



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Turning Handwritten Document into Digitized Version

Swapnil R. Aghadte , Vansh R. Vyawahare , Talha Ahmad Khan
Student Student Student

Pranav R. Zade , Tejas P. Kharkar
Student Student

Prof P. N. Umekar
Assistant Professor , PRPCEM

Computer Science And Engineering
P . R . Pote College Of Engineering and Management , Amravati , India

Abstract Use a scanner or a smartphone with a high-quality camera to capture clear images of the handwritten documents. Save the scanned images in a format that supports image-to-text conversion (e.g., JPG, PNG, or PDF). Employ Optical Character Recognition (OCR) software or online tools to convert the scanned images into editable text. Many OCR tools are available online, some free and some paid. Choose one that suits your needs and provides accurate results. After the OCR process, review the converted text for any errors or inaccuracies introduced during the conversion. Manually correct any mistakes to ensure the accuracy of the digitized text. Sort the digitized documents into appropriate folders or categories to facilitate easy access and reference. If the handwritten documents are from external sources, make sure to cite them properly in the project. Keep the digitized documents in a secure location to prevent data loss or unauthorized access. Create backups to ensure the preservation of valuable information.

Keywords: Handwriting recognition, Optical Character Recognition (OCR), Document digitization, Handwritten to digital conversion, Manuscript transcription, Digital transcription tool, Handwriting capture, Pen-to-digital conversion, Document scanning and conversion, Paper-to-digital conversion, Handwritten text extraction, Digital document creation from handwritten input, Automated handwriting transcription, Handwriting digitization service

I. INTRODUCTION

This paper refers to the system which convert the hand written text or a conversation into a machine coded language, text document or scanned images etc. Character recognition and handwritten numbers are becoming more and more significant in today's digital world as their practical uses in numerous daily tasks appear to be growing. As we know the handwritten character recognition is a bit difficult job to do

because of different variation of the writing styles. Systems that read handwritten letters, characters, and numbers enable humans to complete complicated tasks that would otherwise be time consuming and expensive to complete. As we know the handwritten character recognition is a bit difficult job to do because of different variation of the writing styles. Identification of various style of text using different kind of method

is one of the key features in the area of text recognition. As of right now, the technology of handwritten text recognition is becoming a must in modern planet. The system must be robust to improve the extraction and performance of the system. Nowadays the handwritten text recognition has gained a lot of interests of the industries sowing its application in various fields. This system uses the Intelligent Word Recognition (IWR) technology which makes the system less complex and makes the process easier to recognize the meaningful word or sentences. Because of these technologies the digitization of the text has become simpler. The group of neurons helps capturing the pixel of the input image of the text and extracts a meaning full text or digit by using the intelligent word recognition. Further these data are passed on and transformed into a meaningful manner according to the designed network. Again, the different processes are done to extraction/receive the expected output or show that character was read at the input side. User Interface (UI): The UI provides an interactive platform where users can upload images or scans of handwritten documents. It may also include options for adjusting image quality, selecting processing options, and viewing the converted digital documents. OCR Engine: Optical Character Recognition (OCR) is the core technology responsible for recognizing text within images. This component analyzes the scanned documents to identify characters and convert them into editable text. Text Editor: Once the text is extracted from the handwritten documents, a text editor allows users to view and edit the digital versions. It may include features like formatting options, spell check, and saving capabilities. Output Formats: The application may support various output formats such as plain text, PDF, or editable document formats like DOCX. Users should have the flexibility to choose the format that best suits their needs.

II. FUNCTIONAL COMPONENTS

Converting handwritten documents into digital versions via a web application involves several key concepts and components:

Image Processing: The handwritten documents need to be captured or scanned to create digital images. Image processing techniques may then be applied to enhance the quality of the images, correct distortions, remove noise, and improve readability.

Optical Character Recognition (OCR): OCR technology is used to recognize and convert the text contained in the images into editable and searchable digital text. This involves analyzing the image, identifying individual characters, and interpreting them into machine-readable text.

Web Application Interface: A user-friendly web interface is needed to interact with the application. This interface should allow users to upload scanned handwritten documents, monitor the processing progress, and access the converted digital versions .

File Management: The web application needs a mechanism to manage the uploaded handwritten documents and the corresponding digital versions. This may involve storing the files securely, organizing them into a database or file system, and providing methods for retrieval and deletion.

