IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635

Real World Uses of OCR for Modernization of Traffic system

Dr. Saurabh A Ghogare[1], Prof. Abhijit G Kalbande[2]

Assistant Professor, Adarsha Science J.B.Arts & Birla commerce Mahavidyalaya Dhamangaon Rly, India.[1]

Assistant Professor, Prof Ram Meghe College of Engineering & Management, Badnera-Amravati, India[2]

Abstract: Normally complete OCR methodology for recognizing documents, either printed or handwritten without any knowledge of the font, is presented. This methodology consists of three steps: The first two steps refer to creating a database for training using a set of documents, while the third one refers to recognition of new document images. First, a preprocessing step that includes image binarization and enhancement takes place. At a second step a top-down segmentation approach is used in order to detect text lines, words and characters. A clustering scheme is then adopted in order to group characters of similar shape. This is a semi-automatic procedure since the user is able to interact at any time in order to correct possible errors of clustering and assign an ASCII label. After this step, a database is created in order to be used for recognition. Finally, in the third step, for every new document image the above segmentation approach takes place while the recognition is based on the character database that has been produced at the previous step.

Keywords: :- Digital conversion, Handwritten or printed text, Image conversion.

I. INTRODUCTION

Optical Character Recognition (OCR) is the mechanical or electronics conversion of images of typewritten or printed text into Binary. Optical character recognition usually abbreviated to OCR, involves a computer system designed to translate images of typewritten text usually captured by a scanner into machine editable text or to translate pictures of characters into a standard encoding scheme representing them. OCR began as a field of research in artificial intelligence and computational vision.

OCR technology has been applied throughout the entire spectrum of industries, revolutionizing the document management process. OCR has enabled scanned documents to become more than just image files, turning into fully searchable documents with text content that is recognized by computers. With the help of OCR, people no longer need to manually retype important documents when entering them into electronic database. Instead, OCR extract relevant information and enters it automatically. The result is accurate, efficient information processing in less time. There are many recognition techniques are in use these are as follows,

1) Optical mark recognition (OMR) – The OMR technology recognize simple checkmarks, grouped checkmarks, model checkmarks and checkmarks with correction made by hand in different variations like checkmarks in square frame, checkmarks against the empty background and non-standard checkmarks type. OMR delivers accuracy rate of 99.99%.
2) Optical barcode recognition (OBR) - The OBR technique recognize 1D and 2D barcode type. This feature enables automated detection and recognition of barcodes at any angle on a document.
3) Magnetic ink code recognition (MICR) - At bottom of the bank cheque, magnetic ink code is printed. Magnetic ink consists of magnetic particles which is useful for information extraction. After recognition of magnetic ink code we get information about the cheque as account holder name, account number, cheque serial number etc.
4) Optical music recognition (OMR) – OMR recognize printed sheet which can be edited into playable form with the help of electronic method. An OMR system has many applications like processing different classes of musical data large sc, ale of digitization of musical data and also it can be used for diversity in musical notation. There are many character recognition techniques, but OCR is the best technique because it real time converts handwritten or printed text into digital form.

Types of character and word recognition

1) Optical character recognition (OCR)
2) Intelligence character recognition (ICR)
3) Optical word recognition (OWR)
4) Intelligence word recognition (IWR)

In optical character recognition and optical word recognition, the characters are recognizing with the help of archives and database of system. Database contains reference images of characters and alphabets. In this, two methods, one character is recognizing at a time.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635

II. COUNSTRUCTION & WORKING PRINCIPAL

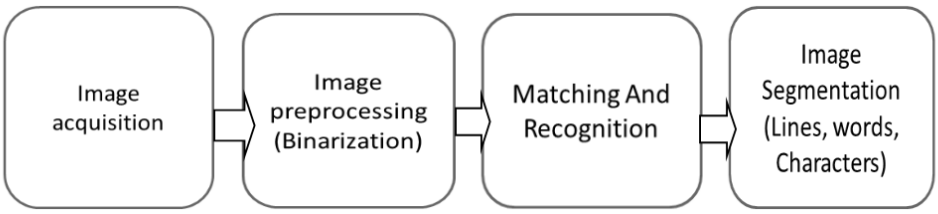


Fig. 1 Steps involved in word, character recognition

Image processing and pattern recognition plays significant role in handwritten character recognition. There are various types of classification of feature extraction methods like statistical feature-based methods, structural feature-based methods and global transformation techniques. Statistical methods are based on the planning of how the data should be selected. It uses the information of statistical distribution of pixels in image, it can be mainly classified in three categories: 1) Partitioning in regions, 2) Profile generation and projections 3) distances and crossing. Structural features are extracted from structure and geometry of character like number of horizontal and vertical lines, aspect ratio, number of cross points, number of loops, number of branch points, number of strokes and number of curves etc.

