

# Robust Assistive Reading Framework for Visually Challenged

**Avinash Verma**

Research Scholar, BBD University Lucknow, India  
Email: avinash.verma93@yahoo.com

**Dr. Deepak Kumar Singh**

Associate Professor, Integral University Lucknow, India  
Email: deepak.iita@gmail.com

Received: 20 April 2017; Accepted: 13 May 2017; Published: 08 October 2017

**Abstract**—The Main objective of this assistive framework is to communicate the textual Information in the image captured by the Visually Challenged person as Speech, So that the Visually Challenged person can acquire knowledge about the surrounding. This framework can help Visually Challenged person to read books, magazine, warnings, instructions and various displays as well by taking their image along with the surrounding. Then the Optical Character Recognition (OCR) extracts and recognizes the text in the image and generates the text file. This text file is further converted to Speech with the help of Text to Speech (TTS) Synthesis. The inherent problem with the previous approach was if the acquired image is affected with the issues of different lighting conditions, noise and issue of Skew and Blur, as the image is captured by Visually Challenged person. Then the overall accuracy of the system was at stake due to inefficient OCR leads to improper Speech output of TTS Synthesis. In this paper we have introduced two more processes that are deblurring using Blind Deconvolution method and Pre-processing operation to remove the effect of noise and blur. Thus it prepares the image for efficient result of the framework for Visually Challenged. The proposed approach is implemented in Matlab with the image captured manually and taken from the internet and the result along with the OCR text file and corresponding output Speech shows that our framework is better than the previous framework.

**Index Terms**—Assistive framework, Visually Challenged, Optical Character Recognition, Text to Speech and deblurring.

## I. INTRODUCTION

A lot of text information can be extracted from the images of the Surrounding Environment having text. The Surrounding having text information are name plates having the name and address of the individuals outside their residences, Boards containing the name of street, Area and distance, Various sign board and warning

boards and the diversion boards are embedded with text, we also see various displays at airports, railways and shopping malls and most of the text is found in the magazines, newspaper and books. So in our day to day life we are surrounded by text which we see from our eye and then we acquire the knowledge based on this textual information. For instance if we are searching for an address and we find a name plate of the same address outside a residence then we are sure that we have found the correct address. Hence textual information plays a vital role in our life. On the other hand if we think of a person which is Visually Challenged he cannot acquire the knowledge of the surrounding whether it be a warning sign, name of street and various types of displays because he cannot see these things. With the advancement of the Information Technology we have thought of the solution to this problem [1] can be with the help of Image Processing and Signal Processing using Optical Character Recognitions and Text to Speech conversion.

### A. Input Capturing by Visually Challenged

Image of the text can be taken either with the help of a scanner or with the help of a portable camera or smart phone camera. In today's era with the advent of technology we have high resolution portable camera with many features such as autofocus, wide angle and high picture clarity it is easy to capture text with clarity. If the text in the captured image is clear then its segmentation and recognition [2] will be efficient and overall accuracy of the system will be high. But for a Visually Challenged person to capture the image with the help of a portable camera and with clarity will be a herculean task. We have assumed that the image captured by the Visually Challenged will be affected with the issues of Blur [2] and skew [3]. Problem associated with the Scanner is that for Visually Challenged person, it is difficult to place a text document in a proper way as the document need to be placed in the scanner in proper orientation, for the scanner to scan efficiently. A lot of research is done on the image acquired by portable camera for efficient optical character recognition [4]. But the image acquired using a portable camera can be affected with problem of

Blur, Skew and variation in the lighting conditions. In this paper we are dealing with the problem of blurred textual image that too affected with a Uniform Motion Blur.

### B. Image Blurring and Deblurring Model

Image Degradation model [5][6] is modeled in which we apply degradation operator on the Input Image along with the additive noise which is random in nature to obtain a Degraded Image. The degraded image is  $g(x, y)$  is obtained by applying the degradation operation  $H$  over the input image  $f(x, y)$  along with the additive noise  $n(x, y)$ .

$$g(x, y) = H[f(x, y)] + n(x, y) \quad (1)$$

Blur in the image is a type of degradation. Blur is the most common degradation which will be associated with our system as the image will be captured while moving. So the Image will be mostly effected with the Uniform motion Blur. Blurred Image is commonly modeled as

Where  $g(x, y)$  is Blurred Image and  $f(x, y)$  is the actual input or the image without blur  $n(x, y)$  is noise  $h(x, y)$  is the Blur Kernel also known as point spread function (PSF) [6] and  $*$  is the convolution operator.

$$g(x, y) = f(x, y) * h(x, y) + n(x, y) \quad (2)$$

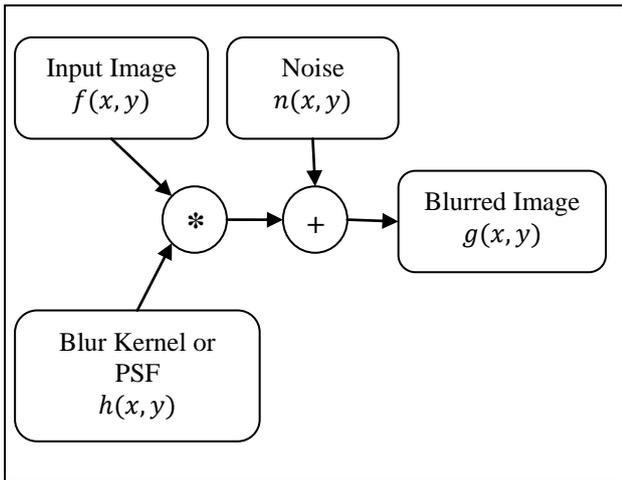


Fig.1. Image Blurring Model

Uniform Motion Blur [21][22] is caused due to the motion of either the object or the capturing device at the time of image capture. The Blur Kernel or the Point spread Function  $h(x, y)$  for the motion Blur is based on two parameters that is the Length of motion Blur ( $L$ ) and the angle of motion Blur ( $\theta$ ). When the object with text information is to be captured translates with a relative velocity  $V$  in respect to the camera, the blur length  $L$  in pixels is  $L = VT_{\text{exp}}$  where,  $T_{\text{exp}}$  is the time duration of the exposure. The expression for motion blur is given as,

$$h(x, y) = \begin{cases} \frac{1}{L}, & \text{if } 0 \leq |x| \leq L \cos \theta ; y = L \sin \theta \\ 0, & \text{otherwise} \end{cases}$$

