Acknowledgment in Pictures and Switching Perceived Text Over Completely To Discourse

[1]Amit Kumar, [1]Ashish Patel
[1]Dept. of Computer Science & Engineering RDEC, Ghaziabad
Email- amit.kumar@gmail.com

Abstract — Themaingoalofthisresearch is to foster a program that can recognize text characters from transform any normal picture into a voice signal. The program need to do the indistinguishable activity for any transferred picture and PDF record. The application ought to likewise havetoolsforpacemodulation, voicechoosingoptions,and capacity ability for picture to message yield. The interest group for this program can be extended to incorporate individuals with extraordinary requirements who likewise have learning disabilities, small kids, and a few other cultural gatherings. The text is removed fromthe picture utilizing optical person acknowledgment (OCR), and the Windows Programming interface is used to transform the text into discourse. The programming language for advanced picture handling isMATLAB.

Keywords-- Digital image processing, optical character recognition, speech modulation, MSER Regions, stroke width algorithm, and image character recognition

1. INTRODUCTION

A famous area of PC innovation is picture to-discourse transformation. It lays out a urgent calculate how we draw in with the framework and connection points on numerous stages. It has for quite some time been an objective to repeat human capacities like perusing throughmachines. Machine perusing, notwithstanding, has created from a pipedream toarealityduringthe past 50years. Themost successful type of human correspondence is in all probability discourse. Oneofthemostpopularuses oftechnologiesinthe fields of example acknowledgment and man-made consciousness is optical person acknowledgment.

The device helps with changing over printed data that is implanted in a picture or scene into discourse. This isn't the main use it could be put to. It is helpful to take text from PDF documents and transform it into discourse. The gathered text can be all put away as a text document in any area on the PC. While the text is being perused so anyone might hear, it likewise offers the choice to look into equivalents for words. Various speeds perhaps agreeable for clients to understand the language. Thus, a provision is added that takes into consideration discourse beat balance. Furthermore, clients can choose from different male and female speakers' voices as well as accents.

OCR, or optical person acknowledgment, is a procedure we use to remove text from photos. Afterthat, a text-to-discourse (TTS) module transforms the text into sound. We can seethatthisprocedurewassplitintotwomodules.Thefirst is picture acknowledgment, and the second is discourse change for thatimage optical person acknowledgment Optic person acknowledgment is alluded to as OCR. Through this methodology, the application will actually want to perceive a person naturally utilizing an optical method.OCR is the transformation of caught photos of printed or typewritten text into carefully variable data. Discourse union: Without straightforwardly utilizing a human voice, discourse blend makes discourse that is more human-like than mechanical. A voice synthesizer is, all the more as a general rule, kind of innovation that makes counterfeit discourse through the formation of images and signals. Altering the discourse's cadence has been conceivable. Furthermore, the application incorporates different voices and accents.
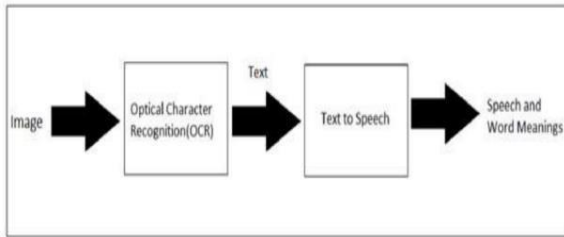
The above can be depicted by the following illustration:

Figure 1 An overview of the system

I. FRAMEWORK ANDARCHITECTURE

*2.1) HIGH LEVEL DESIGN:*THE HIGH LEVEL OPERATION OF THE SYSTEM IS DEPICTED IN THE DIAGRAMBELOW:

The graphic demonstrates that the system is made up of two primary modules:

The system's two primary components are depicted in the figure as:

(a)   image-to-text conversion and PDFgeneration
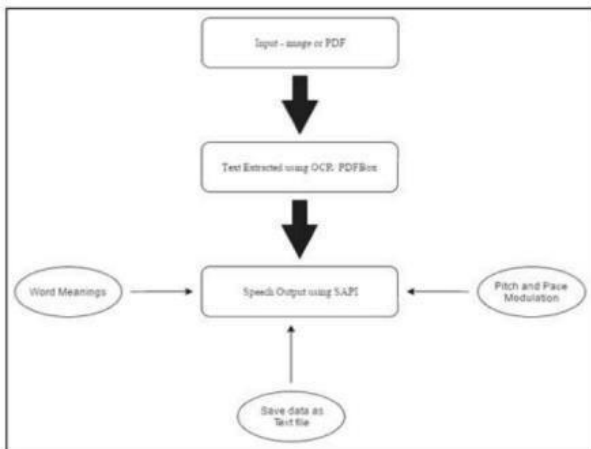
(b) Conversion from text tospeech

Figure 2 High Level Diagram of the System

The conversion of a picture into speech is handled by these two modules. The text to speech modulealso includes a number of other features, such as word definitions and synonyms, pitch and tempo adjustments for speech output, and the ability to store extracted data as text files or in the.txt format.

