ISSN 2395-1621

Proofreading/Quality Checking Tool

Sagar Agarkar, Abhishek Jadhav, Ganesh Raut, Prof. Megha Jadhav

sagaragarkar200@gmail.com
abhijadhav8001@gmail.com
ganeshraut3938@gmail.com
meghajadhav004@gmail.com

Computer Engineering
JSPM's, BSIOTR, Wagholi, Pune, India.

ABSTRACT

Text recognition is a technique that recognizes text from the paper document in the desired format (such as .doc or .txt). This document describes techniques for converting the textual content of a paper document into a machine-readable format using OCR (optical character recognition) principles and concepts. Text Recognition used in official task in which the large data have to type like post offices, banks, colleges etc., in real life applications where we want to collect some information from text written image. So by using this concept anyone can scan a document and have the text of that document available in a .txt or .docx format.

General Terms: Optical Character Recognition Algorithm

Keywords: OCR, Proofreading, Quality Checking.

ARTICLE INFO

Article History

Received: 6th June 2020

Received in revised form : 6th June 2020

Accepted: 9th June 2020

Published online : 9th June 2020

I. INTRODUCTION

There is a lot research work has done on Pattern Recognition which comes under Machine Learning, Artificial Intelligence. OCR well known as Optical Character Recognition is one of the leading branch of the Pattern Recognition [1].

The system reduces human efforts along with time. It can also be helpful for the person who doesn‟t know the language / pronunciation of particular words. The system can also be helpful for visually impaired or person with weak visual ability. Optical character recognition is the mechanical or electronic conversion of images of typed, handwritten or printed text into machine -encoded text, whether from a scanned document or a photo of document [2].

It is widely use as form as a form of information entry from printed paper data records, whether passport documents, invoices, bank statements, computerized receipts, business cards, mail, printouts of static data or any suitable document. OCR is a field of research in pattern recognition, artificial intelligence, computer vision [3].

The app uses a camera of an Android mobile device to take an input. Input is a binary image scanned by the camera. The OCR engine processes the image data and converts it into a text [4].

The respective text is then sent to Android Text-to-Speech. Android text-to-speech is an engine which has ability to convert the text into a speech. The system uses machine learning, it takes a training data and learns from it, hence the accuracy of the output grows down the pages, pass b y pass [5]

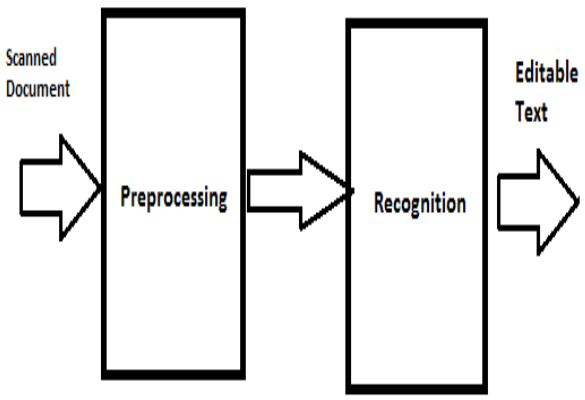

Fig 1. OCR engine

II. PROBLEM STATEMENT

Now-a-days Machine learning has become one of the peak of technology. Previously it was not possible to compute data at higher or faster rate, with the help of leading technology it is now possible to process data at higher rate to get optimized hence better result.

To image to text conversion is challenging task for maintain accuracy words.

III. LITERATURE SURVEY

[4]Esmeralda C. Djamal projected Autography movement emulate the written element of each individual's periodicity and design. By analyzing all fundamentals of handwriting and interpreting them, using typical of graphology author could initiate a chart of the writer's character attribute, sentimental constitution and gracious design. In graph logical analysis's, an image is separated into two accession that graphics attributes and partition digit each character. In this research, author employ graphical accession based on signature and digit of character of consumption scheme using many-frame algorithms and artificial neural networks (ANN). The image crack into two space: the signature occupied on nine appearance and consumption scheme of letters digit space. Each space had performed preprocessing to improve the recognition accuracy.