Text Processing and Editing: Once the handwritten documents are converted into digital text, the application may offer features for text processing and editing. This could include spell checking, formatting options, annotation tools, and the ability to save or export the edited documents.

Security: Given the sensitive nature of some documents, security measures should be implemented to protect user data and ensure the confidentiality and integrity of the documents being processed and stored.

Scalability and Performance: As the application may need to handle large volumes of documents and concurrent users, scalability and performance considerations are crucial. This involves designing the application architecture to efficiently process requests and handle increased loads.

Accessibility: Ensuring that the web application is accessible to users with disabilities is essential. This involves following accessibility standards and guidelines to make the application usable by all individuals, regardless of their abilities.

By incorporating these basic concepts into the development of a web application for turning handwritten documents into digital versions, you can create a robust and user-friendly solution that meets the needs of your users

III OBJECTIVES

The objective is to recognize online handwritten documents, which includes characters, words, lines, paragraphs etc. There is extensive work in the field of handwriting recognition, and a number of reviews exists. The main objective of this project is to develop a robust and efficient system that can convert handwritten documents, such as notes, letters, or other text-based content, into a digitized and machine-readable format.

Develop a user-friendly interface for uploading handwritten documents: Create a feature-rich web application interface where users can easily upload their handwritten documents in various formats (e.g., scanned images, PDFs).

Implement Optical Character Recognition (OCR) technology: Integrate OCR functionality into the web application to accurately recognize and convert handwritten text into digital text. This involves choosing and implementing suitable OCR algorithms or leveraging existing OCR libraries.

Ensure accuracy and reliability of OCR results: Focus on enhancing OCR algorithms to handle various handwriting styles, languages, and document layouts to improve accuracy and reliability of the converted digital text.

Support multiple file formats: Enable the web application to accept and process handwritten documents in various file formats such as PDF, JPEG, PNG, TIFF, etc., ensuring compatibility with a wide range of user inputs.

Provide real-time feedback: Implement a feedback mechanism to inform users about the OCR process status and accuracy, allowing them to review and correct any errors in the converted digital text.

Enable text editing capabilities: Incorporate text editing features into the web application, allowing users to make adjustments or corrections to the converted text as needed before finalizing the digital version.

Ensure security and privacy: Implement robust security measures to safeguard user data and ensure privacy during the document conversion process, including encryption of uploaded documents and adherence to data protection regulations.

Optimize performance and scalability: Design the web application architecture to handle high volumes of document uploads and processing requests efficiently, optimizing performance and scalability to accommodate increasing user demand.

Integrate with cloud storage services: Enable seamless integration with popular cloud storage platforms (e.g., Google Drive, Dropbox) to allow users to easily save and access their converted digital documents from anywhere.

Gather user feedback and iterate: Continuously gather feedback from users to identify areas for improvement and iterate on the web application features and functionalities, ensuring it meets the evolving needs and expectations of users

IV. METHODOLOGY

This project proceeds in the following steps to generate the desired output:

1. Data Collection
2. Input
3. Feature Extraction
4. Character Separation
5. Pre-Processing
6. Segmentation
7. Data Base

8. Post Processing

1. Data Collection

To develop a web application for turning handwritten documents into digital versions, you'll need to collect data for various purposes, including training machine learning models, improving user experience, and ensuring accuracy in text recognition. Here's a methodology for data collection:

- Handwritten Document Samples
- Data Annotation
- Continuous Improvement
- Privacy Considerations
- Benchmarkin

2. Input

- Upload an image that we have to convert.
- Check there is any error in the program. If there is no error
- then upload the input image.

3.Feature Extraction

The most intriguing stage is feature extraction. To determine the locations of the black pixels, each row of the image is scanned from left to right and top to bottom. With consideration for the variations in these features that may result from variations in handwriting, the features horizontal line, vertical line, left slant, right slant, higher curve, lower curve, left curve, and right curve are intelligently recognised based on their places. When one feature could be confused for another, it may be prudent to detect the features in a certain order and mark the visited black pixels to prevent confusions

4.Character Separation

By examining a vast number of examples, we were able to conclude that capital letters are typically written with various characters spaced apart. To split a line into individual characters to enable feature extraction and matching easier.

5.Pre-Processing

- Line segmentation check every line by line
- Word segmentation check every word and symbol.