When the angle of blur  $\theta = 0$ , it is called horizontal motion blur. Point spread function can be represented in discrete as,

$$h(m, n, L) = \begin{cases} \frac{1}{L}, & \text{if } m = 0, |n| \leq \left\lfloor \frac{L-1}{2} \right\rfloor \\ \frac{1}{2L} \left\{ (L-1) - 2 \left\lfloor \frac{L-1}{2} \right\rfloor \right\}, & \text{if } m = 0, \\ & |n| \leq \left\lfloor \frac{L-1}{2} \right\rfloor \\ 0, & \text{elsewhere} \end{cases}$$

Deblurring [23][24][25] is the inverse operation of Blurring process which comes under Image Restoration [6]. Deblurring model can be given as the Deconvolution operation  $**$  of the degraded image  $g(x, y)$  and the estimated Point Spread function or Blur Kernel  $h(x, y)$  to obtain the Input image  $f(x, y)$ . It is practically impossible to get the exact  $f(x, y)$  but we will get some degraded version of input Image  $f'(x, y)$  due to the random noise present in the image which gets amplified will Deconvolution operation. So we call the obtained  $f'(x, y)$  as Latent Image.

$$f'(x, y) = g(x, y) ** h(x, y) \quad (3)$$

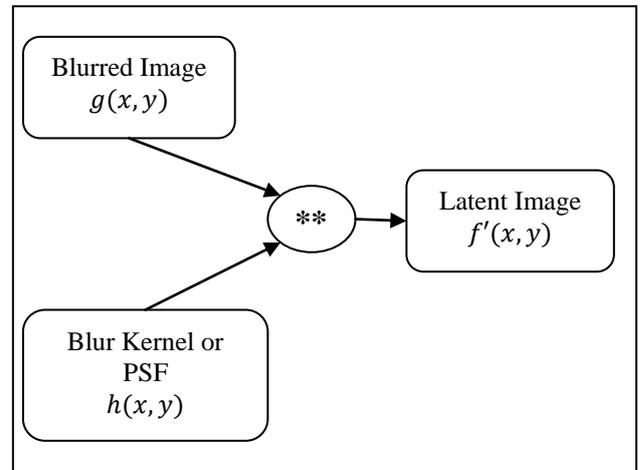


Fig.2. Image Deblurring Model

### C. Optical Character Recognition (OCR) Algorithms

Most of the Reading framework for Visually Challenged till date have used work used Optical Character Recognition [4] in or the other way. OCR is image processing technology in which we extract the text present in the image and then output it as a Text file. Input to the Optical Character Recognition module is the acquired Image ideally it should be an uncompressed Bitmap Image. With the advent in the technology knower days we can extract the text from even compressed Images. On the input image we have to perform various pre-processing operations such as Binarization, Blur removal, skew correction and noise removal step [2] in order to increase the accuracy of the OCR and to prepare the image for further processing. Then text and non-text segmentation [7][8] is performed to isolate text information from graphics, segmented text is then segmentation into various lines and is known as Line Segmentation. From each line we segment different words and then these words are segmented as individual characters known as Character level Segmentation. Character recognition of the individual character is performed and output will be the Text file containing the recognized text. The main problem associated with OCR Engine is when the input image to the OCR is affected with any kind of Blur. Blur in the image is caused due to the movement of the capturing device or the object to be captured at the time of Image Capture. This problem was not addressed by the previous Reading Framework for Visually challenged persons. We have taken this problem into consideration and have come up with one of the efficient cost effective technique to remove the Blur in the image using Blind Deconvolution Algorithm in our previous paper Text Deblurring Using OCR performance [9]. Thus we remove the Blur present in the image and then it send for Optical Character Recognition. Thus the recognition rate of the text increases and the overall reliability of the System increase drastically.

#### D. Review of Text to Speech Engine

After the Optical character recognitions of the Image we get the text file as output. Now text has to be converted to speech or Braille So that the visually challenged person can understand it. Text to Speech Engine [10][12] converts the text output from OCR engine to its corresponding speech output. Text to speech Engine initially works by performing the preprocessing operation [11] on the Input text file. This is done in order to increase the efficiency of the text to speech generation. After this step speech generation is performed for the preprocessed text. The preprocessing step prepares the input text file for further processing the operations like text analysis, text normalization and then translating into a phonetic or some other linguistic representation are performed. During Preprocessing the spell checking is performed this helps in correcting some misrecognized text during the Optical Character Recognition on the text based on the punctuation marks the formatting of the paragraph is done. The abbreviations and acronyms are handled in the text Normalization [11] which enhances the speech output. This helps in communicating the

meaningful speech to the visually challenged person. Morphological operations for the proper pronunciation of the word are performed in the Linguistic analysis followed by syntactic analysis to facilitate in handling ambiguities in the Input text. Now the speech generation process involves various steps such as phonetic analysis that is useful in finding the phone level within the word. Each phone level has the information about the sound tagged with it to be produced. Grapheme to phoneme conversion is the next step based on the dictionary. This is followed by the prosodic analysis which attaches the pitch and the duration information for the speech conversion. Speech synthesis is the last step which involves voice rendering to get the speech from the text to speech Engine.