1) Image acquisition

2) Image pre-processing (Binarization)

3) Image segmentation

4) Matching and recognition

The image has been captured using a camera or digital scanner. The flap of the scanner had been kept open during the acquisition process in order to obtain a uniform black background. The image had been acquired using the portable camera with good quality. The configuration of the Image has been done with the help of Image create subvi function. The configuration of the image means selecting the image type and border size of the image as per the requirement. In this work 8 bit image with border size of 3 has been used.

Binarization is the process of converting a grayscale image (0 to 255 pixel values) into binary image (0 to1 pixel values) by using a threshold value. The pixels lighter than the threshold are turned to white and the remainder to black pixels. In this work, a global thresholding with a threshold value of has been used to binarize the image i.e. the values of pixel which are from 175 to 255has been converted to 1 while the of pixel which have gray scale value less than 175 have been converted to 0.

The segmentation process consists of line segmentation, word segmentation and finally character segmentation.

Line segmentation is the first step of the segmentation process. It takes the array of the image as an input and scans the image horizontally to find first ON pixel. The system continues to scan the image horizontally and found lots of ON pixel since the characters would have started. When finally first OFF pixel has been detected the system and check the surrounding of the pixel to find out required number of OFF

pixels. If this happens then the system clips the first line from input image. In this way, all the lines have been segmented & stored to be used for word and character segmentation.

In the word segmentation process the line segmented images have been vertically scanned to find first ON pixel. When this happen the system remember the coordinate. This is the starting coordinate for the word. The system continues the scanning process until successive OFF pixels have been obtained. The system records the first OFF pixel. In this way all the words have been segmented and these segmented words have been used in next step for character recognition.

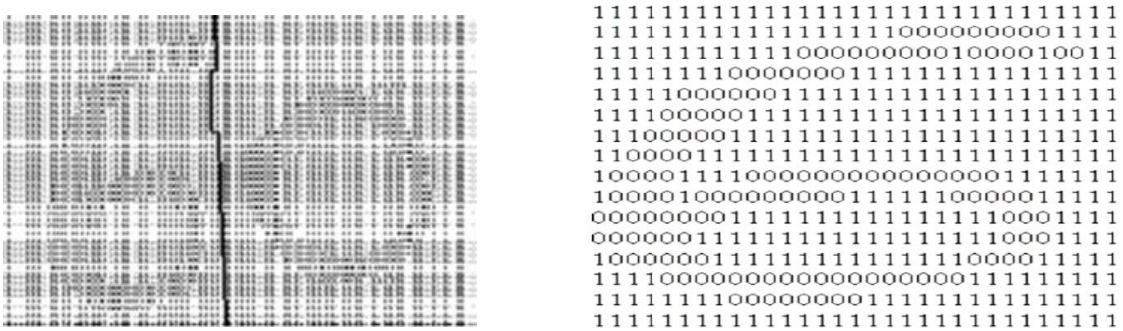
Character segmentation has been performed by scanning the word segmented image vertically. This process is different from the word segmentation in following two ways - Number of horizontal OFF pixels between the different characters are less in comparison to number of OFF pixels between the words and Total number of

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635



(a) Compact character (b) Resize character

Fig.2 Character segmentation

3.5 Matching and recognition

In this process, correlation between stored templates and segmented character has been obtained by using correlation. The correlation determines the correlation between segmented character and stored templates of each character. The value of the

highest correlation recognizes a particular character. In this way in order to recognize the character every pre-defined data stored in the system. Since same font size has been used for recognition, a unique match for the each character has been obtained. characters and their order in the word has been determined so as to reproduced.