6.Segmentation

- In this step were used to extracting the character through words.
- Segmentation algorithm share out with the handwritten
- image and take-out separate characters from handwritten words.
- This transforms into digitized text through machine learning.

7. Data Base

- Check the words from data base in digital version

8. Post Processing

- Digitized words are from after recognizing Analyze words with semantic meanings.

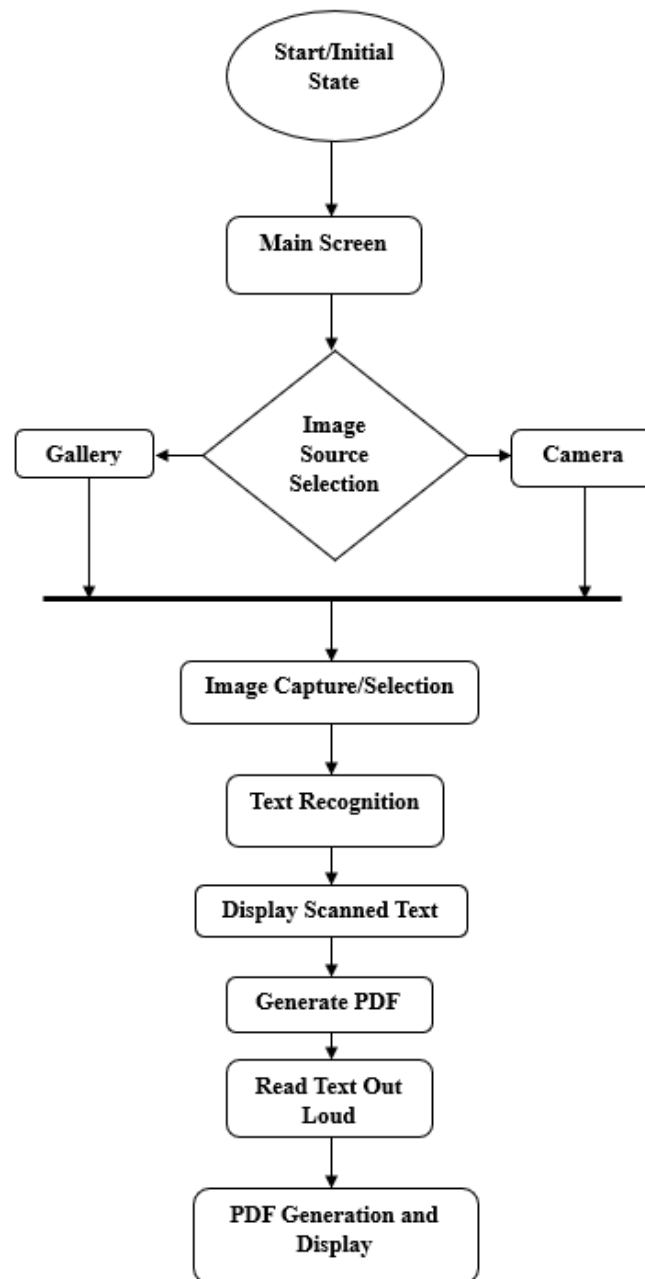


Fig. 01 Workflow of Turning Handwritten Document into Digitized Version

V. ADVANTAGE

Turning handwritten documents into digital versions via a web application offers numerous advantages, including:

- **Accessibility:** Digitizing handwritten documents makes them accessible to a wider audience. Individuals with visual impairments can utilize screen readers to access the content, and the digital format allows for easy resizing and adjusting of font sizes for readability.
- **Searchability:** Digital documents are searchable, enabling users to quickly locate specific information within large volumes of text. This feature significantly improves efficiency and saves time compared to manually scanning through physical documents.
- **Data Analysis and Insights:** Digital documents enable advanced data analysis techniques such as text mining, sentiment analysis, and machine learning. By extracting insights from the textual content, organizations can gain valuable business intelligence and make data driven decisions.
- **Environmental Benefits:** Digitizing handwritten documents reduces paper consumption and contributes to environmental conservation efforts. By adopting digital document solutions, organizations can minimize their carbon footprint and promote sustainability.

• **Remote Access:** With a web-based digital document solution, users can access documents from anywhere with an internet connection, facilitating remote work and collaboration among distributed teams.