## II. RELATED WORK

Reading framework for Visually Challenged has been one of the researched topics under the assistive technology for visually challenge [13]. From the early 90's there have been attempt to create some assistive technology for visually challenged when Xerox launched the device called Reading Edge, which used to scan the printed materials and then it use to read out loud to its users. It also provides is user with the Braille interface so that the blind persons can read out the contents using Braille interface. Reading Edge device has a scanner, speech generation software and a Braille interface equipped with keyboard for editing. Users were also given the facilities of adjusting the reading speed and have the option of choosing among different speaking voices. This was a handy aid for the visually challenged at that time but its usage required significant effort. The reading materials especially the books or the page has to be placed in the proper orientation for it to be scanned. Also the unit consisted of scanner which was large in size and was weighted. So the unit cannot be carried freely. R-Map [14] android application proposed it uses the camera to capture the Image of the Text and then with the help of Optical character Recognition and the text to speech engine it provides the read out loud service. As the mobile phone was used that has lower processing power than a desktop or Notebook. Mobile camera image is affected with various issues such as skew, blur, curved base lines and auto focus mechanism etc. there by making the best possible OCR engine to fail. The most powerful open source Tesseract OCR [20] engine was used for the recognition of the text in the captured image. After that recognized text file is send to Text To Speech (TTS) engine for further text to speech synthesis. With limited mobile screen size it was hard for a visually challenge person to capture the image of the long printed material with accuracy. OCR is developed for scanned documents of high quality to perform text extraction on a low resolution mobile camera captured image it is a challenging job. Whereas the issues of skew blur, different lighting condition and complex background also make the task difficult.

Assistive Reading System for Visually Impaired [15] proposed by Akshay Sharma, Abhishek Srivastava, Adhar Vashishth that uses a Document scanner to scan the image of document text to act as Input to the Optical Character Recognition module, which performs text and Non-text segmentation and recognition of segmented text to generates a text file as output which is converted to its corresponding speech with the help of Text to Speech Module. The difficulty with this proposed system is it also uses a scanner which is not portable and requires accuracy of visually challenged person to put the document into the scanner in proper orientation. Portable Camera-Based Assistive Text and Product Label reading for hand held Objects for Blind persons [16], a framework to help blind person to read the text label and product packaging from hand held object in their day to day life proposed by Chucai Yi, Yingli Tian and Aries Arditi. This system make blind person to feel independent in there day to day life. It isolates the object from its surrounding with the help of motion based Region of Interest (ROI) [2] by asking the user to shake the object. Thus it extracts the moving object from its complex background and then text extraction is performed on the segmented object and this followed by text recognition. The recognized text as speech is communicated to the blind using text to speech mechanism. Hence blind person can get the essence of the object details based on the specification provided on the object as speech. Patrick E.Lanigan, Aaron M. Paulos, Andrew W. Williams, Dan Rossi and Priya Narasimhan proposed Trinetra [17] a cost-effective assistive technology developed for visually challenged person to allow them independent in their life. With the help of this system they can easily do their daily activities. Trinetra system aims at quality improvement of the life of visually challenged by harnessing the collective capability of diverse networked embedded devices to help them to support navigation, shopping, transportation. Trinetra uses a barcode-based solution comprising a combination of off-the-shelf components, such as an Internet and Bluetooth-enabled mobile phone, text-to-speech software and a portable barcode reader. This is a bar code based solution to the problem of visually Challenged person. It have been seen that most of the system proposed were having their own set of merit and demerits. With the advancement in the information technology in recent times and the availability of considerable information processing capability and miniaturized sensors this similar unit can be designed which will have much smaller form factor, for e.g. like a mobile or special goggles. Instead of scanning the printed material it will take a picture and then use it to convert to text and then to speech [18][19].

### III. PROPOSED APPROACH

In the proposed approach input to the system is the image captured by the visually challenged person containing the textual information. Our aim is to communicate the text information in the image to the

Visually Challenged person as Speech, so that he may acquire the knowledge about its surrounding. In order to fulfill our aim we have prepared an overall architecture of robust Reading framework for visually challenged person. In this framework on the input image  $f(x,y)$  the deblurring step is performed. This step is responsible for the blur kernel or the point spread function estimation in the image and then performing the Deblurring operation on the input image  $f(x,y)$  to get the deblurred Image  $f'(x,y)$ . Deblurring is performed using Blind Deconvolution method using OCR performance [9]. After deblurring we get the deblurred image  $f'(x,y)$  then we perform the preprocessing steps in order to prepare the image  $f'(x,y)$  for better Optical Character Recognition the steps involved are noise removal, thresholding and then Binarization operations to obtain the perfect binary black and white image  $f''(x,y)$  which can be sent to the Optical Character Recognition for the extraction of the text information in the image  $f''(x,y)$  and convert the text information into corresponding Output text file. The last and the important step is conversion of Output text file to speech or voice output with the help of Text to Speech. Hence the output speech will be send to visually challenged person so that he can get the knowledge of the text. This will help him to be more independent and dream of providing some level of Independence can be fulfilled through this framework. The Output Speech is feed to the Visually Challenged person so that he/she could listen whatever text is there in the image.

