

# A Review on Optical Character Recognition and Text to Speech Conversion

Swati Vikas Kodgire<sup>1</sup>, G. S. Sable<sup>2</sup>

<sup>1,2</sup>Maharashtra Institute of Technology, Beed By Pass Road, Satara Parisar, Aurangabad 431010, Maharashtra, India

**Abstract:** *The application depending on image and voice with a parallel functioning is suitable to assist physically challenged people. So that dependability of a challenged person is decreased to a improved level. Image acquisition based text reader can help visually challenged people to manage the handheld objects in day to day life. Initially steps involves capturing of image, distinguishing image with text portion and residual regions, image pre-processing on region of interest, after the extraction of characters and words, conversion of text to speech is done. To splinter text from a document it is obligatory to discover all the possible manuscript text regions. Text detection, line detection, character identification, feature extraction, training of extracted features are the steps in sequence that are executed.*

**Keywords:** Region of interest (ROI), Optical character recognition (OCR), Text to speech conversion (TTS), Binary image (BW).

## 1. Introduction

Independence for disabled persons is of chief significance. This paper presents a study of effective prototype system to help blind persons to become self-governing. Assistive technology is a development of a product which can be customized in a way that assists the challenged people to collect more particulars as that of normal people. At present there are abundant assistive technologies devices such as vibrating watch and talker device are common examples. Pictures with transcript act as central communication medium for conveying information. Important phase of consideration are text existence and character recognition in the assistive methods.

Two basic methods for obtaining region of interest is rule based method and learning based method. In rule based method, definite threshold values are preferred as cause of valuation. Opposite to this is learning methods in these neural networks are trained for gathering the information about picture. Printed text information is widely stretched in today's scenarios for example in consumer product labels, bank forms, receipts etc. There are assured devices like magnifying glasses, forms of optical aids that aid blind users for obtaining transcript information. Current environment helps blind user in product text reading is bar code reader with modest complexity in locating the barcode. When it comes for text extraction it is obvious that the blind users will hold the exact manner of product for recognition. There is possibility for holding the product box in upside down or with some orientation of box. So a process to identify text even with some orientation is needed.

In assistive reading devices especially for blind users it is tricky to guess that they will hold the product which is suitable for text extraction or simply region of interest. A hybrid method needs to be adapted helps to detect oriented product labels too.

Initial stage is to identify the location of text in the image. Segmentation is the second stage that isolates the lines, words and ultimately the characters. The text style can vary

in terms of size, font and formats, background and foreground, color etc. After the text fragmentation process, it is necessary to focus on correct character recognition and final part is conversion text into speech.

## 2. Traditional Methods

Traditionally, pen scanners were accessed that were designed for simple background images, standard font styles and small font sizes. The drawback of this system was that the scanner could not handle scene images with complex facilities or other than the specified points. KReader Mobile activates on a handset and allows user to access email, messages and other documents. The short coming of this unit the data must be only on a messed up background also it cannot recognize text with colored background, but the application produces better results for dark colour background with light colour text. KReader cannot access text on cylindrical objects [1].

When the stick is scanned over the printed letters, OCR makes out the text and transforms the information into voice. The voice is then read back and thus helps visually challenged. The unit consists of optical character recognition and text to speech technology and headphones. The system takes a longer time to read, limited publication accessibility, expensive device, also cannot handle huge data [2].

There are currently few systems that propose a promising portable application but cannot handle larger data sets. Bar code readers help blind people to identify the grocery products to access information through speech and braille. The limitation of the system is that blind people find it difficult to locate the barcode scanner exactly on the barcode [3].

A finger worn device assists the visually impaired with effectively and efficiently reading paper-printed text. A novel device is introduce a local-sequential manner for scanning text which enables reading lines, blocks of text for central section's while providing real-time auditory and tactile feedback. The design is stimulated by introductory

Volume 5 Issue 6, June 2016

[www.ijsr.net](http://www.ijsr.net)

[Licensed Under Creative Commons Attribution CC BY](https://creativecommons.org/licenses/by/4.0/)

studies with visually impaired people, and it is small-scale and mobile, which enables a more manageable operation with little setup. Finger Reader is an index-finger wearable device that supports the blind in reading printed text by scanning with the finger. The finger-worn design helps focus the camera at a fixed distance and utilizes the sense of touch when scanning the surface. Also, the device does not have many buttons or parts in order to provide a simple interface for users and easily orient the device. The device developed has some misleads such as if a blind person takes a break in between he is unable to relocate the place and hence delay occurs in finding the correct position [4].