VI. LIMITATIONS

While turning handwritten documents into digital versions via a web application offers numerous advantages, there are also several limitations to consider:

- **Variability in Handwriting Styles:** Handwriting can vary significantly from person to person, making it challenging for OCR systems to accurately recognize and transcribe all handwriting styles. As a result, some handwritten documents may not be effectively digitized, leading to incomplete or inaccurate conversions.
- **Language and Character Set Limitations:** OCR systems may have limitations in recognizing languages other than the ones they are trained on, as well as specific characters, symbols, or writing systems. This can pose challenges for digitizing handwritten documents in languages or scripts that are not well-supported by the OCR technology.
- **Privacy and Security Concerns:** Digitizing handwritten documents raises privacy and security concerns, particularly regarding the protection of sensitive or confidential information contained in the documents. Web-based applications must implement robust security measures to safeguard user data and prevent unauthorized access or data breaches.
- **Cost of Implementation and Maintenance:** Implementing and maintaining a web-based digital document solution involves costs associated with software development, infrastructure setup, licensing fees for OCR technology, ongoing maintenance, and support. Organizations must carefully assess the financial implications and return on investment (ROI) of adopting such solutions.

VII. CONCLUSION

The main perspective of this project is to recognition of handwritten text-lines with the better accuracy. The main focus of this investigation is to improve the system to identify any other character and tried to make our system more font independent that our system can read or recognize any other font. An Intelligent Word Recognition (IWR) and the segmentation algorithm techniques has obtained good results

VIII. REFERENCES

1. D. Kavitha, P. Shamini, "Handwritten Document into Digitized Text Into Using Segmentation Algorithm". Department of Computer Application, Easwari Engineering College, Tamil nadu (ISSN2014)
2. Jamshed Memon, Maira Sami, Rizwan Ahmed Khan. "Handwritten Optical Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)". Faculty of IT, Barrett Hodgson University1 January, 2020.
3. S M Shamim, Mohammad Badrul Alam Miah, Angona Sarker. "Handwritten Digit Recognition Using Machine Learning Algorithms". Global Journal of Computer Science and Technology.
4. Edgard Chammas & Chafic Mokble. "Handwriting Recognition of Historical Documents with few Lebeled Data". University of balamand El-koura, Lebanon
5. Yellapragada SS Bharadwaj, Rajaram P, Sriram V.P, "Effective Handwritten Digit Recognition using Deep Convolution Neural Network." Dept. of Ckoneru lakshmaish education foundation, India.
6. Caiyun Ma & Hong Zhang, "Effective Handwritten Digit Recognition Based on Multi- feature Extraction and Deep Analysis." College of computer scienceand technology, Wuhan
7. Bhowmik S, Roushan M. G, Sarkar R, Nasipuri M, Polley S. & Malakar "Handwritten Bangla Word Recognition Using HOG Descriptor." In 2014 Fourth International Conference of Emerging Applications of Information Technology (EAIT). IEEE, 2014.
8. Kruse R, Borgelt C, Klawonn F, Moewes C, Steinbrecher M & Held P. MultilayerPerceptrons. In Computational Intelligence (pp. 47-81). Springer London 2013.

9. Shubham Sanjay Mor, Shivam Solanki, Saransh Gupta, Sayam Dhingra, Monika Jain, Rahul Saxena, Handwritten Text Recognition : with Deep Learning.
10. Ahmed Mahi Obaid, Hazem M EI Bakry, M. A. Eldosuky, A.I. Shehab, Handwritten Text Recognition Based on Neural Network, IJARCSST, Vol. 4, Issue 1, 2016.
11. A. N. S. Chakravarthy, Penmesta V Krishna raja, P. S Avahani, Handwritten Text Image Authentication IJNSA, Vol. 3, No. 5, 2011
12. R. S. Sabeenian M. Vidhya, Handwritten Text to Digital Text Conversion using Radon Transform and Back Propagation Network (RTBPN), International Conference on Advances in Information .
13. Amandeep Kaur, Seema Bhagla, Sunil Kumar, Study of Various Character Segmentation Techniques for Handwritten off-Line Cursive Words, IJASEAT, 2015.
14. Handwritten Off-Line Cursive Words, IJASEAT, 2015
15. A. N. S. Chakravarthy, Penmesta V Krishna raja, P. S Avahani, Handwritten Text Image Authentication using Back Propagation, IJNSA, Vol. 3, No. 5, 2011.