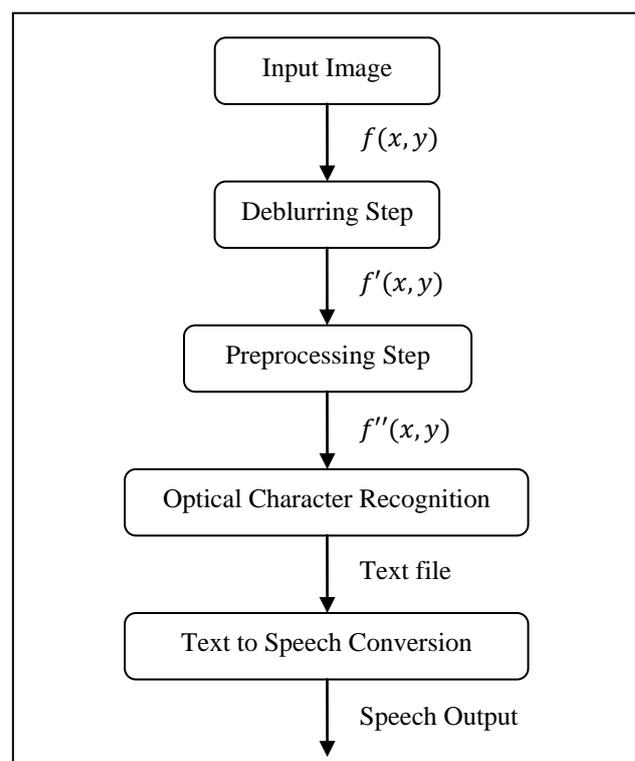


Fig.3. Overall Architecture of Robust Reading Framework for Visually Challenged

### Algorithm for the proposed work

- i. Input to the System is Image Captured by Visually Challenged person  $f(x, y)$ .
- ii. Blur Removal is the most important operation in which we first Estimate the Blur Kernel  $H(x, y)$ . Blur Kernel is estimated with the help of two parameters that are Blur Length  $L$  and Blur Angle  $\theta$ .
- iii. We Iteratively Create the Blur Kernel  $H(x, y)$  for different values of Blur Length  $L$  and Blur Angle  $\theta$ . Then using Deconvolution operation we extract the Latent Image  $f'(x, y)$  for the Input Image  $f(x, y)$  and the Current computed Blur Kernel  $H(x, y)$ .
- iv. In Order to find the best Blur Kernel  $H(x, y)$  which can help in the text information extraction we evaluate each resultant Latent Image  $f'(x, y)$  by calculating the Average Word Confidence  $AWC$  with the help of OCR.
- v. Highest value of  $AWC_{max}$  indicates that Latent Image  $f'(x, y)$  is best recognized with the corresponding Blur Kernel  $H(x, y)$  with two parameters that are Blur Length  $L_{max}$  and Blur Angle  $\theta_{max}$ .
- vi. Thus we De-blur using Blind Deconvolution, the Input Image  $f(x, y)$  with the Blur Kernel  $H(x, y)$  with  $AWC_{max}$  for Blur Length  $L_{max}$  and Blur Angle  $\theta_{max}$  to get the Deblurred Image or the Actual Latent Image  $f'(x, y)$  that can be used for further processing.
- vii. The Preprocessing Operations are applied to the Deblurred Image or the Actual Latent Image  $f'(x, y)$  to obtain the Image  $f''(x, y)$  which is used for further processing. Preprocessing Operations increase the efficiency of the overall system.
- viii. The Image  $f''(x, y)$  thus obtained is feed to an Optical Character Recognition (OCR) Engine for Text recognition. Output of OCR engine is the text file containing the text in the Image.
- ix. Text file Output is send to a Text to Speech (TTS) Engine which performs the preprocessing operation the text file and then converts it to the corresponding Speech output.

#### A. Deblurring step on input image

Input image to the system  $f(x, y)$  is image captured by the visually challenged person with the motive to acquire the text information present in the image as speech, so that they can act accordingly. As already discussed that the image if affected by motion blur that is caused by the movement of the capturing device or the object at the time of image capture then this system accuracy will be at stake as the OCR is going to fail miserably and the next step that is TTS engine will have nothing as input for speech generation. To overcome this problem Deblurring step is being implemented. Average Word

Confidence  $AWC$  metrics [9] is used for the deblurring process.  $AWC$  Value lies between 0 and 1 it is calculated as

$$AWC = \frac{\sum \text{Individual Word Confidences}}{\text{Total number of words}}$$

In this Deblurring step we iteratively based on two parameters of motion blur that are Blur Length  $L$  and Blur Angle  $\theta$  we try to estimate the optimal PSF that will give the best recognition rate based on  $AWC$  value of the OCR.  $AWC$  is the mathematical average of individual word confidences. Word Confidence [9] is the Normalized sum of character level confidence. Character Confidence is the normalized measure of the how effectively the character is recognized. Higher the Character Confidence of recognition Higher is the Word Confidence of Recognition and vice-versa. The Word Confidence is also affected by the dictionary based verification. If a word is found in the dictionary, it increases the Word Confidence value of that word. The longer the word, the higher will be the confidence value if it is found in the dictionary. For example if a long word of around 15 characters is found in dictionary it is pretty sure that the word is correct and will yield a higher word confidence, while on wrongly detected character a match against the dictionary by mistake is unlikely to occur. Short words like 'add' or 'odd' will both be found in dictionary. Therefore for smaller words there is a probability that we can get the dictionary match. Hence to overcome this problem words with 2 or less characters are not checked against the dictionary. The word confidence is normalized to an interval of 0.00 to 1.00 where 1.00 is the best and 0.00 is the worst word confidence.