### 3. Literature Review

The overall review is separated into four subsections. The first section presents the idea behind pre processing concept and need of pre processing and post processing. The second section consists of review of isolation of non text and textual regions of an image. The third section focuses on various techniques used for image to text conversion. The fourth section covers the brief review on text to speech conversion methods.

#### 3.1 Basic preprocessing steps

Pre processing is a series of necessary to get a good image in terms of filtering, area of interest and quality. It essentially enhances the image rendering it suitable for further process. Database created from camera or scanner must be analyzed through all these phases. The next step is to convert color image to gray and then from gray to binary, process is named as binarization, mostly all digital image processing is done on bi-tonal images rather than multi-tonal by using some threshold value. Morphological operations such as thinning, connected components label, erosion and dilation etc can be performed. Authors have used some or the other filter to remove noise. Smoothing and skew correction is also done in this phase.

Poovizhi P worked on handwritten text recognition techniques. The main objective is to pre process the raw image for the development of OCR. As compared with printed character recognition, handwritten character recognition has challenging issue due to following factors such as individual styles of writing, speed of writing, size of letters, physical and mental condition of the writer, overlapping of letters etc. Noise in the image gets added due to poor photocopy. Binarization, normalization, sampling, de-noising, thinning are the steps carried out. Gray scale image is converted to binary i.e. bi-level, two levels. Otsu's method is used to perform threshold based on cluster. The process of changing the intensity value of pixel to the range of [0,1] is called normalization. The process of selecting the subset of individuals from the large sample of population. Author has studied comparative results of filters such as medfilt2, salt and pepper, ordfilt2. Thinning results in single pixel width image to recognize the handwritten characters [5].

Megan Elmore proposed image pre preprocessing suite, through text detection, auto rotation, and noise reduction improves the accuracy of OCR analysis in a camera based

translation system. The system determines background and fore ground and skew estimation using morphological edge analysis. The system translates a color image to binary i.e. it classifies image pixels as text (black) and background (white). So that it helps in segmentation of text region with the background. Binarization provides sharp contrast. There are two major classes of thresholding: i) global thresholding decides one partition for the entire image and ii) local that considers the properties within smaller regions. After that, text detection is done based using method like i) color based text detection and Heuristic connected component filtering ii) by edge based text detection and mathematical morphology method iii) texture based text detection [6].

Maya R. Gupta et.al worked on problem of document binarization as a pre processing step for optical character recognition (OCR) for the purpose of keyword search of historical printed documents. They tried to implement a number of promising techniques opted from the literature for binarization, pre-filtering, and post-binarization denoising along with newly developed methods for binarization: an error diffusion binarization, a multi-resolution version of Otsu's binarization, and denoising. Binarization approaches such as global fixed threshold, Otsu threshold, multi resolution Otsu, Chang's method, Sauval-Niblack method, margin error diffusion method, Markov model for OCR binarization [7]

Safoora O.K. et.al proposed Optical character recognition (OCR) method that is one of the most successful applications of pattern recognition and image processing. Character geometry is a useful feature for identifying characters in images. The geometric feature extraction techniques proposed in literature are complex and requires extensive effort in implementation. They proposed a preprocessing technique which is simple and makes the feature extraction process easy. The proposed system is intended for preprocessing color images containing printed English alphabets. The image is fed to a preprocessing unit where it is first converted to gray and then threshold to obtain a binary image. Next step is to filter the noise present in the image. Now the area containing text will be identified and segmented. Each character is extracted from the segmented area in the image and applied morphological thinning operation to make the feature extraction process easy [8].