*2.2 IMAGE/PDF TO TEXT CONVERSION:*THE DIAGRAM SHOWS THE VARIOUS

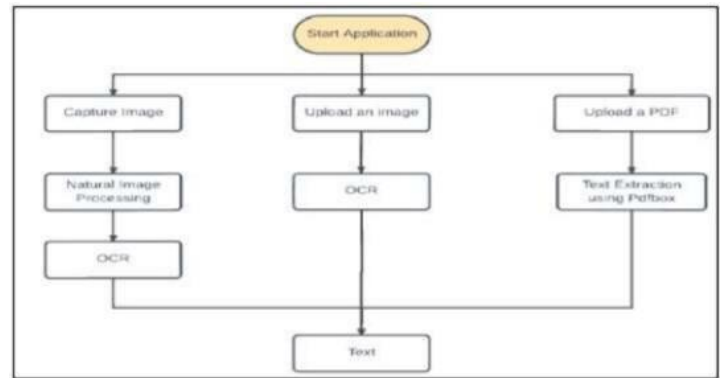FUNCTIONALITIES OF THISMODULE.

Figure 3 Detailed Diagram of Image/PDF to Text Conversion

The Image to Text Conversion module is further subdivided into three sub-modules that are incharge of Image to Text conversion.

- Capture animage
- Upload animage
- Upload a PDFfile

The technique of extracting text from various sources differs. As seen in the illustration, word extractionfroma taken image necessitates natural image processing. It begins by finding the text-containing portions of the picture. MSER (Maximally Stable ExtremalRegions) and Stroke-Width algorithms are utilized to perform this as well as locate characters. The recognized letters are then combined into words and phrases using Optical Character Recognition (OCR). Finally, a text file with the extracted textiscreated.
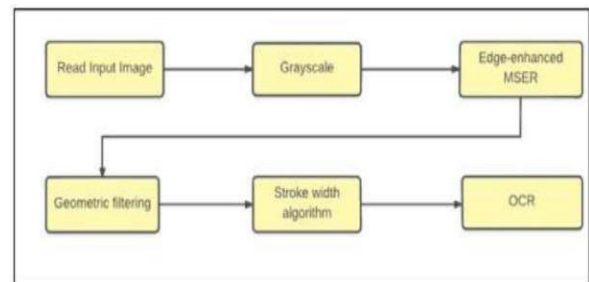
The graphic below depicts this process:

Figure 4 Natural Image Processing for Captured Image

To improve efficiency, all noise in the image is clipped before to processing. The picture is then

2

OCR' d. The extractedtextissubsequentlywrittentoafileintheworking directoryforfurtherprocessing.ThePDFBox.jarfileis used to extract text from PDFs. It is integrated with the program to allow for smooth extraction from PDFfiles.

It is important to remember that the algorithms used to analyze natural images are designed in such a way that the quality of the image does not interfere with the precision withwhichthetextisretrieved.TheOCRworkswellenough for photos with non-straight text orientations and low resolution.InanimageoraPDF,textcanbeinanyfontstyle andsize.The algorithms have been designed to correctly recognize text in any style or size. When it is converted to text format, however, these details are lost because the algorithmsareonlyequippedtoreadthetextandnotitsfont size and font style. The entire module has been designed to minimize latency and maximizeefficiency.

2.2) Text to Speech Conversion: Text retrieved from any type of input is saved in the working directory as a text file. Microsoft supplies a Speech API that is utilized to supply different voices, and an algorithm is created to translatethis text intospeech.
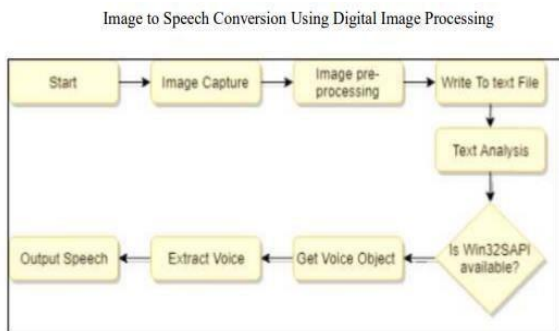
Image to Speech Conversion Using Digital Image Processing

Figure 5 Text to Speech Conversion Flowchart

The first image is taken, or, in the case of a PDF, it is uploaded. Using the prior module, text from the input is extracted and transformed to text format. The text is then analyzed to divide into sentences.