For different value of Blur Length  $L (L_1 L_2 L_3 L_4 \dots L_n)$  and different value of Blur angle  $\theta (\theta_1 \theta_2 \theta_3 \theta_4 \dots \theta_n)$  we make the PSF  $H(x, y)$  and then for every PSF we perform the deconvolution operation which is inverse operation of the blur to get the latent Image  $f'(x, y)$  and the using OCR we calculate the  $AWC$  value for the latent Image. The value of Blur length, Blur Angle and  $AWC$  are tabulated. The highest value of  $AWC_{max}$  is easily identified from the table and the corresponding values of the Blur length  $L_{max}$  and Blur Angle  $\theta_{max}$  are found. These values constitute the best PSF  $H(x, y)$  which can deblur the given input Image  $f(x, y)$  to get the maximum recognition rate of the text information in the Image. After that we perform the deconvolution operation on the Input Image  $f(x, y)$  with the obtained value of Blur kernel or the PSF  $H(x, y)$  to obtain the latent Image  $f'(x, y)$  which is given as

$$f'(x, y) = f(x, y) ** H(x, y)$$

Where  $**$  is the deconvolution operation which helps in deblurring and is the inverse of Blurring operation. The resultant Image as known as Latent image  $f'(x, y)$  as it will be a distorted version of actual after deblurring step as the noise in the image get amplified during this

process. Latent image  $f'(x, y)$  will be used for the further processing. Fig 4 shows Input Image  $f(x, y)$  to this system is Blurred Image. Now we estimate the PSF for which we will get the recognition of the text. Hence for every estimated PSF we perform the deblurring operation and calculate  $AWC$  value for them. The Highest value indicates that the PSF used will give the best recognition result for the Blurred Input Image. Fig 5 shows the Deblurred Image  $f'(x, y)$  with PSF created with Blur Length  $L = 15$  pixel and Blur Angle  $\theta = 19$  degree and the yellow boxes with values in the image indicate the word confidences of the words. They help in the calculation of  $AWC$ . Similarly Fig 5 shows the Deblurred Input Image  $f'(x, y)$  with PSF obtained for Blur Length  $L = 18$  pixel, Blur Angle  $\theta = 24$  degree  $AWC = 0.586364$  and Fig 6 shows the Deblurred Input Image  $f'(x, y)$  with PSF obtained for Blur Length  $L = 17$  pixel, Blur Angle  $\theta = 19$  degree  $AWC = 0.646726$ . For different values of Blur Length  $L$  and Blur Angle  $\theta$  the highest value of  $AWC_{max} = 0.646726$ . Hence the corresponding values of the Blur length  $L_{max} = 17$  pixels and Blur Angle  $\theta_{max} = 19$  degree are found. The Latent image thus obtained  $f'(x, y)$  thus obtained in shown in Fig 7 that is used for further processing in next steps.



Fig.4. Blurred Input Image  $f(x, y)$

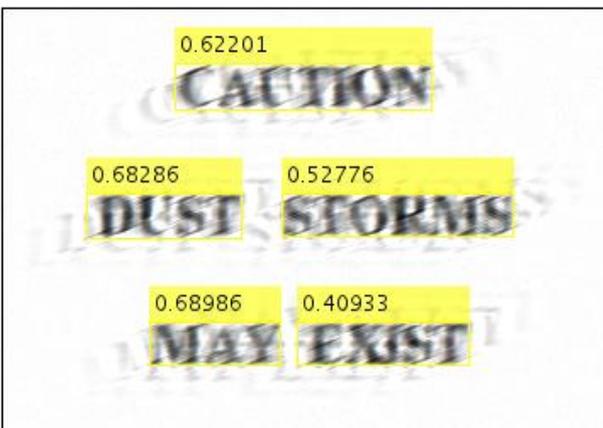


Fig.5. Deblurred Input Image  $f'(x, y)$  with PSF obtained for Blur Length  $L = 18$  pixel, Blur Angle  $\theta = 24$  degree  $AWC = 0.586364$

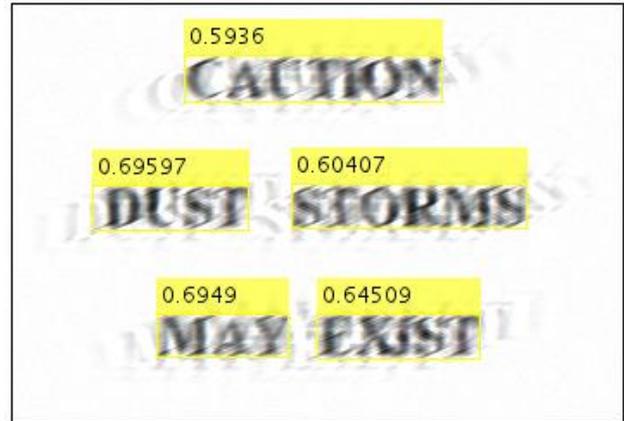


Fig.6. Deblurred Input Image  $f'(x, y)$  with PSF obtained for Blur Length  $L = 17$  pixel, Blur Angle  $\theta = 19$  degree  $AWC = 0.646726$

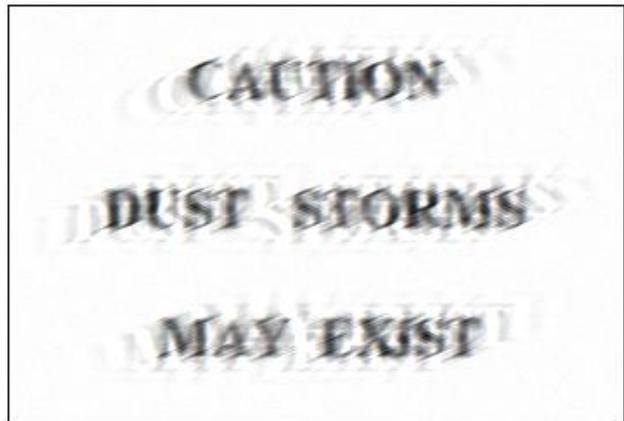


Fig.7. Latent Image  $f'(x, y)$  with Optimal PSF obtained for Blur Length  $L = 17$  pixel, Blur Angle  $\theta = 19$  degree  $AWC = 0.646726$