#### 3.2 Review on isolation of non text and textual regions

Vandana Gupta et.al proposed a system based on region identification of script from an image. This implies division and classification of the parts of the document into blocks of text and graphics. Separation of the document components into blocks of text and graphics increases the performance of the document investigation system. They used variance property of feature extraction that reduces the dimensionality. Transformation of input data into the set of features is called feature extraction. Step variance based approach is implemented, variance is calculated for each row of entire image and it is stored in one array of dimension number of row \*1. The difference is found out between two calculated consecutive variance values of array and it is stored in new array. A graph is drawn between number of

rows and variance values calculated from previous step. Same process is applied to calculate column wise variance. Then a common threshold value is calculated from these two graphs (row wise graph and column wise graph) such that all pixel variance values those are greater than and equal to this threshold value i.e text part and those pixel variance values are lower than threshold value i.e non-text part [9].

Ankush Gautam et.al proposed text segmentation method to segregate text from embedded images. He employed SVM let wavelet and 2-mean classification for segmentation of text from image document. He has used morphology operation like as dilation and erosion in post processing. The wavelet transform provides a multi-resolution representation. Symlet wavelets are in a group of wavelets. They are a modified version of Daubechies wavelets with increased symmetry. Symlet Wavelet $[n]$  is defined for any positive integer  $n$ . The scaling function ( $\phi$ ) and wavelet function ( $\psi$ ) have compact support length of  $2n$ . The scaling function has  $n$  vanishing moments. Symlet wavelet can be used with functions as Discrete Wavelet Transform. The k-means algorithm is an evolutionary algorithm clusters observations into  $k$  groups, where  $k$  is provided as an input parameter. They used 2-means classification in our implementation. One group of white pixel and second group of black pixel are used in 2-means classification [10].

Danial Md Nor et.al introduced a method for the investigation of text segmentation from a picture. The image captured is set to a fixed resolution using sampling approach. The team used RLSA (Run length smoothing algorithm) that is applied row-by-row and column-by-column to the mentioned binary document images representing white pixels by 1's and black pixels by 0's. The RLSA transforms a binary sequence  $x$  into an output sequence  $y$ . RLSA is used to connect the neighboring black areas when they are separated by less than threshold pixels. The degree of connectivity depends on threshold, the distribution of white and black pixels in the document and the 'dpi' (dots per inch) resolution of the document. The two distinct bit-maps are generated using the RLSA in both horizontal and vertical directions. The spacing between the components in the document image tends to differ horizontally and vertically. So two different thresholds  $T_h$  and  $T_v$  are used for the RLSA in respective horizontal and vertical direction. ( $T_h = 80$  and  $T_v = 100$ ). The two bit-maps of respective RLSA output in horizontal and vertical directions are combined using a logical AND operator to detect various components in the document images [11].

Rodolfo P. Dos Santos et.al used text line algorithm that contains a morphological operator to obtain the features of the images. A sequence of histogram projection and recovery is proposed to obtain the line segmented region of the text. Y histogram projection is performed which results in the text lines positions. To divide the lines in different regions a threshold is applied. Second threshold is used to eliminate false lines. Sometimes it causes some loss on the text line area. So, a recovery method is proposed to minimize this effect. In order to detect the extreme positions of the text in the horizontal direction, an X histogram projection is applied. Now in the Y direction, another threshold is used to

eliminate false words. In order to optimize the area of the manuscript text line, a text selection is carried out. Experimental results are described for false line detection and false word detection and exclusion [12].

Neha Gupta et.al contributed to image segmentation element. They employed discrete wavelet transform (DWT) for extracting text information from complex images. For extracting text edges, the sobel edge detector is applied on each sub-image. The resultant edges so obtained are used to form an edge map. They used morphological operations that can be applied on the processed edge map and thresholding is done to improve the performance. By using HAAR discrete wavelet transform most of the textured images are well characterized by their contained edges. The team used 2 D DWT that decomposes input image into four components or sub bands one average component and three detail component. The drawback of this system is that there might be incidence of missing of characters when the image has poor contrast with respect to background [13].

Nikolaos G. Bourbakis et.al presented a methodology for document processing, by separating text paragraphs from images. The methodology is based on the recognition of text characters and words for the efficient separation text paragraphs from images by keeping their relationships for a possible reconstruction of the original page. The text separation and extraction is based on a hierarchical framing process. The process starts with the framing of a single character, after its recognition, continues with the recognition and framing of a word, and ends with the framing of all text lines. A technique for text-paragraphs and images separating has been presented in this paper. The main advantages of this method are: accurate extraction and recognition of text-paragraphs from documents; extraction and recognition of handwritten unstructured text from documents; The weak point of this method is that it is slower than the other document- processing methods due to words recognition effort which the other methods do not use [14].