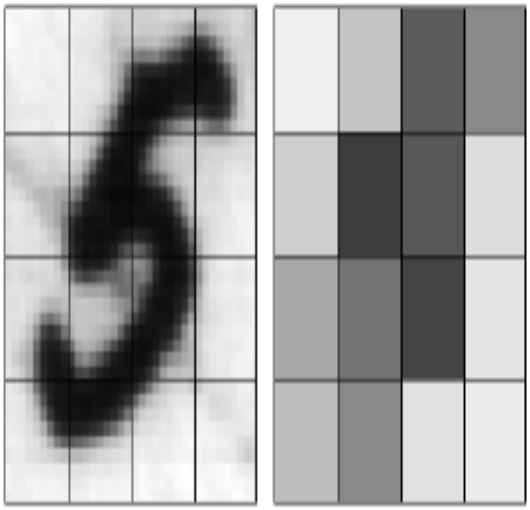


Fig.3 Zoning features of gray image

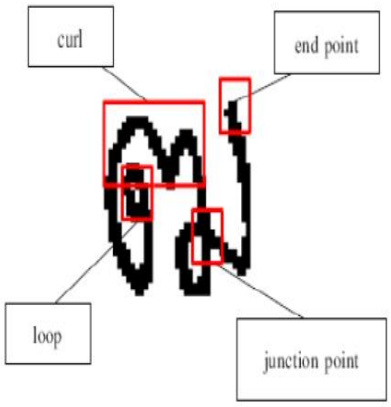


Fig.4 Structural features extraction

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635

III. APPLICATION

A. Text and product label reading

To identify the text from the input image first the input image is segregated into frames and each frame is converted into Gray image and then into binary image. Feature extraction is applied on the binary image to localize the text from the background. The output is given to the OCR.

B. OCR documents scanner

Optical character recognition (OCR) software works with scanner to converts printed characters into digital text. We can remove unwanted text from scanned characters which is converted into digital one. If we lost or erase important text from digital data we can easily recover the lost data by scanning the hard copy of printed document. One of the most important feature of this application is that we can converts the printed data into any type of digital documents.

Fig. 5 flatbed optical scanner

C. Automatic cheque verification

cheque without human involvement. Cheque can be inserted into machine the writing on it is scanned instantly and correct amount of money is transferred. This technology fairly accurate for handwritten cheque through occasionally requires manual confirmation One widely known OCR application is in banking, where OCR is used to processed. Overall, this reduces wait times in many banks.

Fig. 6 Automatic bank passbook and cheque verification machine

D. Information extraction

In school, collages, hospitals and government as well as private offices large amount of documents are gathered. Due to large amount of documentation, it is very complicated to sort out the required information such as name, address, date

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635

of birth etc. With the help of OCR we can easily extract the information from document. This can be save time and human resource.

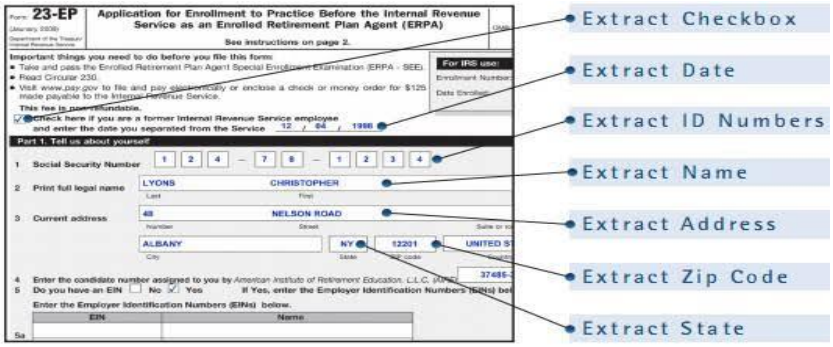


Fig. 7 Extraction of information on form using OCR

E. Barcode recognition

A barcode is an optical machine-readable representation of data relating to the object to which it is attached. Barcode can be defined as symbology that automates the storage and retrieval of data without any involvement of human. Barcode technology is considered as one of the most important part of Automatic Identification and Data Capture (AIDC) [1]. The two dimensional (2D) barcodes have been developed to enhance the capability of the traditional one dimensional (1D) barcode, especially in data capacity. In the Two-dimensional (2D) symbols, data are encoded in both the height and width of the symbol, and the amount of data that can be contained in a single symbol is significantly greater than that stored in a one dimensional symbol. Over thousands alphanumeric characters can potentially be placed in a single symbol the size of a large postage stamp [2]. Obviously, the main advantage of using 2D bar codes is that possibly a large amount of easily- and accurately-read data can "ride" with the item to which it is attached. There are new applications being created for 2D bar code technology.