**B. Preprocessing step on resultant image**

Resultant Image  $f'(x, y)$  obtained from the Deblurring step is a mostly degraded image. As we know that during the Deblurring process to obtain the Latent Image to the extent that its text information can be extracted with higher degree of Word confidence the deconvolution operation with the estimated Blur Kernel  $H(x, y)$  is performed. During this process the noise gets amplified in order to increase the efficiency of the system we need the preprocessing steps. On the Image  $f'(x, y)$  the thresholding operation [2] is applied in order to increase the speed of processing. After thresholding step next operation is the binarization. Binarization [2] operation converts the image into perfect black and white pixels. In the image each and every pixel has either two values that are 0 or 1. Now the resultant Image of the previous stage  $f'(x, y)$  now gets converted to binary image having only black and white pixels. This binary image thus obtained  $f''(x, y)$  is the input to the next stage OCR engine which will give the best possible recognition result.

**C. Optical character recognition**

After the preprocessing steps the Resultant Image  $f''(x, y)$  is feed to the OCR engine. OCR engine used is Tesseract [20]. It is an open source OCR engine developed by HP between 1984 and 1994. Tesseract [20] works on improving the rejection efficiency than on base-level accuracy. The input image to the Tesseract is a binary Image  $f''(x, y)$  obtained in the previous preprocessing step. Connected component analysis is performed and the resultant Image  $f''(x, y)$  outlines of the components are stored. This step had the advantage of being simple to detect the inverse text and to recognize it as easily as black and white text. At this stage the outlines are gathered together by nesting the blobs. Blobs are organized into text lines, and the lines and regions are analyzed for fixed pitch or proportional text. Text lines need to be segmented into words this is done according to the type of character spacing. Fixed pitch text is chopped immediately by character cells. Proportional text is broken into words using definite spaces and fuzzy spaces. After the segmentation process next step is Recognition. Recognition process then proceeds as a two-pass process. In the first pass, each word is recognized. Each satisfactory recognized word is passed to an adaptive classifier as training data in order to learn the words for further recognition. The adaptive classifier then gets a chance to more accurately recognize text lower down the page. Since the adaptive classifier may have learned something useful too late to make a contribution near the top of the page. A Second pass is run over the page, in which words that were not recognized well enough are recognized again. This increase the accuracy of the recognition as two passes of recognition are run on the same input text.

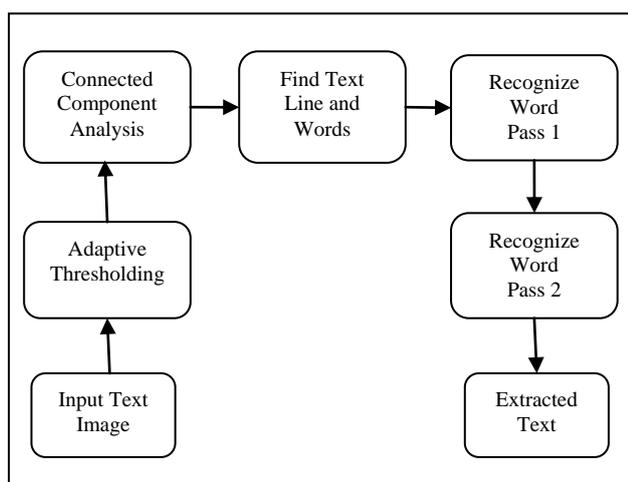


Fig.8. Tesseract OCR Architecture

normalization is also performed in order to handle the abbreviations and acronyms to enhance the speech output. Various Morphological operations of text to speech are performed for proper pronunciation of word in Linguistic and syntactic analysis. After preprocessing step the next step is speech generation which involves the phonetic analysis to find the phone level present in the words. Different phone levels have the information about the sound tagged with it to be produced. Graphemes are then converted to phonemes based on the dictionary. The Prosodic analysis is one of the important steps of speech generation in this step pitch and the duration information is attached for speech conversion. After this speech synthesis is done which involves voice rendering to get the speech for the text input. In addition to that there are various options available for visually challenged person which can help him in order to understand the communicated speech output. In the programming implementation of the text to speech function named TTOS is created which synthesizes the text input. The audio format is mono, 16 bit, 16k Hz by default which can be varied based on the requirement. We can also select the voice available in the list based on our preference by default first voice is set form the List. We also can set the pace of the speech output that is -10 as the slowest and +10 as the fastest. In our case we have used the default value that is 0. We also have to set the sampling rate of speech in Kilo Hertz that are 8000, 11025, 12000, 16000, 22050, 24000, 32000, 44100 and 48000. We have used the default value 1600. The function also requires the Microsoft Win32 Speech API (SAPI) for the voice output.