### 3.3 Review on image to text conversion methodologies

Li Zhao et.al presented their work on Optical Character Recognition (OCR) analyze a reference implementation of the OCR workload on a low power general purpose processor and identify the primary hotspot functions that incur a large fraction of the overall response time. They implemented and analyzed several software/algorithmic optimizations such as i) Multi-threading, ii) Image sampling for a hotspot function and iii) miscellaneous code optimization. Results show that up to 2X performance improvement in execution time of the application and almost 9X improvement for a hotspot can be achieved by using various software optimizations. They have used OCRopu0.3 as it is open source software and supports modularity. RAST based analysis is used that includes column finding, text line identification and reading order determination. OCRopus incorporates Tesseract as text line recognition engine. The developed system requires high clarity images for efficient results [15].

Yafang Xue et.al introduced two implementations in optical character recognition using template matching method and

feature extraction method followed by support vector machine classification. After pre processing, the texts are segmented into isolated characters and the correlations between a single character and a given set of templates are computed to find the similarities and then identify the input character. In the second method, features extracted from the segmented characters are used to train the SVM classifiers. In Template Matching method, the recognition is based on measuring the similarities between the structure of the input image and a given set of templates. This method is sensitive to template mismatch when the input characters are not exactly the same font as the templates. Errors occur when the font of the input image (Ariel) is different from the template. The system misreads letter I, Q and R to O, O and P, respectively. Because of the slightly change in the structure of the input characters, the highest matches of certain characters are not found in their true corresponding templates of letters or digits. The recognition accuracy is highly affected by the font of the input characters. The classifiers created by feature extraction seem to have a relatively low accuracy in recognition with the highest accuracy being 83.3%. A larger dataset might improve the performance of the classifiers. Although with the relatively poor classification accuracy of the two sets of classifiers, the advantage of the HOG features extraction over raw pixels values is still notable [16].

Nadira Muda et.al. described a system prototype that has its own scopes which are using Template Matching as the algorithm applied to recognize the characters, characters to be tested are alphabet (A – Z), grey-scale images were used with Times New Roman font type, using bitmap image format with 240 x 240 image size and recognizing the alphabet by comparing between two images. Matlab R2011a is the software tool that was used in developing the system prototype. The team used a fixed set of database alphabets (A to Z), with times new roman font type, of size 240 X 240. Provide the input image to the system, flow proceeds as it rescales the image irrespective of the size, then computes the matching range, if the highest match is found system stores the value, it checks whether the character is bounded and then stores the best match of the recognized character finally terminate the loop. The system is limited to font category, if capital letter templates are used it will not work for small letter matching and vice versa. They collected the database of about hundred different Sanskrit characters. SVM based and kNN method has been employed to propose a two stage multi classifier for about one hundred classes. kNN is assigned as an input query sample to a group containing k possible labels. Second stage SVM multi classifier helps to classify between labels in the group. The correct recognition rate is 85.6% using one-against-all and 84.57% using one-against-one multi classifier, respectively [17].

S.K.Thilagavathy et.al worked on the edge detection algorithm that has a list called traverse list. It is the list of pixel already traversed by the algorithm. The Edge Detection algorithm terminates when it has covered all the pixels of the character as every pixel's position would be in Traverse List so any further call to Edge Detection is prevented. The Edge Detection algorithm terminates when it has covered all the pixels of the Character as every pixel's position would be in

Traverse List so any further call to Edge Detection is prevented. Diagonal feature extraction scheme for recognizing off-line handwritten characters is proposed in this work.

Create a Traverse List - List of pixels which have been already traversed. This list is initially empty. Scan row Pixel-by-Pixel. Whenever a black pixel is found check whether the pixel is already in the traverse list, if it is simply ignore and move on else apply Edge detection Algorithm. Add the List of Pixels returned by Edge detection Algorithm to Traverse List [18].