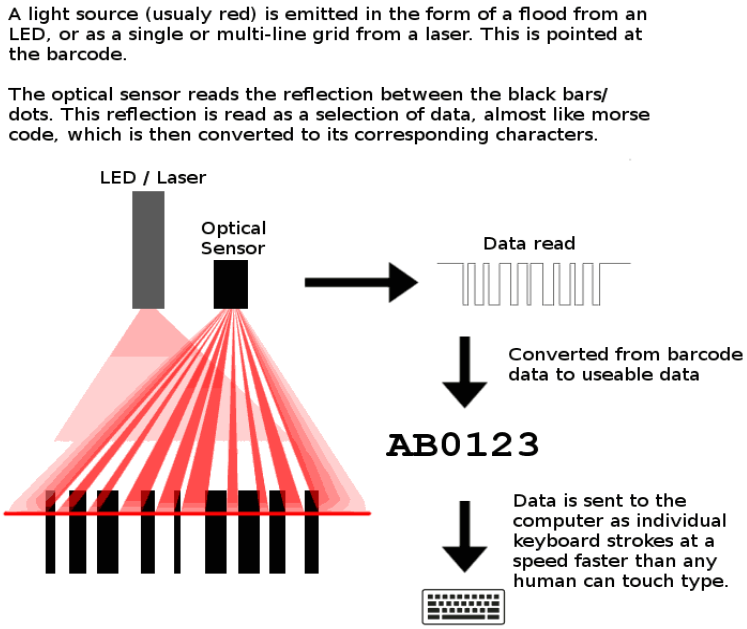


Fig. 8 Barcode recognition

F. Automatic number plate recognition

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

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 6, June 2025

DOI: 10.17148/IJIREEICE.2025.13635

Fig. 9 Automatic number plate detection

G. Optical music recognition

Optical music recognition (OMR) recognize printed sheets which can be edited into playable form with the help of electronic and electrochemical methods. An OMR system 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.

H. Invoice imaging

It is widely used in many businesses applications to keep track of financial records and prevent a backlog of payments [100] from piling up. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes. As the technology continues to develop, more and more applications are found for OCR technology, including increased use of handwriting recognition. Furthermore, other technologies related to OCR, such as barcode recognition, are used daily in retail and other industries.

I. Legal Industry

In the legal industry, there has also been a significant movement to digitize paper documents. In order to save space and eliminate the need to sift through boxes of paper files, documents are being scanned and entered into computer databases. OCR further simplifies the process by making documents text-searchable, so that they are easier to locate and work with once in the database. Legal professionals now have fast, easy access to a huge library of documents in electronic format, which they can find simply by typing in a few keywords.

J. Healthcare

It Healthcare has also seen an increase in the use of OCR technology to process paperwork. 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. As a result, healthcare providers can focus on delivering the best possible service to every patient.

K. Captcha

CAPTCHA is a program that can generate and grade tests that human can pass but current computers programmers' cannot. Hacking is a serious threat to internet usage. Now a day's most of the human activities like economic transactions, admission for education, registrations, travel bookings etc are carried out through and all this requires a password which is misused by hackers. They create programs to like dictionary attacks and automatic false enrolments which lead to waste of memory and resources of website. Dictionary attack is attack against password authenticated systems where a hacker writes a program to repeatedly try different passwords like from a dictionary of most common passwords. In CAPTCHA, an image consisting of series of letters of number is generated which is obscured by image distortion techniques, size and font variation, distracting backgrounds, random segments, highlights and noise in the image. This system can be used to remove this noise and segment the image to make the image tractable for the OCR (Optical Character Recognition) systems.

