iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35371

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Transcript Anatomization with Multi-Linguistic and Speech Synthesis Features

Rohan Modi[1], Ridhima Tyagi[2], Rashi Chaudhary[3], Preetish Gaur[4]
Student, CSE Department, Raj Goel Institute of Technology, Ghaziabad

***Abstract*—Handwriting Detection is a process or potential of a computer program to collect and analyze comprehensible input that is written by hand from various types of media such as photographs, newspapers, paper reports etc. Handwritten Text Recognition is a sub-discipline of Pattern Recognition. Pattern Recognition is refers to the classification of datasets or objects into various categories or classes. Handwriting Recognition is the process of transforming a handwritten text in a specific language into its digitally expressible script represented by a set of icons known as letters or characters. Speech synthesis is the artificial production of human speech using Machine Learning based software and audio output based computer hardware.**

While there are many systems which convert normal language text in to speech, the aim of this paper is to study Optical Character Recognition with speech synthesis technology and to develop a cost effective user friendly image based offline text to speech conversion system using CRNN neural networks model and Hidden Markov Model. The automated interpretation of text that has been written by hand can be very useful in various instances where processing of great amounts of handwritten data is required, such as signature verification, analysis of various types of documents and recognition of amounts written on bank cheques by hand.

Keywords— HTR - Handwritten Text Recognition, NLP - Natural Language Processing, CRNN - Convolutional Recurrent, Neural Networks, OCR - Optical Character Recognition, HMM - Hidden Markov Model.

I. INTRODUCTION

Handwriting Recognition or HWR, also known as Handwritten Text Recognition (HTR), is the process in which understandable handwritten input from various sources for example paper documents, photographs, touch- screens and other devices are received by the computer and accordingly it interprets it[4].

Handwriting Recognition, also known as Handwritten Text Recognition (HTR), is the procedure during which comprehensible handwritten input from various sources, for instance paper documents, photographs, touch-screens and other devices are received by the system and accordingly it interprets it. HTR can be broadly classified into Online HTR and Offline HTR. Online HTR comprises of the automated transcription of scripts while it is being written down on the basis of a unique digitizer or PDA, where a sensory device picks up the movements of the tip of the pen along with the pen-up and pen-down shifting. The image of the written text may either be detected "off-line" from a piece of paper by optical scanning (optical character recognition) or by the intelligent word recognition. Alternatively, the movements of the pen tip could also be sensed "on-line", for instance by a pen- based display screen surface, a generally easier task as there are more clues available.

Offline handwriting recognition is relatively difficult, as different people have different handwriting styles. A handwriting recognition system adheres to composition and correct character segmentation of handwritten letters and their formatting, and finds the foremost plausible words. Offline handwriting recognition involves the automated transcription of text in a picture into character symbols that are ready to be used within computer based as well as text- processing software programs. During conduction of this procedure, many models are framed to map the extracted attributes to various classifications and thus recognizing the characters or words that are represented by the features.

Speech synthesis is the creation of artificial replication of human voice based speech. The device used for this purpose is known as a speech synthesizer, which can be integrated within the software or hardware components of a system[1]. Since the last few years, the implementation of text-to-speech conversion technology has grown over a variety of applications such as the the rapidly growing use of speech based response systems and the digital voice storage systems for voice mails.

The application of the text to speech engine in furnishing symbolic linguistic renditions like phonetic transcriptions into speech has been discussed within the scope of this research paper.

II. LITERATURE REVIEW

HTR systems have used the Hidden Markov Models (HMM) [1] for the transcription task, but recently, through Deep Learning, the Convolutional Recurrent Neural Networks (CRNN) approach has been used to overcome some limitations of HMM. To exemplify a CRNN model, we bring the model (Fig. 1):