Sameeksha Barve used ANN that is trained using Back Propagation algorithm. In this, each English alphabet is represented by binary numbers that are used as input to feature extraction system, output of this and original input is feed to ANN. Back propagation algorithm comprises of training, calculating errors and modifying weights. Initially neural network is trained and knowledge is stored within inter neuron connection strengths known as synaptic weights. Here she have used two hidden layer multilayer perception (MLP). The inputs are fed to input layer and get multiplied by interconnection weights. Within first hidden layer they get summed then processed by a non linear function. Finally data is multiplied by interconnection weights then processed last time within the output layer to produce the neural network output [19].

Pardeep Kaur et .al used neural network for character recognition increases performance or accuracy of character as compared to conventional scanning devices. Linear discriminant analysis method is used for data classification and dimensionality reduction. MDA is preferred because it is capable of deriving discriminant subspace from large scale training data for classification. It seeks and subspace where samples from same class are gathered while samples from different classes are separated, thus simplifying classification. They have used sobel edge detection algorithm. After the initial preprocessing steps features such as crossing and distances, cropping, area perimeter are determined. They have developed a GUI that shows comparative results of both MDA and NN technique. Experiments show the results of mean square error and peak signal to noise ratio. Also matching time and matched and miss matched character results are displayed. Matching time required is less [20].

Sandeep Tiwari et.al contributed to OCR process on the basis of performance evaluation of handwritten text. They followed the complete process of pre and post processing. Segmentation is performed by isolating each connected component. Some drawback of segmentation is that consecutive character if joined are detected as a single character, noise present rarely cannot be filtered and may result in undesirable outputs. Feature extraction is categorized in distribution of points, transformation and series expansions and structural analysis. Performance rates investigated are recognition rate, rejection rate, error rate. Recognition rate is 85 to 90 % due improper handwritten text but it increases to 90 to 95 % for printed text [21].

Sandeep Tiwari et.al have proposed SURF feature and neural network based OCR has the capability of strong robustness performance and good distinction between feature points that improve computing speed. For neural networks classification pixels derived from the resized character in segmentation stage form the input to classifier. Classifier consists of two hidden layers beside input and output layer. SURF (Speeded Up Robust Features) feature matching approach is used to match the corresponding features between two images. Euclidean distance measures the nearest and second nearest neighbors. But the method only matches feature points from a single direction. The disadvantage is error rate is high and one-to-many matching are often appeared in the results. Correct classification rate is calculated for both the methods Surf and NN. CCR table summarizes the SURF, NN, and proposed SURF + N, PNSR and MSE values. NN has lowest rank and SURF is in between and by combining of these both techniques system outsources better results [22].

Reetika Verma et.al presented an offline handwritten alphabetical character recognition system using multi layer feed forward neural network. Initially a 40 point feature extraction method is used so the output becomes the input to the NN. The entire preprocessed image is divided into 40 zones. They have collected dataset of 1040 images of upper case English alphabets of various appearances divided into training and testing sets. The objective was to determine from given text of A to Z. The different attributes and character morphology of single alphabets are highlighted by the 40 point feature extraction technique and stored in the Matlab created Neural Network. When testing is done the features extracted from the tested character are simultaneously matched with those previously stored in the neural network. The maximum percentage of matching of the features extracted from the training and the testing characters give the resultant alphabet as output in a graph [23].

### 3.4. Review on text to speech conversion methods

Chaw Su Thu et.al worked on two aspects OCR and TTS. After system recognizes the character the output is given to speech conversion process. Win 32SAPI is used for getting the voice object and load the library. It initializes the wave player for the conversion of text to speech. They have implemented graphical user interface. The disadvantage of the system is that single character is converted to text to speech system is not much friendly for number of words [24].

K. Kalaivani et.al proposed text to speech synthesizer. The theme is based on horizontal text extraction of text documents and then conversion into speech. The text book or a document is placed under the mechanical set which consists web camera, captures the images from the text book. Captured image is loaded in the GUI. GUI have the facility to preview image and then capture the perfect image, the text in the image is spontaneously displayed on the text block. After the basic pre-processing step text is extracted and converted into speech using text to speech synthesizer [25].

