Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 11, No. 3, November : 2023

AUTO SPEECH AND MULTIPLE LANGUAGE TRANSLATOR

Dr. T. V. Hyma Lakshmi, Associate Professor, Dept. Of ECE, S R K R Engineering College.

Abstract
As per the World Health Organization (WHO), nearly 285 million people were suffering from visual impairment, and nearly one-fourth of the total population was facing an issue of language barrier. So in order to eliminate the risk of individual development for people with visual impairment and for those who are facing issues of language barrier, a system is being proposed through our work as an "auto speech and multiple language translator". It is a computer-based application that follows four modules: image processing, text extraction from images, text translation to multiple languages, and conversion of text to audible speech (saved as a.mp3 file). The extraction of text from images, language translation, and text-to-speech conversion are done by using OCR and deep learning techniques. Subsequently, by using this system, the text information to English language translation acts as a dictionary and promotes effective communication; the conversion of text information to audible speech acts as an intelligent text-reader, which helps the visually challenged.

Keywords:
Image Processing, OCR technology, Language Translator, image extraction, text to speech.

I. Introduction

There is a huge demand for character recognition when an image is scanned. Recognition of text and its extraction from an image is one of the most complex processes because of the different backgrounds [1-3]. There are different methods for text recognition. In traditional days, clustering and pattern matching techniques were used. Clustering is an unsupervised machine learning task that groups the unlabeled dataset. It involves automatically discovering natural groupings in data. But there are some disadvantages to clustering when dealing with a large dimensions and data items because of time complexity [4]. Similarly, there are two issues with pattern matching. This kind of recognition is difficult to execute, and it is an extremely slow method. Secondly, it requires a larger data set to acquire enhanced accuracy[5-7].

The recognition of text from images is still an active research area. But there is a way that software like optical character recognition (OCR) can help us recognise characters from a given image. The purpose of the "Auto speech and multiple language translator" is to recognize the text(in any language) from an image and convert it into a base language(English). The Python API is used for translation to English from any other language. Then the text is converted into audible speech which is saved under .mp3 file. In image processing, text recognition can be done by using OCR. Firstly, the images must be scanned properly. In order to detect text from an image, OCR checks lines by lines and words by words by using a unit test framework. The extracted text can also be edited or deleted manually.

II. Proposed System

2.1 Objective of the Proposed System

Optical Character Recognition (OCR) includes Pre-Processing technique, Feature extraction, Segmentation and Recognition. As a result, when we capture and scan an image, OCR detects and extracts text from image by eliminating other data. It also translates text from other languages to English language.

2.2 Methodology of the proposed system

The suggested approach consists of four interconnected sections- Image processing, text extraction, text to speech conversion and multiple language translation.

2.2.1 Image detection

Image processing is a type of signal processing in which the input is an image and the output may be an image or features related to that image [8]. There are various components to be taken into consideration when performing operations on an image in order to obtain an enhanced image or extract some useful information from it. Those are image acquisition, image enhancement[9-12], image restoring, color image processing, wavelets[13-15] & multiresolution processing[16-19], compression, morphological processing, segments, representation & description, object recognition.

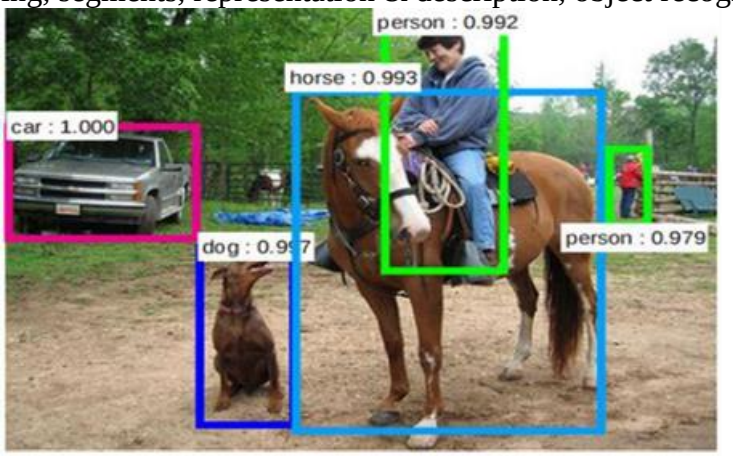

Figure 1. Image detection using deep learning

Detection of the image and detection of the object in the image including words, character, and lines in an image is important to extract text from an image. There are two types of detections: Image detection and Line detection. The detection of an image using deep learning is shown in figure1.