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

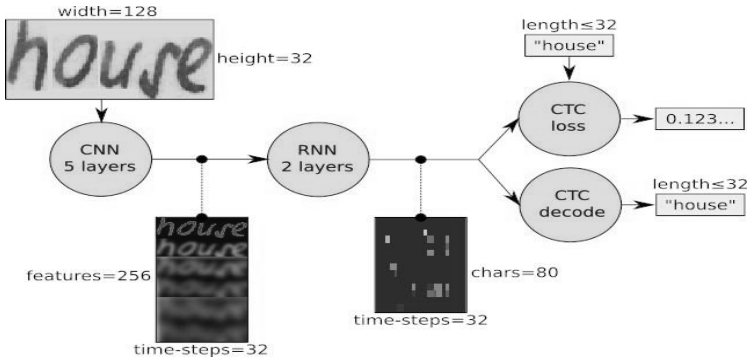


Fig. 1 Overview of a CRNN

The workflow can be divided into 3 steps [2]:
Step 1: the input image is provided into the CNN layers to extract characteristics. The output is a feature map.
Step 2: through the implementation of the process of Long Short-Term Memory (LSTM), the RNN is able to move information over longer distances and provide more robust features to training.
Step 3: with RNN output matrix, the CTC, calculates loss value and also decodes into the final text.
In addition, the char-set for encoding text is also the same for all datasets. So, the list used consists of 95 printable characters from ASCII table (Fig. 2) by default and doesn't contain accented letters.

```
!"#$%&\'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\\]^_`abcdefghijklmnopqrstuvwxyz{|}~
```

Fig. 2 Char-set adopted to encoding text

Now the question arises why to employ the identical CTC [3] and charset for the structures and data-sets. For this experiment a more uniform approach between the models and databases can be considered. Since the utilization of agenda specific dictionaries and character sets can end up in delivering more appreciable results in some scenarios, but not as well in some others. For instance, a charset consisting of only 60 characters is "easier-to-hit" (1/60) than another with 125 (1/125). In fact, the proposition here becomes tougher for text recognition by the models. The fundamental significant way in some identification of handwritten English character scheme is pre-processing followed by segmentation procedure and feature extraction procedure.
Handwritten alphabets recognition system's accuracy [1] of any photograph is dependable upon the sensitivity of the selection of features and class of categorization employed. Therefore, such a large amount of feature removal and classification techniques may be implemented within the literature. This research work executes handwritten alphabets recognition of the English language's phrases.
To identify cursive letters we can take into consideration the Holistic Technique. This technique is brought in to action for signifying phrases via various different transformation phases like phrases, features, points, letters, contours, and phrases. The Feature vector is exclusively constructed from the image to build arithmetical relationships among letters and features. Partially calculated characters are recognized via evaluation through lexicon. The lexicon comprises of only 130 words; thus, limited number of words and features are identifiable.
The vector of the features is constructed from the borderline data of characters which includes position of the boundary in contact with 4 situation lines, its node, event, degree and curve to the boundary etc. 10 dimensional vectors of the attributes is built. HMM for each of the characters of the alphabet is structured and by amalgamation of those HMMs, the HMM for each thesaurus alphabet is structured. Limited ranged thesaurus is employed.
HMM is framed to acknowledge both cursive as well as isolated handwritten characters. The writer utilizes recognition technique by applying HMM. Hybrid technique is utilized to get the best use of the efficiency of the HMM Algorithm. To acknowledge alphabet features, it analyzes the averages of black run in every scanned line. Alphabet image is examined in 4 various directions to mine out the features from it. Average of each direction raises a sparse directional frame of the alphabet. The isolated compactness from left to right HMM method is utilized for interpretation.

III.PROBLEM STATEMENT AND SOLUTION APPROACH

Since a very long time, human used to write their thoughts in the form of letter, transcripts etc.; in order to convey them to others. But since the development of computer technology the format of handwritten text changed rapidly to computer

generated digital text and so people feel a need of such method that can transcribe the text written by hand to digital text because it enables the processing of this data quite easy and fast.

The methodology utilized in this project can be understood with the following flow chart(Fig. 3).

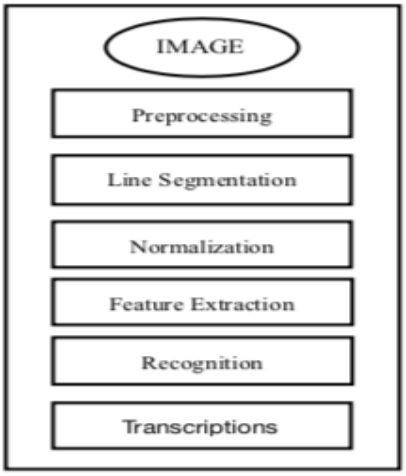


Fig. 3 Handwritten recognition system overview text

The various stages of the HTR methodology utilized in the scope of this research paper have been explained as follows:

A. Pre-Processing

Preprocessing [4] is an actions chain which is carried out on the scanned input image. In preprocessing, the goal is to discard irrelevant information. Irrelevant information includes duplicate points, noise, etc. The approach consists of removing irrelevant data and performing normalization. As a result it can easily detect the symbols. It basically enhances the image, furnishing it appropriately for segmentation.