These sentences must appear in the correct order alongside the voice output. When a full stop ("."), a question mark ("?"), or an exclamation mark ("!") is
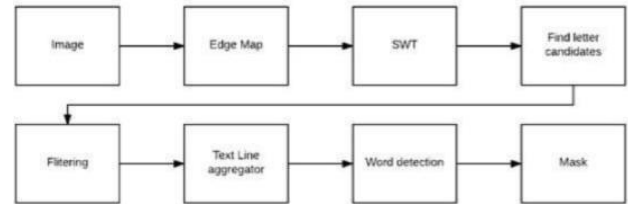

Figure 6 Stroke Width Algorithm Flowchart

encountered,thetextisbrokenintosentences.IfWindows32 SAPI is not available,the default configuration for the speech module is utilized. This module also allows users to alter the tempo and timing of their speech. Areal-time play and pause feature is alsoavailable.

Thismoduleincludesit.Thismodulealsoincludes highly useful features such as word definitions, which can be used offline because it is supported by the Microsoft Word API. The extracted text canbestored in any format for the user's future reference. Finally, this module allows the user to store the document in a variety of formats and easily access it for futurereference.

2.3) MSER Regions: For blob detection in photos, the program uses the Maximally Stable Extreme Regions algorithm.

MSER changes with the picture threshold; given a threshold value, pixels below that threshold value are

"white,"whilethoseoverorequaltothatvalueare"black." Because of the uniform color and great contrast of the text, the MSER feature detector performs well for locating text sections. The first step in constructing MSER is toexecute

a basic luminance threshold of the picture by sweeping the intensity threshold from black to white. After that, the extracted related components

or Extreme Regions are executed. Following that, a threshold is determined when an external region is maximally stable. Finally, the descriptions of the areas as MSER features are acquired.

Although the MSER algorithm detects the majority of the text, it also detects several other stableregions in the imagethatarenottext.Thestrokewidthalgorithmcanassist in resolvingthis.

(2.5) Stroke Width Algorithm: This algorithm has been implemented to retrieve the most likely stroke containing the required pixel. The algorithm receives an RGB image and returns an image of the same size, where the regions ofsuspectedtext are marked.

The first step is the stroke width transform which is an operator which determines the width of the most likely stroke containing the pixel for each and every pixel. The output produces by the SWT is an image of the same size asoftheinputimagewhereeachelementcontainsthewidth of the stroke associated with that pixel. We have now obtained a map of the most likely stroke-widths for each pixelin the originalimage.The next step is to group all these pixels into letter candidate which is done by selecting two neighboring having similar stroke width, and then applying several rules to distinguish the letter candidates.

For this reason, we have modified the classical Connected Component algorithm by alteringtheassociationrulefromabinarymaskto apredicatewhichcomparestheSWTvaluesofthe pixels. To increase the efficiency and reliability westriveforwardtogrouptheletters.Sincesingle letters are not expected to appear in images, closely positioned letter candidates are gathered together into regions oftext.

(2.5)Optical Character Recognition: The method of translating text from a digital picture into editable text is known as optical characterrecognition. It allows a machine to recognize the characters using optical methods.

The OCR result should preferably be the same as the input in formatting. The procedure begins withthe picture file being pre-processed, followed by the acquisition of critical information about the textualtext. MSER and the Stroke Width Algorithm aid in this process.

II. ALGORITHMSUSE

(3.1) Tokenization: It's the process of breaking down the text into sentences and phrases. The work entails breaking down a text into smaller chunks (known as tokens) while discarding some characters, such as punctuation.

Consider the following example:

Text input: Potter walked to school yesterday.

Potter went to school yesterday, according to the text output.

The major disadvantage of this strategy is that it works better with some languages and worse withothers. This is particularly true when it comes to tonal languages like Mandarin or Vietnamese.

Depending on the pronunciation, the Mandarin term ma can signify "a horse," "hemp," "a scold," or"a mother." The NLP algorithms are in grave danger.

(3.2) Bag of Words: This paradigm represents a text as a bag (multiset) of words, neglecting syntax and even word orderwhile keeping multiplicity. In essence, the bag of wordsparadigm generates a matrix of incidence.