2.2.2 Line detection

In image processing, line detection is an algorithm that takes a collection of 'n' edge points and finds all the lines on which these edge points lie. The most popular line detectors are the Hough transform and convolution-based techniques.

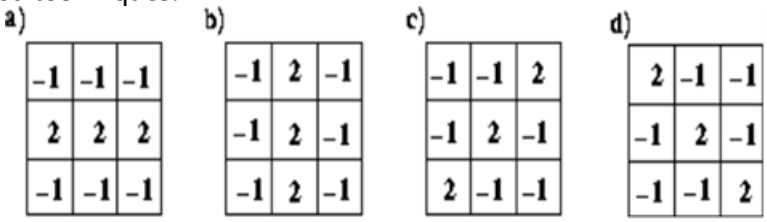

Figure.2. Line detection using 3x3 matrix

As shown in the above figure 2, four-line detection kernels that respond maximally to horizontal, vertical, and oblique (+45 and -45 degree) single-pixel wide lines. In line detection, filters are used to detect an exact line in an image, or else different methods are used to perform line detection. And those function in the format of scanning an image and using the unit test framework, which will be suitable with the Python programming compilers, and those compilers will understand the importance of the unit test framework and those that perform line detection.

Figure.3. Block diagram of the text extraction process

The module for text extraction will come next. If necessary, the text that was retrieved from the images is then sent to the text-to-speech component. The OpenCV.dll and Emgu.CV wrappers are used in this module. To call OpenCV functions from VB.Net, use Emgu.CV. In EmguCV, there are six DLLs. The dll files Emgu.CV.UI, Emgu.CV.GPU, Emgu.CV.ML, Emgu.CV.OCR, and Emgu.CV.Util make up Emgu.CV. For user interfaces like image boxes, use Emgu.CV.UI.dll. The Tesseract-OCR library is used by Emgu.CV.OCR.dll to carry out optical character recognition. The Tesseract object is used in this module to separate the text from the images and show it on the screen.

2.2.3 Text translation

In text translation, the text extracted from the image in the previous module is taken as input and converted into any another language. First, the input text in any language is converted to the base language. The base language we considered here is English (source). Then, the user has to select a destination language (desired language). The text in English is split into words, and then it is searched in the dictionary based on the destination language given by the user, and the corresponding matched text or word from the dictionary is displayed in the desired language.

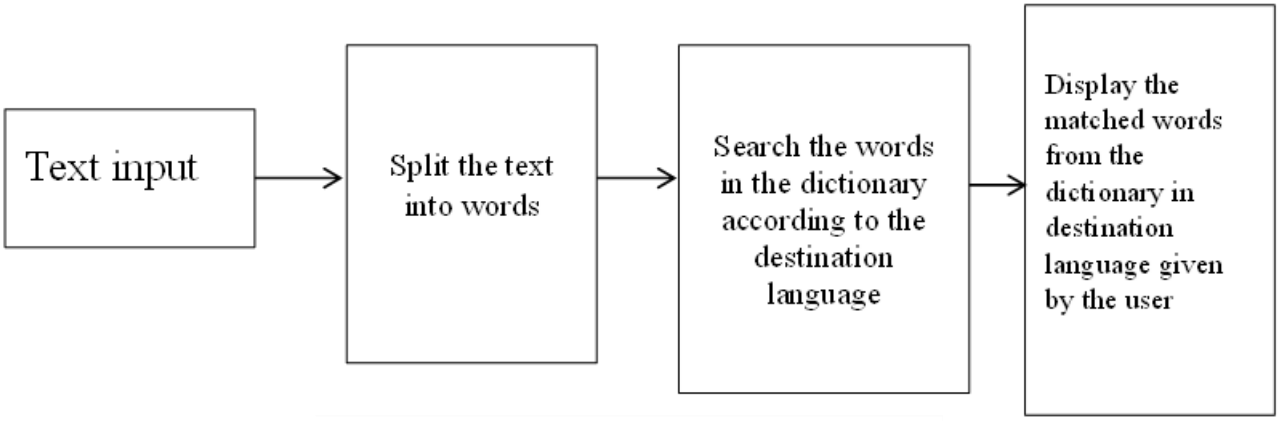

Figure.4. Block diagram of text translation.

2.3 Text to speech conversion

Any wording, whether chosen or chosen at random, can be transformed into voice using a text-to-speech system. Integrating taped voice that is saved in a handlist will produce voice as an output. There are primarily two parts: the first is the input text synthesis, and the other is the conversion of the text to speech. Speech integration is the term used to describe the procedure of turning text into voice, and the computer device used to achieve this is known as a voice integrator. It primarily consists of two activities: word standardisation and tokenization. As a result of this technique words, phrases, clauses, and sentences are converted into the spoken transcription.

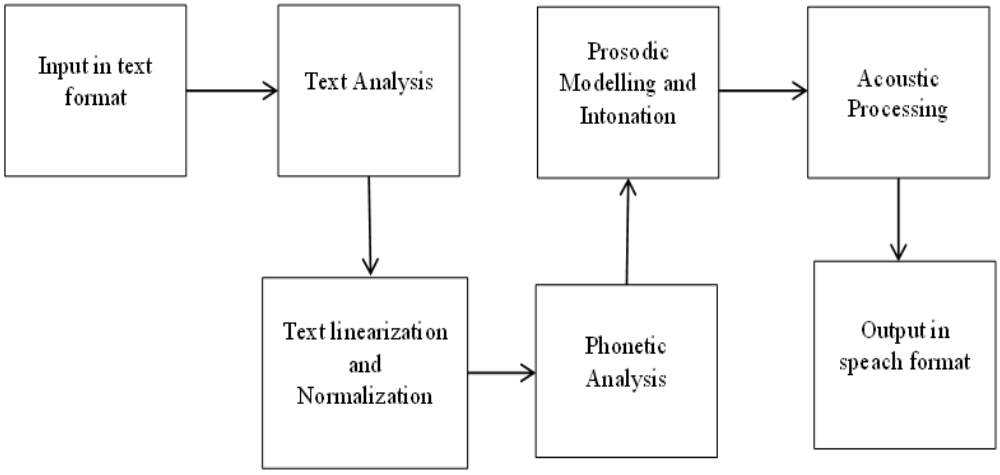

Figure.5. Block diagram of text to speech conversion

2.3 Text to speech conversion

The technique of supervision and detecting significant details and standards in textual data is known as text estimation and disclosure. It might include tasks like sentiment study, language identification, named entity understanding, and topic modelling. Textual analysis and detection are frequently used in marketing, customer service, news media, and other industries to derive relevant information from text data to aid in decision-making and to better accept text data, and the results of these analyses and detections can be used in these contexts.

2.4 Text Linearization and Normalization

Text normalisation, in its simplest terms, is the process of changing words into a pronounceable form. Eliminating punctuation and changing both uppercase and lowercase characters are both done during the text normalisation process. When contrasting characters have the same meaning, it works effectively: for example, "can't" and "cannot," "I have" and "I've," and "don't" and "do not." Normalisation includes stages such as word segmentation, acronym conversion, number conversion, and abbreviation conversion.

2.5 Prosodic Modelling & Intonation

In addition to the sound of the words being said, prosody is a unit of speech that contains numerous other aspects of speech. Prosodic analysis and its modelling take into account the words spacing, the speech's pitch and pace, as well as its duration. The shift in our facial expression as we utter particular words or sentences is known as tonal variation.

2.6 Phonetic Analysis

Utterances and their observable traits are explored through phonetic analysis. It deals with connected speech and how users hear it. This includes how speech sounds are produced, conveyed, and received. Combining expertise from language studies, speech science, and computer programming, phonetic analysis is a diverse disciplinary field. It is employed to construct voice control systems in addition to analysing speech patterns and transforming spoken words into graphic text. Additionally, phonetic analysis has been used in the creation of speech synthesisers, speech therapy, and research on linguistic and cultural variations in speech.

2.7 Phonetic Analysis

The final phase involves producing the speech waveform for each word and sentence using both prosody and phonemes. The first processing technique involves joining recorded voice samples together. A chunk is essentially a group of words. The second method produces formants by using signal processing. It is possible to construct the required software programme or code module using the VB.net Speech Software Development Tool. Users can immediately insert text into the editor by using a text entry area. Based on the input text, the synthesiser component changes the text into speech. Speech output produces the desired sound for the text it is associated with.

2.8 Block diagram of proposed system

Figure.6 shows describes the entire set of process steps included in the proposed model.

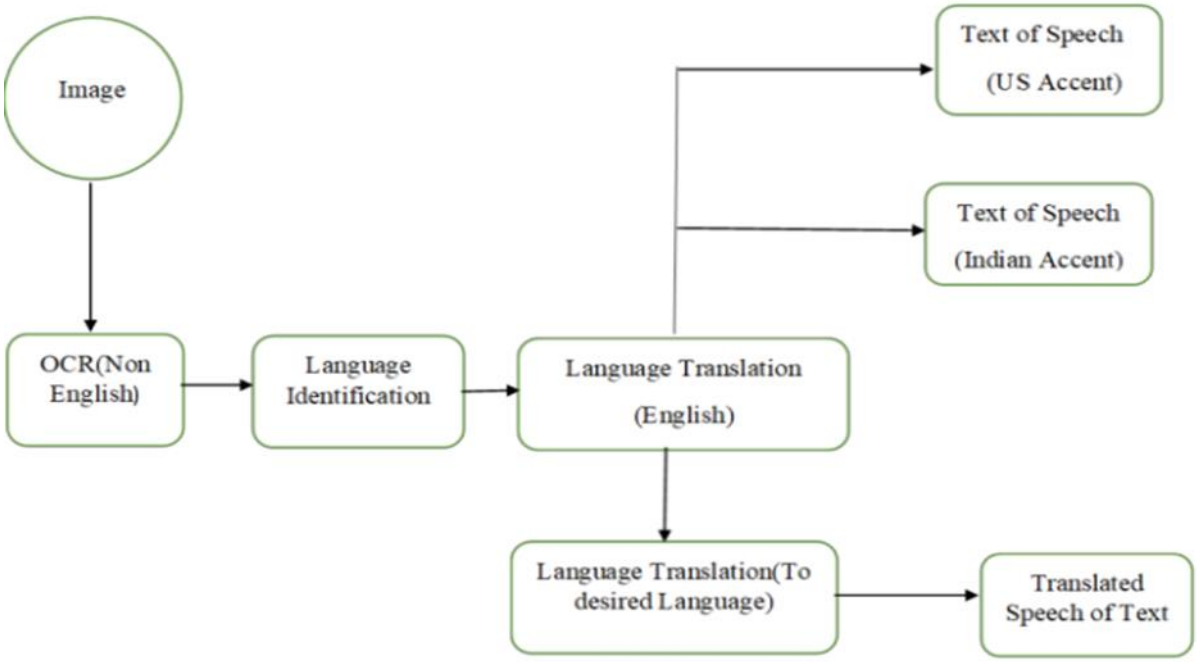

Figure 6. Block diagram of proposed system.

2.9 Applications

2.9.1 Easy Translation at Bus stops

The proposed system is useful in that if there are people at a bus stop and all destination locations are written in Telugu but they only know English, they can take a picture of the destination location board and use this application to easily generate an English prompt and easily reach the destination.

2.9.1. Document Making

A Newspaper of any Country or State of any language can be easily translated into an English language by scanning the paper and the user can get output in the form of speech if needed.

If there are 100 pages of important information that he/she needs but that information is available in a different language, in such cases, our application will solve the problem immediately by translating any language into English for easy understanding. If the information is in an image and the user requires it in text format, the proposed system allows the user to easily extract the text from the image.

III. Conclusion

In the suggested system, we built in support for users who could experience issues owing to a language barrier, and the user interface is also designed to be simple enough for users to engage with it without

difficulty. In other words, this approach eliminates the need for dictionaries, which instantly lessens the user's effort in learning the languages needed for communication.