During preprocessing of an image, a variety of operations are executed on it, as discussed below:

1) *Removal of Noise:* The preprocessing stage enhances the superiority of the raw image and puts the data of significance. It is also acknowledged as pixel level or low-level processing, which is arranged on the detained image to manage it for additional analysis.

Fig. 4 Pre-processing phase results, including noise removal and skew correction

2) *Skew Correction:* Skew Correction is used to correct the text line in scanned document images if they are not (left-and-right) matched up during the scanning process.

B. Line Segmentation

In the Segmentation phase under pre-processing, the image structured as a sequence of alphabets is decomposed into sub-images of single character each. The input image which was preprocessed is fragmented into separated alphabets by conveying a number to each alphabet by implementing a labelling procedure. This labelling system provides knowledge regarding the number of characters present within the image. Then, every single character is restructured in a homogeneous form, conforming it into a 90*60 pixels format for the next stages of categorization and recognition phases.

C. Normalization

Once the lines are extracted from the document image, the Normalization [6] step is performed before extracting the features. The foremost intent of the Normalization phase is to eliminate the differences that might otherwise complicate the classification and reduce the recognition rate of characters that are having similar structures with different words. The major generic foundations of variability in handwritten text based images is the volume and incline of the letters.

1) *Slant Correction:* Slant Correction rectifies the inclinations or deviations of the slant of text written by hand. By applying a shear transformation, the writing's slant is transformed into its upright position.

D. Features Extraction

Features play an essential role in Handwriting Recognition(HTR) Systems. Their main goal is to extend the interpretation rate by effectively categorizing the data. Feature extractions concerns mainly with extracting the major aspects of the data from an image's pixels based upon the ability to be efficiently refine and the utilize the extracted data.

E. Recognition

Interpretation of characters written by hand is a fairly demanding task. The alphabets might be scripted in varying directions, measurements, arrangement, widths and sizes. That may provide a lot of differences. The efficiency of neural network (NN) in simplification and insensitivity to the misplaced data is very beneficial in recognizing handwritten alphabets.

F. Transcription

In this stage of pre-processing, preciseness of classification is amplified additionally by linking the scheme with the dictionary with an object of achieving Syntactic Analysis and Semantic Analysis like higher-level categorization, that pertains in the identified alphabet's verification.

G. Speech Synthesis

Speech Synthesis can be defined as the artificial synthesis of human speech by computer systems. Synthesized speech can be generated by merging pieces of recorded speech that may be stored in a database.

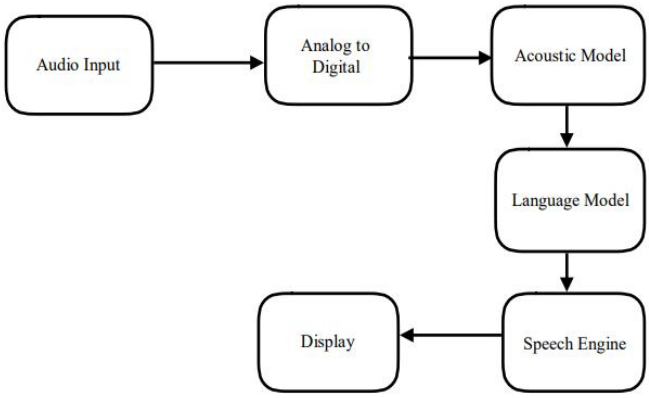


Fig. 5 Speech Synthesis

IV. CONCLUSION

Handwritten Character Recognition (HCR) is a cumbersome task, that isn't easily solvable. The need surrounds datasets and databases. This model is developed to read the text and convert it into voice based audio formats.This application is applicable in many sectors of healthcare and consumer sector. This type of model used in health application can save understanding perspectives of people and store each and every record digitally.

This application is usable in various fields of Consumer and Healthcare Sectors. On implementation in the medication and healthcare sector, this model saves the tedious task of understanding as well as comprehending the perspectives of different individuals physically by enabling the storage of each one digitally.

It can be refined and improved further so that the system can recognize any character, words or an entire document which is written by any writer. Document retrieving may also be done as a future job.