[5]Sandeep dhang on Handwriting Analysis of Human Behaviour Based on Neural Network, Graphology or Handwriting analysis is a scientific method of identifying, evaluating and understanding of anyone personality through the stroke and pattern revealed by handwriting.
Handwriting reveals the true personality including emotional outlay, honesty, fears and defenses and etc. Handwriting stroke reflects the on paper draw of each individual's rhythm and Style. The image split into two areas: the signature based on three features and application form of letters digit area.

[6] Javier Galbally, Julian Fierrez, Marcos Martinez-Diaz, R´ejean Plamondon E´cole Polytechnique de Montre´al focus on "Quality Analysis of Dynamic Signature Based on the Sigma- Lognormal Model". In this paper author distinct that various personal ethics can be precisely illuminate as a set of influential describe sequenced together by a Markov chain. To diagnose personal ethics from sensible data and to deduce personal ethics over a few seconds time, author then use these influential Markov layout. To ensure the virtue of this designing avenue, creator report an experiment in which, author was able to achieve 95% precision at predicting automobile driver's subsequent actions from their starting preparatory movements.

IV. PROPOSED SYSTEM

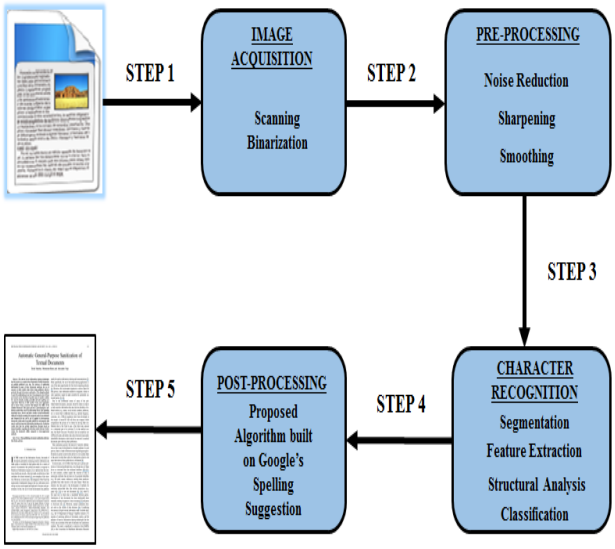

Fig 2. System architecture

4.1 GUI Application:

Three tier Desktop applications consist of presentation logic, business logic and data logic. Presentation logic is where user interface (UI) is developed using which users initiate requests. Business logic is where the validations and web service functionalities are written. Data logic is related with all the database queries generated as a result of web requests.

4.2 Upload Module:

User can use system and upload image to convert text.

4.3 OCR Module:

As per the optical character Recognisation convert the image to text with limited accuracy.

4.4 ALGORITHM PROCESS

Begin
Step 1: Upload scanned image and respective editable file in the system
Step 2: Compare every scanned data with editable file and if not matched then edit it
Step 3: Perform below operations for required output:-

a) Insert line
b) Remove line
c) Uppercase to lowercase conversion and vice versa
d)Font settings
e) Read out loud the text file to compare text file with scanned image
f) Fnd and replace for particular string in the text file
g) Save file

Step 4: End

V. RESULT AND DISCUSION

Fig 3. Input and output result

Input: Here the user can input the image to the given system.
Output: The text output data will generated based on OCR process.

VI. APPLICATION

Optical character recognition has been applied to a number of applications. Some of them have been explained below.

Legal Industry
OCR is used in Legal industry for digitize documents, and directly entered to computer database. Legal professionals can further search documents required from huge databases by simply typing a few keywords [12].

Healthcare
Healthcare professionals always have to deal with large volumes of forms for each patient, including insurance forms as well as general health forms. To keep up with all of this information, it is useful to input relevant data into an electronic database that can be accessed as necessary. Form processing tools, powered by OCR, are able to extract information from forms and put it into databases, so that every patient's data is promptly recorded [12].

Optical Music Recognition
Initially it was aimed towards recognizing printed sheets which can be edited into playable form with the help of electronic methods. It has many applications like processing of different classes of music, large scale digitization of musical data and also it can be used for diversity in musical notation [12].

Automatic Number Recognition
Automatic number plate recognition is used as a technique making use of optical character recognition on images to identify vehicle registration plates. They are used by various police forces and as a method of electronic toll collection on pay-per-use roads and cataloging the movements of traffic or individuals [12].

Handwriting Recognition
It is the ability of a computer system which scans the image of handwritten text by scanner and extracts only handwritten character from that image [13].

VII. CONCLUSION

It gathers data in binary image format, those images are processed for the Optical Character Recognition. The OCR recognizes the text from a binary image, and converts it into machine generated text. After text we apply the voice generation. The proposed system is also able to recognize the printed as well as handwritten text with limited accuracy.

VIII. FUTURE SCOPE

In future scope use the recognized data is sent further for converting it into speech using Android text to Speech. The system reduces human efforts along with time. It can also be helpful for the person who doesn‟t know the language / pronunciation of particular words. The system can also be helpful for visually impaired or person with weak visual ability.

REFERENCES

[1] Champa H N, Dr. K R Ananda Kumar, “Artificial neural network for human behaviour prediction through handwriting analysis”, International Journal of Computer Applications(0975 – 8887)Volume 2-No.2,May 2010.

[2] “Automated Human behavior prediction through handwriting analysis” By Champa H N and K R Anandakumar

[3] S. M. E. Hossain and G. Chetty,”Human Identity Verification by Using Physiological and Behavioral Biometric Traits”, International Journal Bioscience, Biochemistry and Bioinformatics, Vol. 1, No.3 ,September 2011

[4] Esmeralda C Djamal, Sheldy Nur Ramdlan, Jeri Saputra, “Recognition of Handwriting Based on Signature and Digitof Character Using Multiple of Artificial Neural Networks in Personality Identification”, Information Systems International Conference (ISICO), 2 – 4 December 2013

[5] Sandeep Dang ,Prof. Mahesh Kumar, Mahesh, “Handwriting Analysis of Human Behaviour Based on Neural Network”,

ISSN: 2277 128X, Volume 4, Issue 9, September 2014

[6] Javier Galbally, Julian Fierrez, Marcos Martinez-Diaz, R´ejean Plamondon E´cole Polytechnique de Montre´al,"Quality Analysis of Dynamic Signature Based on the Sigma- Lognormal Model", 2011 International Conference on Document Analysis and Recognition

[7] Albert Ali Salah1, Theo Gevers1, Nicu Sebe2, and Alessandro Vinciarelli3," Challenges of Human Behavior Understanding"

[8]Handwriting Personality Profile. http://handwritingpro.com (1998)

[9] S. M. E. Hossain and G. Chetty,"Human Identity Verification by Using Physiological and Behavioral Biometric Traits", International Journal of Bioscience, Biochemistry and Bioinformatics, Vol. 1, No. 3, September 2011

[10] Hugo Gamboaa and Ana Fredb," A Behavioral Biometric System Based on Human Computer Interaction"

[11] Ajzen I, Fishbein M. (1999) Theory of reasoned action/Theory of planned behavior. University of South Florida.

[12] "Survey of OCR Applications" by Amarjot Singh, ketan bacchuwar, Akshay bhasin.

[13] M.D. Ganis, C.L. Wilson, J.L. Blue, "Neural network-based systems for handprint OCR applications" in IEEE Transactions on Image Processing, 1998, Vol: 7, Issue: 8, p.p. 1097 – 1112.