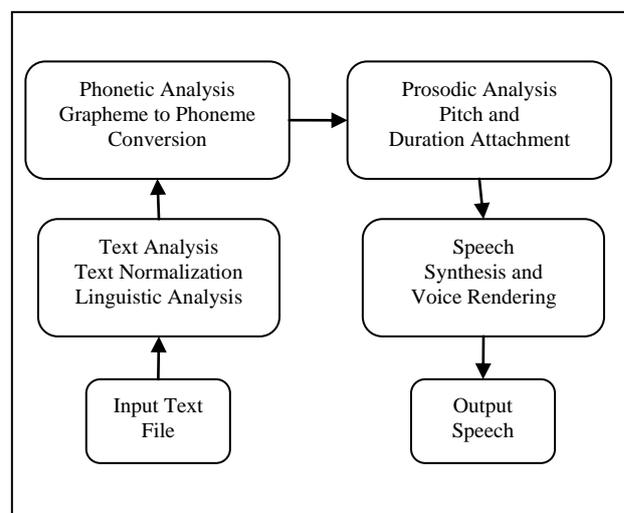


Fig.9. Overall Architecture of Text to Speech Engine

#### D. Text to speech conversion

Optical Character Recognition extracts the text information present in the input image  $f(x, y)$  in the form of text file. In order to communicate it to the Visually Challenged person this text file need to be convert to speech. The text file is initially preprocessed for the spelling check as during recognition process few words can be misrecognized they are corrected. Text

## IV. RESULT ANALYSIS

In the proposed work robust assistive reading framework for visually challenged on the input image  $f(x, y)$  to the framework in Fig.4 the deblurring step is applied. Deblurring step is responsible for the removal of the blur present in the input image  $f(x, y)$ . The

Deblurring step iteratively estimates the PSF of the motion blur based on the Average Word Confidence  $AWC$ . In the conventional frameworks if the input image is affected with blur then there will be no recognition of text in the OCR this will lead to no speech output of the TTS, these were the inherent problems with the conventional framework that led to the failure of conventional frameworks. In Fig. 5 and Fig. 6 the deblurred images with the values of blur length  $L$ , blur angle  $\theta$  and Average Word Confidences  $AWC$  are shown. The highest value of the  $AWC=0.646726$  for the blur length = 17 pixels and blur angle = 19 degree is obtained for all the iterative values of blur length and angle in the range. Thus the Latent image  $f'(x,y)$  obtained after deblurring with  $AWC=0.646726$  is shown in the Fig. 7. The obtained Latent image  $f'(x,y)$  is the restored image for which highest recognition rate can be obtained. As it is known that during the deblurring process the noise present in the input image gets amplified. On the obtained Latent image  $f'(x,y)$  the preprocessing steps are applied that are thresholding and binarization for the removal of amplified noise to obtain the image  $f''(x,y)$ . Image  $f''(x,y)$  is sent to the Tesseract OCR engine which recognizes the text present in the image and generates an output text file. Then this text file is converted to speech output after the series of steps applied to the text file. For the speech conversion various attributes can be set based on the preference of the visually challenged persons otherwise the default attributes of speech output are used for the speech API. Hence the speech output is communicated to the visually challenged person so that he can get the knowledge about its surrounding. This framework will also increase the efficiency and reliability of the system as the blurred image affected with motion blur can be handled which was not possible with previous conventional framework.

## V. CONCLUSION AND FUTURE WORK

Robust reading framework for visually challenged person can help them in getting the knowledge of the surrounding in the form of speech. In this framework we have overcome the inherent problem associate with all the previous frameworks that was issue of blur in the Input image captured by the visually challenged person. We have initially checked the input image  $f(x,y)$  to the system for the motion blur. If the blur is present in the image we have iteratively based on two parameters that are Blur length  $l$  and Blur angle  $\theta$  created the PSF that is the Blur kernel. For every PSF in the range we have found the optimal PSF that will give the best Deblur image for further processing based on the Average Word Confidence  $AWC$  metrics of the OCR. This Deblurring step was not present in all the previous systems. Without this step the overall framework was of no use if the image had any kind of Blur. As blurred image will give incorrect recognition of text and the incorrect text will be communicated as voice which will be of no use for visually challenged person. In the future work we want to propose new Deblurring algorithms which can be used

for this framework based on Computational Neural Networks. We have only worked on the reading aspect for the visually challenged person this can also be extended for the Vision and Navigation of visually challenged person. Any type of assistance to visually challenged person can help them to be independent in their life.

## ACKNOWLEDGMENT

I would like to express my gratitude towards almighty God the mother father of this universe for giving me strength and positive thinking to work on this research topic. I would like extend my gratitude towards my PhD guide Dr. Deepak Kumar Singh for encouraging me to write this paper. He has motivated me whenever I was down and has helped me in writing this paper. With his deep knowledge of the subject he has time and again helped and given his valuable suggestion which have helped to improve my work. I would also like to thank my parents and family for being my strength.

## REFERENCES

- [1] Marion A.Hersh, Michael A.Johnson "Assistive technology for Visually Impaired and Blind people" Book by Springer ISBN 978-1-84628-866.
- [2] C. Gonzalez Richard E. Woods "Digital Image Processing" Book Third Edition Rafael Interactive Pearson International Edition prepared by Pearson Education PEARSON Prentice Hall.
- [3] Jian Liang, Daniel DeManthou "Geometric rectification of Camera-Captured Document Image" IEEE Transaction on Pattern Analysis and Machine Intelligence Vol 30 No 4 April 2008.
- [4] U. Pal, B.B. Chaudhuri "Indian script character recognition: a survey" Pattern Recognition 37 (2004) 1887 – 1899 published by ELSEVIER
- [5] J.Amudha, N.Pradeepa, R.Sudhakar "A Survey on Digital Image Restoration" ELSEVIER Procedia Engineering, Volume 38, 2012, Pages 2378-2382.
- [6] Ratnakar Dash "Parameter Estimation for Image Restoration" PhD thesis NIT Rourkela, Orrisa,India March 2012.
- [7] Rainer Lienhart, Member, IEEE, and Axel Wernicke "Localizing and Segmenting Text in Images and Videos" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 12, NO. 4, APRIL 2002.
- [8] Honggang Zhang,KailiZhao,Yi-ZheSong,Jun Guo " A Text extraction from natural scene image:A survey" Neurocomputing 122 (2013) ELSEVIER.
- [9] Avinash Verma, Deepak Kumar Singh " Text Deblurring Using OCR Word Confidence", International Journal of Image, Graphics and Signal Processing (IJIGSP), Vol.9, No.1, pp. 33-40, 2017, DOI :10.5815/ijigsp.2017.01.05.
- [10] Kaveri Kamble, Ramesh Kagalkar "A Review: Translation of Text to Speech Conversion for Hindi Language" International Journal of Science and Research (IJSR) Volume 3 Issue 11, November 2014
- [11] Suhas R. Mache,Manasi R. Baheti,C. Namrata Mahender "Review on Text-To-Speech Synthesizer" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 8, August 2015.
- [12] Itunuoluwa Isewon,Jelili Oyelade,Olufunke Oladipupo