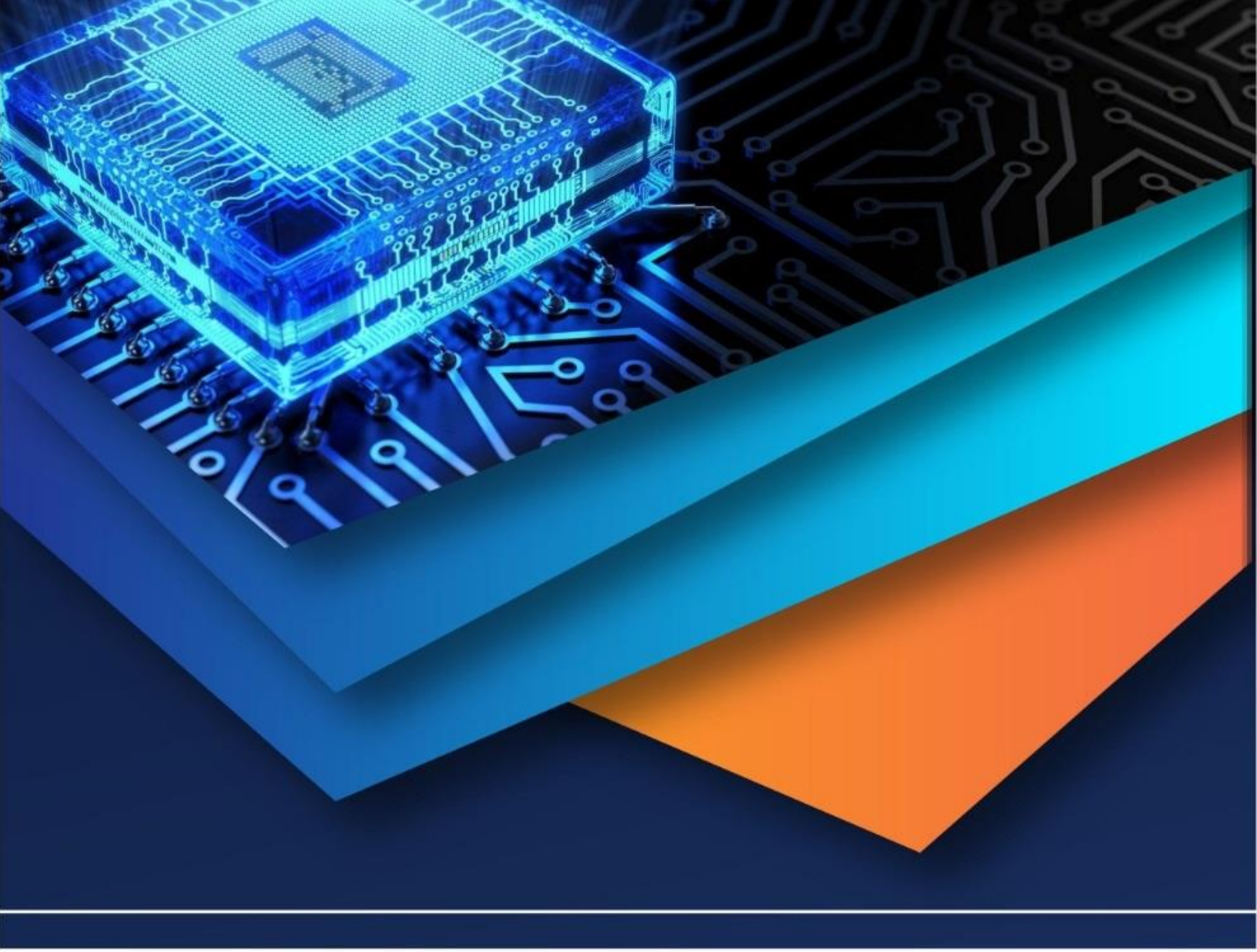
REFERENCES

[1] Sonkusare, M. and Sahu, N., 2016. A survey on handwritten character recognition(HCR) techniques for English alphabets. Advances in Vision Computing: An International Journal (AVC), 3(1).

[2] icea.org.in/wp-content/uploads/2020/07/Contribution-of-Smartphones-to-Digital-Governance-in-India-09072020.pdf

[3] Preetha, S., I. M. Afrid, P. Karthik Hebbar, and S. K. Nishchay. "Machine Learning for Handwriting Recognition." International Journal of Computer (IJC) 38, no. 1 (2020): 93-101.

[4] Islam, N., Islam, Z. and Noor, N., 2017. A survey on optical character recognition system. arXiv preprint arXiv:1710.05703.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY