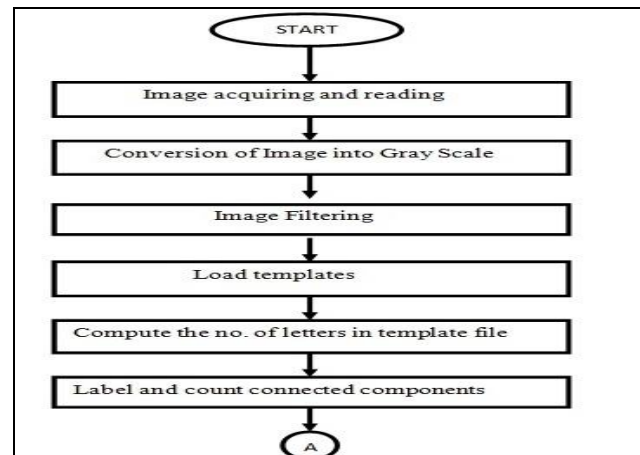
Feature extraction stage extracts a set of features and this helps in maximizing the recognition rate. Template matching which are the main techniques for extraction of features has been exploited. Template matching find the location of a sub image called a template inside an image. For recognition to occur templates are loaded. The existing input character is match up to each template to find either an exact match, or the template with the closest depiction of the input character. Letters are extracted and resized them by imresize command. Then open the text.txt as file for write by the using fopen command.

4. Classification

The classification identifies each character and assigns it to the correct character class. Classification performed is based on the extracted features. For classification and recognition, Artificial Neural Networks technique has been explored. Artificial neural networks are the arrangement of interconnected "neurons" which can compute values from inputs. Neural networks do the mapping by setting of input values to setting of output values.

3. FLOWCHART

Below flowchart illustrates the conversion of text image into text file and then converting it into speech.



4. CONCLUSIONS

This paper gives an approach for converting text image of both Devanagri and Roman Script into readable text using optical character recognition technique and then converting this text into speech by text to speech technology. Therefore this technology will help People with bad vision or totally blindness goan people to read a text loud. This approach will also help in reading roman script document.

REFERENCES

- [1] Ambadas B. Shinde, yogesh H. Dandawate., "Shirorekha Extraction in Character Segmentation for Printed Devanagri Text in Document Image Processing classifier," IEEE , Dec. 2014, pp. 1-7. DOI: 10.1109/indicon.2014.7030535.
- [2] S.Farkya ,G.Surampudi,A.Kothari, "Hindi Speech Synthesis by concatenation of Recognized Hand Written Devnagri using support vector machines classifier," IEEE , Dec. 2014, pp. 0893 - 0898. DOI: 10.1109/ICCSP.2015.7322625.
- [3] Pooja Chandran, Aravind S, Jisha Gopinath and Saranya S S, "Design and Implementation of Speech Generation System using MATLAB," IJEIT, vol.4, Dec. 2014.
- [4] Ravina Mithe, Supriya Indalkar, Nilam Divekar, "Optical Character Recognition," IJRTE, vol.2, March. 2013.
- [5] Mustafa Saifee, "Devanagari Font Design for Optical Character Recognition," thesis, May. 2012.
- [6] N.Swetha and K.Auradha, "Text to Speech Conversion," IJATCE, vol.2, Dec. 2013, pp.269-278.
- [7] Mrs. S. D. Suryawanshi, Mrs. R. R. Itkarkar, Mr. D. T. Mane, "High Quality Text to Speech Synthesizer using Phonetic Integration," IJARECE vol.3, Feb. 2014.