References

[1]Ray, Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, "Object Detection with Deep Learning: A Review" Fellow, IEEE.
[2]Swapnil Sinha, Harsh Kumar Kataruka, Vijayakumar Kuppusamy, "Image to Text Convertor and Translator using Deep Learning and Image Processing", International journal of Innovativetechnology and exploring engineering (IJITEE), ISSN:2278-3075, Volume Issue-8 June 2020.
[3]Pratik madhukar manwatkar, Shashank H. Yadav, "Text Recognition from Images", IEEE sponsored 2nd international conference on innovations in information, Embedded and communication systems (ICIIECS), 2015.
[4]Implementation of Text Recognition and Text Extraction on Formatted Bills using Deep Learning - International Journal of Control and Automation Vol. 13, No. 2, (2020), pp. 646 – 651.
[5]Yang, Jufeng, Kai Wang, Jiaofeng Li, Jiao Jiao, and Jing Xu, "A fast adaptive binarization method for complex scene images," 19th IEEE InternationalDr.K.CH.Srikavya, "Satellite Image Enhancement based on the Half-Band Polynomial Sub Bands Fusion and CLAHE", Conference on Image Processing (ICIP), 2012.
[6]Shrey Dutta, Naveen Sankaran, PramodSankar K., C.V. Jawahar, "Robust Recognition of Degraded Documents Using Character N- Grams," IEEE, 2012.
[7]T.V.Hyma Lakshmi, G.Anusha, "An Efficient Guidance for Robots Using Artficial Intelligence ", (SCOPUS) June-2019, International Journal of Advanced Science and Technology (IJAST), Vol-127, pp:327-336.
[8]T.V. Hyma Lakshmi , "A Road Safety System For Hill Stations With Automatic Controls" International Journal Of Research And Analytical Reviews (Ijrar.Org) vol.9, Issue. 4, 2022, E-ISSN 2348-1269, P- ISSN 2349-5138.
[9]T.V. Hyma Lakshmi, K.Ch. Sri Kavya, T.Madhu, Sarat K Kotamraju, " Image resolution and contrast enhancement with optimal brightness compensation using wavelet transforms and particle swarm optimization" IET Image Processing, vol. 15, Issue:49, pp: 2833 -2840.
[10] T.V.Hyma Lakshmi, Ramesh Nayak.D , "Denoising of Image and Audio by using Wavelet Transform", International Journal of Advanced Research in Computer Science, Volume 4, Issue 5, May 2013, Page:88-92
[11] T. V. Hyma Lakshmi, T. Madhu, K. Ch Sri Kavya. "Gaussian Mixture Model and Lifting Wavelet Transformed Base Satellite Image Enhancement" International Journal of Recent Technology and Engineering, vol.8, Issue.2, pp.29-34.
[12] T.V.HymaLakshmi, Dr.T.Madhu, (SCOPUS), April,2018, Jour of Adv Research in Dynamical & Control Systems, vol:10, Issue:4, pp:333-338.
[13] T. V. Hyma Lakshmi, T. Madhu K. Ch. Srikavya, K. Sarat Kumar, "Study on Image Enhancement Using Lifting and Stationary Wavelet Transforms and Contrast Limited Adaptive Histogram Equalization" Book Chapter, September 2022, New Visions in Science and Technology , vol.4, Issue:23, pp:50-59.
[14] T.V. Hyma Lakshmi, K.Ch. Sri Kavya, T.Madhu, Sarat K Kotamraju, "Satellite Image Enhancement using Optimized Wavelet Decomposition and Bicubic Interpolation", January, 2020, "International Journal of Recent Technology and Engineering (IJRTE)", vol: 8, Issue:5, pages: 4430-4434, ISSN: 2277-3878.
[15] T.V.HymaLakshm, Dr.T.Madhu, Dr.K.CH.Srikavya, "Image enhancement technique using lifting and stationary wavelet transforms and contrast limited adaptive histogram equalization",

2018, International Journal of Engineering &Technology, vol:7, Issue: 4,pages:6711-6714, https://doi.org/10.14419/ijet.v7i4.16833.

[16] T.V.Hyma Lakshmi, Dr.T.Madhu, Dr.K.CH.Srikavya, Shaik Esub Bhasha, " Novel Image Enhancement Technique Using CLAHE and Wavelet Transforms", Nov-2016, International Journal of Scientific Engineering and Technology, vol: 5, Issue:6, pages: 507-511, ISSN (Online) : 2277-1581

[17] T.V.Hyma Lakshmi, Dr.T.Madhu, Dr.K.CH.Srikavya, "Image Resolution and Contrast Enhancement Using Wavelet Transform and Contrast limited adaptive histogram Equalization", International Journal of Computer Science and Information Security, vol:14, Issue: 9, pages: 969-981, Sept-2016.

[18] T.V.Hyma Lakshmi, Dr.T.Madhu, Dr.E.V.Krihna Rao, V.Lakshmi Mounica Satellite Image Resolution Enhancement Using Discrete Wavelet Transform and Gaussian Mixture Model, International Research Journal of Engineering and Technology (IRJET), vol. 2, no., 4 95-101, July,2015.

[19] T.V.Hyma Lakshmi and D.Ramesh Naik, "Denoising of Image and Audio by using Wavelet Transforms ", International Journal of Advanced Research in Computer Science, vol:4, Issue: 5, pp:88-92, May 2013, ISSN: 0976-5697.

Dr.Tirumani V Hyma lakshmi obtained Ph.D from K L UNIVERSITY on the area of Image processing, M.Tech.(ECE) from JNTUCE, Ananthapur and B.E. (ECE) from S.R.K.R.Engg college, Bhimavaram. Presently working as an Assistant Professor in S.R.K.R.Engg college, Bhimavaram. Published papers in various SCI (IET Image Processing), SCOPUS, reputed international journals and Springer conferences. She has patents, completed various NPTEL courses, organized FDPs and various student activities as an IETE student advisor, women empowerment cell coordinator, Internal Complaint Committee member, etc. Won the Global Faculty Award. Her research interests are image processing, Radar Engineering, VLSI, etc. Fellow ship in ISTE and IETE.