These word frequencies or instances are then employed as features in the training of a classifier.

Unfortunately, there are some drawbacks to this paradigm. The worst is the lack of semantic meaning and context, as well as the fact that such terms are not appropriatelyweighted(for example,inthismodel,theword "universe" weighs less than the word"they").

(3.3) Text Summarization: As the name

implies, NLP approaches can assist in the summarization of big volumes of text. Text summarization is commonly utilized in situations such as news headlines and research studies.

Text summarization can be done in two ways: extraction and abstraction. By deleting bits from the text, extraction methods create a rundown. Abstraction tactics produce summaries by constructing newtext that conveys the essence of the original content.

Different NLP algorithms can be used for text summarization, such as Lex Rank, Text Rank, and Latent Semantic Analysis. To use Lex Rank as an example, this algorithm ranks sentences based on their similarity. Because more sentences are identical, and those sentences are identical to other sentences, asentence is rated higher.

(3.4) Lemmatization and Stemming: Two of the strategies that assist us to develop a Natural Language Processing of the tasks are lemmatization and stemming. It works nicely with a variety of other morphological variations of aword.

These strategies allow you to limit a single word's variability to a single root. We can, for example, reduce "singer,""singing,""sang,"and"sang"toasingularversion of the word "sing." We can quickly reduce the data space required and construct more powerful and robust NLP algorithms by doing this toall the terms in a document or text.

(3.5) Keyword Extraction: Keywords Extraction is one of the most important tasks in Natural Language Processing, and it is responsible for determining variousmethodsforextractingasignificantnumberof wordsandphrasesfromacollectionof texts.Allof this is done to summarise and assist in the relevant and well-organized organization, storage, search, and retrieval ofcontent.

There are numerous keyword extraction algorithms

available, each of which employs a unique set of fundamental and theoretical methods to this type of problem.There are various types of NLP algorithms, some of which extract only words and others which extract both words and phrases. There are also NLP algorithms that extract keywords based on the complete content of the texts, as well as algorithms that extract keywords based on the entire contentof the texts.

The following are some of the most prominent keyword extraction algorithms:

Text Rank: This algorithm operates on the same idea as Page Rank. Google uses this method torank the importance of various websites on the internet.

Term Frequency – Inverse Document Frequency (TF- IDF):Thefullversion ofTF-IDFisTermFrequency–Inverse Document Frequency, which tries to better define the importance of a term in adocument. Also, take into account the relationships between texts from the samecorpus.

RAKE: RAKE stands for Rapid Automatic Keywords Extraction and is a type of NLP method. This can extract keywords and key phrases from a single document's content without taking into account other documents in the same collection.

(3.6) Knowledge Graphs: Knowledge graphs are a collection of three items: a subject, a predicate, and an entity that explaina method of storing information using triples.

The subject of approaches for extracting knowledge- getting ordered information from unstructured documents includes awareness graphs.

Knowledge graphs have recently become more popular, particularly when they are used bymultiple firms (such as the Google Information Graph) for various goods and services.

Building a knowledge graph requires a variety of NLPtechniques(perhapseverytechnique

coveredinthis article), and employing more of these approaches will likely result in a more thoroughhandeffective knowledge graph.

*(3.7)WordsCloud:*Sometimesknownasatagcloud, is a data visualization approach. Words from a text are displayed ina table, with the most significant terms printed inlargerlettersandlessimportantwordsdepictedinsmaller sizes or not visible atall.

BeforeapplyingotherNLPalgorithmstoourdataset,we can utilize word clouds to describeourfindings.

(3.8) Named Entity Recognition: Another significant technique for analyzing natural language space is named entity recognition. It's in charge of classifying and categorizing persons in unstructured text into a set of predetermined groups. This includes individuals, groups, dates, amounts of money, and so on.

There are two sub-steps to named entity recognition. Named Entity Identification (the identification of prospective NER algorithm candidates) and Named Entit Classification are two of these phases (assignment of candidates to one of the predefined categories)

(3.5)Sentiments Analysis: Sentiment analysis is the most often used NLP technique. Emotion analysis is especially useful incircumstances where consumers offer their ideas and suggestions, such as consumer polls, ratings, anddebates on social media.

In emotion analysis, a three-point scale (positive/negative/neutral)isthesimplesttocreate.Inmore complex cases, the output can be a statistical score that can be divided into as many categories asneeded.

Both supervised and unsupervised algorithms can be used for sentiment analysis. The most frequentcontrolled model for interpreting sentiments is Naive Bayes.