- "Design and Implementation of Text To Speech Conversion for Visually Impaired People" International Journal of Applied Information Systems Foundation of Computer Science FCS, New York, USA Volume 7– No. 2, April 2014
- [13] Avinash Verma, Deepak Kumar Singh, Nitesh Kumar Singh " Review on Assistive Reading framework for Visually Challenged" accepted for publication by International Conference on Computational Intelligence in Data Mining (ICCIDM-2016) in the AISC Series, Springer.
- [14] Akbar S. Shaik, G Hossain "Design and Development and performance Evaluation of Reconfigured Mobile Android Phone for People Who are blind or Visually Impaired" In Proceedings of the 28th ACM International Conference on Design of Communication, New York, NY, USA, 2010. ACM. October 2010.
- [15] Akshay Sharma, Abhishek Srivastava, Adhar Vashishth " An Assistive Reading System for Visually Impaired using OCR and TTS" International Journal of Computer Applications, Vol 95, No. 2 June 2014.
- [16] Chucai Yi, Yingli Tian and Aries Arditi "Portable Camera-Based Assistive Text and Product Label Reading from Hand Held Objects for Blind Persons" IEEE/ASME Transaction on Mechatronics.
- [17] Patrick E.Lanigan, Aaron M. Paulos, Andrew W. Williams, Dan Rossi and Priya Narasimhan "Trinetra: Assistive Technology for Grocery Shopping for the Blind" Published by Cylab in 2006.
- [18] Bindu Philips and R.D. Sudhaker Samuel " Human Machine Interface – A smart OCR for the visually challenged", International Journal of Recent Trends in Engineering, Vol 2, No. 3 November 2009.
- [19] Roberto Neto, Nuno Fonseca "Camera Reading For Blind People" HCIST 2014 International Conference on Health and Social Care Information Systems and Technologies, Procedia Technology 16 ( 2014 ) 1200 – 1209 published by ELSEVIER.
- [20] Ray Smith Google Inc. theraysmith@gmail.com "An Overview of the Tesseract OCR Engine" Google Inc OSCON 2007.
- [21] Shamik Tiwari, V. P. Shukla, and A. K. Singh "Review of Motion Blur Estimation Techniques" Journal of Image and Graphics Vol. 1, No. 4, December 2013.
- [22] Kishore R. Bhagat, Puran Gour "Novel Approach to Estimate Motion Blur Kernel Parameters and Comparative Study of Restoration Techniques" International Journal of Computer Applications (0975 – 8887) Volume 72– No.17, June 2013.
- [23] Lu Fang, Haifeng Liu, Feng Wu, Xiaoyan Sun, Houqiang Li "Separable Kernel for Image Deblurring" 2014 IEEE Conference on Computer Vision and Pattern Recognition Pages: 2885 - 2892, DOI: 10.1109/CVPR.2014.369.
- [24] Jing Wang, Ke Lu, Qian Wang, and Jie Jia "Kernel Optimization for Blind Motion Deblurring with Image Edge Prior" Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2012, Article ID 639824, 10 pages doi:10.1155/2012/639824.
- [25] Long Mai, Feng Liu "Kernel fusion for better image deblurring" 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Pages: 371 - 380, DOI: 10.1109/CVPR.2015.7298634.
- [26] Taeg Sang Cho, Sylvain Paris, Berthold K. P. Horn, William T. Freeman "Blur Kernel Estimation using the Radon Transform" Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on June 2011.
- [27] Zohair Al-Ameen, Ghazali Bin Sulong, Md. Gapar Md. Johar "Computer Forensics and Image Deblurring: An Inclusive Investigation" IJMECS Vol.5, No. 11, November 2013 PP.42-48, DOI: 10.5815/ijmeecs.2013.11.06.

#### Authors' Profiles



**Avinash Verma** born in India 1988 has Done B.tech in Computer Science and Engineering, M.Tech in Information Technology from Centre for Development of Advanced Computing (CDAC) Noida. Qualifying GATE 2010 and 2016 and currently perusing PhD in Computer Science and Engineering. His area of interest is Image processing and Optical character Recognitions. He has published and presented papers in National and International Journals and conferences.



**Dr. Deepak Kumar Singh** from India did his PhD from IIT Kanpur. He is currently Associate Professor at Integral University Lucknow India; his Area of Interest includes Artificial Intelligence, Expert system. He has guided many M. Tech theses and currently various PhD scholars are enrolled under his able guidance. He has Published and presented Papers in various national, International journals and conferences.

**How to cite this paper:** Avinash Verma, Deepak Kumar Singh, " Robust Assistive Reading Framework for Visually Challenged", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.9, No.10, pp. 29-37, 2017.DOI: 10.5815/ijigsp.2017.10.04