Tapas Kumar Patra et.al proposed text to speech conversion using matrix operations. By the use of microphone some similar sounding words are recorded using record program in MATLAB and recorded sounds are saved in .wav format in the directory. These recorded sounds are sampled and the sampled values are taken and separated into their constituent phonetics. The separated syllables are then concatenated to reconstruct the desired words. Sound recording is initiated through the Matlab graphical user interface (GUI) by clicking on the record button. The duration of the recording can be adjusted to be anywhere from 1 to 6 seconds. Once recorded, the time data is normalized to maximum amplitude of 0.99 and displayed on the upper plot in the GUI window. In addition to the time domain waveform, a spectrogram is computed using Matlab's built in spectrogram function. When a sound is recorded, its spectrogram i.e. time domain representation can be viewed. By studying the various amplitude (power) peaks in the spectrogram, the region of presence of phonemes can be detected. Using the wav read commands sounds of .wav extension stored in the Matlab directory, can be imported into Matlab workspace. The sounds imported into the workspace are a sampled version. Other details like sampling frequency, number of bits per sample, etc. can also be stored [26].

G. D. Ramteke et.al proposed text to speech for Marathi numerals written in devanagari script. The system is based on rule based approach. The database is created with recording of numerals through speech of various people (male / female) and with different age groups. Features of generated speech signals are extracted by pitch detection algorithm. Mean and standard deviation is calculated for pitch detection of each voice samples. A TTS process is to determine the pronunciation of a word based on its spelling which is often called Grapheme-to-phoneme (G2P). Grapheme is written form in specific language. G2P conversion by rule is perhaps the classic application of tradition knowledge based rules in TTS [27].

R. Shantha Selva Kumari et.al worked on the conversion of English text into speech is done by using a stored speech signal data. Text to speech conversion module is designed by the use of matlab. By the use of microphone the phonemes (alphabets, numbers, and words) are recorded using gold wave software. The recorded .wav (sounds) files are saved as a database separately. The phonemes are extracted from the text file. For text to speech conversion the concatenation method is proposed. The recorded speech are concatenated together to produce the synthesized speech. The resulting speech output is assessed by listening test. The Mean Opinion Score (MOS) value is calculated for the synthesized speech output as the performance measure [28].

Chucui Li et.al proposed a camera-based assistive text reading framework to help blind persons read text labels and product packaging from hand-held objects in their daily lives. To automatically localize the text regions from the object ROI, they used a novel text localization algorithm by learning gradient features of stroke orientations and distributions of edge pixels in an Adaboost model. Text characters in the localized text regions are binarized and recognized by optical character recognition (OCR) software.

The system framework consists of three functional components: scene capture, data processing and audio output. The scene capture component collects scenes containing objects of interest in the form of images or video. In their prototype, it corresponds to a camera attached to a pair of sunglasses. The data processing component is used for deploying our proposed algorithms, including i) Object-of-interest detection to selectively extract the image of the object held by the blind user from the cluttered background or other neutral objects in the camera view; and ii) Text localization to obtain image regions containing text, and text recognition to transform image-based text information into readable codes.

They used a min-laptop as the processing device in current prototype system. The audio output component is to inform the blind user of recognized text codes. A blue tooth earpiece with mini-microphone is employed for speech output [1].

#### 4. Discussions

The paper reviews different techniques on OCR and TTS techniques available and summarizes them. But it is found that the techniques that provide better results are slow in nature while fast techniques result inefficient and complex design. Survey suggests that OCR techniques based on neural networks provide enhanced results than others. This review establishes a system that converts scanned images of optical characters to text documents. Various techniques that are discussed have its own uniqueness and level of accuracy; in future researchers can do different modifications of the system.