A sentiment-labeled training corpus is required, from which a model can be trained and then utilizedto define the sentiment. Naive Bayes isn't the only machine learningmethodthat can be used; it can also employ random forest or gradient boosting.

III. RESULT ANDDISCUSSION

MATLAB is used to implement the method. The method has been tested on a variety of scanned andprinted document pictures. Various images/documents were considered, with minimal overlapping and fewer brokenand connected characters. As predicted, the program delivered a less efficient text extraction with small details for handwritten notes with moderately blurredtext.

The application is capable of identifying text form natural images as shown below:

Figure 7 Captured Image of a street Sign

Image to Speech Conversion Using Digital Image Processing

Figure 8 Applications correctly identifies the text

The application is capable of identifying hand written text as shown below:
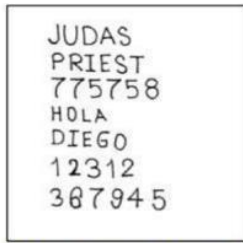
Figure 9 Captured Image of a handwritten text

Figure 10 Application detects the correct text

The application is capable of identifying text form uploaded PDF as shown:
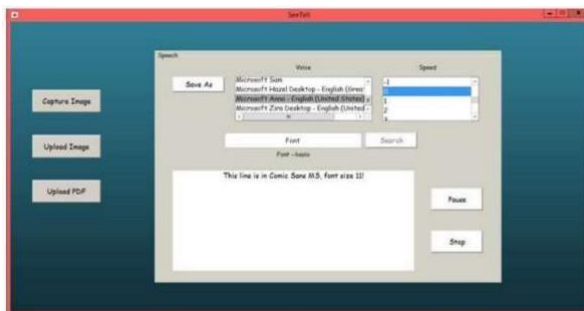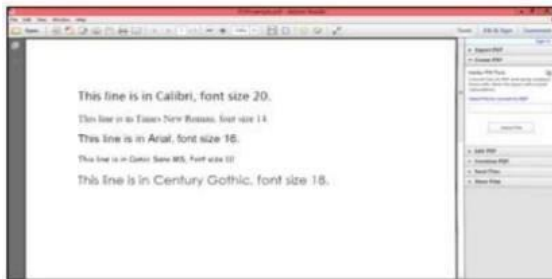
Rozelle Jain and Shantanu Das

Figure 12 Application detects the correct text from PDF file

The intermediate text is presented sentence by phrase alongside the voice in this example.

Other features are available. The pitch is set to Microsoft Anna - United States, and the tempo is set to 0, which is the typical pace. The definition of the word "font" is looked up and shown. The system is tested using the methods outlined above. Testing is essential to offer information about the quality of the developed system. It determines if the system is fit for usage.

Test case	Description	Expected Result	Result
1.	Capture Blur or Low- Resolution Image from Camera	Error: "Text cannot be extracted, please retry"	Pass
2.	Give Incorrect Path to Upload Image	Error: "Image does not exist, please retry"	Pass
3.	Give Incorrect Path to Upload PDF File	Error: "File does not exist, please retry"	Pass
4.	Search for a non-existent word meaning	Pause/Stop recitation Error: "Incorrect spelling, please recheck"	Pass
5.	Start application, Capture Image from Webcam, Recite Text	Display text sentence wise and recite the same Provide options of Pause/Play, Stop, Save	Pass

REFRENCES

[1] ArchanaA.Shinde,D.G.Chougule"TextPre-processing andTextSegmentationforOCR"IJCSET,January2012.

[2] Benjamin Z. Yao, Xiong Yang, Liang Lin, MunWaiLee and Song-Chun Zhu, "I2T: Image Parsing to Text Description".

[3] Bernard GosselinFacultéPolytechnique de Mons, Laboratoire de Th´eorie des Circuitset.

[4] Huizhong Chen1, Sam S. Tsai1, Georg Schroth,David M. Chen, RadekGrzeszczuk and Bernd

[5] JishaGopinath, Aravind S, PoojaChandran, SaranyaS S, "Text to Speech Conversion System using OCR", International Journal of Emerging Technology and Advanced Engineering,

January2015.

[6] Itunuoluwasewon, JeliliOyelade, OlufunkeOladipupo, "Design and Implementation of Text To Speech Conversion for Visually Impaired People", International Journal of Applied Information Systems(IJAIS,2014).

[7] Yao Li and HuchuanLu, "Scene Text Detection via Stroke Width", 21st International Conferenceon Pattern Recognition (ICPR 2012) November 11-15, 2012. Tsukuba,Japan.