#### References

- [1] Chucai Yi, YingLi Tian, Aries Ardit. Eason, B. Noble, and I. N. Sneddon, "Portable Camera-based Assistive Text and Product Label Reading from Hand-held Objects for Blind Persons", *Submit to IEEE/ASME Transactions on Mechatronics*.
- [2] <http://www.tuvie.com/voice-stick-portable-text-scanning-device-for-the-visually-impaired>
- [3] Douglas Chai and Florian Hock, "Locating and Decoding EAN-13 Barcodes from Images Captured by Digital Cameras", 2005 IEEE
- [4] Roy Shilkrot, Pattie Maes, Jochen Huber, Suranga C. Nanayakkara, Connie K. Liu "FingerReader: A Wearable Device to Support Text Reading on the Go".
- [5] Poovizhi P, "Study on Preprocessing Techniques for Character Recognition", *International Journal of Open Information Technologies*, 2014
- [6] Megan Elmore, Margaret Martonosi, "A Morphological Image Pre-processing Suite for OCR on Natural Scene Images".
- [7] Maya R. Gupta, Nathaniel P, Jacobson, Eric K Gracia, "OCR binarization and image pre-processing for searching historical documents", 2006 *Pattern Recognition Society*, Published by Elsevier Ltd.
- [8] Vandana Gupta, Kanchan Singh, Ashok K Sinha, "A Novel approach for classification of Text and Non Text data from an image", *International journal of computing*, 2014.
- [9] Ankush Gautam, "Segmentation of Text from image document", *International journal of computer science and information technologies*, 2013.
- [10] Daniel Md Nor, Rosli Omar, M. Zazar M Jenu, Jean Marc Ogier, "Image Segmentation and Text extraction: Application to the Extraction of Textual Information in Scene images", *International seminar on application of science mathematics*, 2011.
- [11] Rodolfo P dos Santos, Gabriela S. Clemente, Tsang Ing Ren, George D. C. Calvalcanti, "Text line Segmentation based on Morphology and Histogram projection", *International Conference on document analysis and recognition*.
- [12] Neha Gupta, V. K. Bang, "Image Segmentation for Text Extraction", *International Conference on Electrical, Electronics and Civil Engineering*, 2012.
- [13] Nikolas G. Bourbakis, "Methodology for document processing: Separating Text from images", *Engineering Applications of Artificial Intelligence*, Elsevier.com., 2001
- [14] Sadagopan Srinivasan, Li Zhao, Lin Sun, Zhen Fang, Peng Li, "Performance Characterization and Acceleration of Optical character recognition and handheld platforms", *IEEE 2010*.
- [15] Yafang Xue, "Optical character recognition", *submitted to University of Michigan*.
- [16] Nadira Muda, Nik Ismail, Siti Azami Abu Bakar, "Optical character recognition by using Template Matching", *submitted to University of Malaysia Pahang*.
- [17] Yichang Shih, Donglai Wei, "Machine learning: Handwritten Sanskrit recognition using multi class SVM with k NN guidance".
- [18] S. K. Thilagavathy, Dr. R. Indra Gandhi, "Recognition of distorted character using edge detection algorithm", *International Journal of Innovative Research Computer and Communication Engineering*, 2013.
- [19] Sameeksha Barve, "Optical character recognition using Artificial Neural Network", *International Journal of Advanced Research in Computer Engineering and Technology*, 2012
- [20] Pardeep Kaur, Pooja Choudhary, Varsha Sahni, "English scanned document character recognition and matched and missed matched analysis using NN and MDA", *International Journal of Computer Applications*, 2015.
- [21] Sandeep Tiwari, Shivangi Mishra, Priyank Bhatia, Praveen Yadav, "Optical character recognition using MATLAB", *International Journal of Advanced Research in Electronics and Communication Engineering*, 2013.
- [22] Reetika Verma, Mrs. Rupinder Kaur, "Enhanced character recognition using SURF feature and neural network technique", *International Journal of Computer Science and Information Technologies*, 2014
- [23] Sandeep Saha, Nabarag Paul, Sayam Kumar, Sandip Kundu, "Optical character recognition using 40 point feature extraction and artificial neural network", *International Journal of Advanced Research in Computer Science and Software Engineering*, 2013.
- [24] Chaw Su Thu, Theingi Zin, "Implementation of Text to Speech conversion", *International Journal of Engineering rresearch and Technology*, 2014.

- [25] R. Shantha Selva Kumari, R. S. Sangeetha, "Conversion of English Text to Speech conversion using Indian speech signal", *Mathematical and computational methods in electrical engineering*.
- [26] Taapas Kumar Patra, Biplab Patra, Pushpanjali Mohapatra, "Text to Speech conversion with Phonematic Concatenation", *International Journal of Electronics and Computer Technology*, 2012.
- [27] G. D. Ramteke, R. J. Ramteke, "Text to Speech Synthesis of Marathi Numerals", *International Journal Engineering and Technical Research*, 2015.
- [28] R. Shantha Selva Kumari, R. Sangeetha, "Conversion of English Text to Speech using Indian Speech signal, *International Journal of Scientific Engineering and Technology*, 2015 .