DOI: 10.17148/IJIREEICE.2025.13635

L. Institutional Repositories and Digital Libraries

Institutional repositories are digital collections of the outputs created within a university or research institution. It is an online locale of intellectual data of an institution, especially a research institution where it is collected, preserved and aired. It helps to open up the outputs of an institution and give it visibility and more impact on worldwide level. Enables and encourages interdisciplinary approaches to research and facilitates the development [103] and sharing of digital teaching materials and aids. It is basically a collection of peer reviewed journal articles, conference proceedings, research data, monographs, books, theses and dissertations and presentations. Their first role is to provide the Open Access literature. Practical implementation of this includes setting up a system which consists of scanner which scans the documents. This scanned document is then fed as an input to an Optical Character Recognition system where information is acquired and retained in digitized form.

IV. ADVANTAGES & DISADVANTAGES

ADVANTAGES

1. Eliminating manual data entry, a process that drains profit and productivity.
2. Improving data accuracy by reducing human error, which can occur while the data is being input.
3. Improve customer service.
4. Aid compliance with secure, controlled access to sensitive documents.
5. Importing application and form submissions.

DISADVANTAGES

1. If image quality is low, recognition of characters should not take place properly.
2. If Input image to the system is not clear and dark, then errors will be occurred.

V. CONCLUSION

Optical character recognition technique is the most efficient. This technique efficiently converted handwritten or printed text into digital data which is advantageous. Data transformation takes place with minimum errors. Data stored more compactly and less documentation. Today optical character recognition is most successful for constrained material, that is documents produced under some control. However, in the future it seems that the need for constrained OCR will be decreasing. The reason for this is that control of the production process usually means that the document is produced from material already stored on a computer. Hence, if a computer readable version is already available, this means that data may be exchanged electronically or printed in a more computer readable form, for instance barcodes.

REFERENCES

[1] Ali Mir Arif Mir Asif 1, Shaikh Abdul Hannan2, Yusuf Perwej3, Mane Arjun Vithalrao4 International Journal of Advance Research in Science and Engineering http://www.ijarse.com IJARSE, Vol. No.3, Issue No.7, July 2014 ISSN-2319-8354(E)

[2] B.M. Sagar et al., “OCR for printed Kannada text to machine editable format using database approach,” WSEAS Trans. on Computers, vol. 7, pp. 766-769, 2008.

[3] N. Arica and F. Yarman-Vural, An Overview of Character Recognition Focused on Offline Handwriting, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 31, No.2, pp. 216--233, 2001.

[4] Applications of OCR System to Processing and Digitization of Paper Gaur Priyanka, and Shamik Tiwari. "Barcode Localization using Curvelet Transform and Neural Network." International Journal of Computer Applications. 85(6):6-9, January 2014. Documents.

[5] Shrivakshan, G. T., and C. Chandrasekar. "Comparison of various Edge Detection Techniques used in Image Processing." International Journal of Computer Science Issues (IJCSI) 9.5 (2012).

[6] Nair, R. R., Kota, B. U., Nwogu, I., & Govindaraju, V. (2016, December). Segmentation of highly unstructured handwritten documents using a neural network technique. In Pattern Recognition (ICPR), 2016 23rd International Conference on (pp. 1291-1296).

[7] Singh, G., & Sachan, M. (2014, December). Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition. In Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on (pp. 15).

DOI: 10.17148/IJIREEICE.2025.13635

[8] Wikipedia: The free encyclopedia. (2004, July 22). FL: Wikimedia Foundation, Inc. Retrieved August 10, 2004, from https://www.wikipedia.org

[9] N. Arica and F. Yarman-Vural, An Overview of Character Recognition Focused on Offline Handwriting, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 31, No.2, pp. 216--233, 2001.

[10] C. C. Tappert, Cursive Script Recognition by Elastic Matching, IBM Journal of Research and Development, vol. 26, no. 6, pp.765-771, 1982.

BIOGRAPHY

Dr. Saurabh A. Ghogare Working as An Assistant Professor at Adarsha Science, J.B. Arts and Birla Commerce Mahavidyalaya, Dhamangaon Rly. He has more than five years of teaching and academic experience. He published three research papers in reputed journals and authored four books in my subject area. His work reflects a deep commitment to education, research, and student development

Prof. Abhijit G Kalbande Working as An Assistant Professor in Department of Information Technology at Prof Ram Meghe College of Engineering & Management, Badnera. He Has an 11 Years of experience in Teaching & Research. He is working on Internet of things, Image Processing Domain. He Publish 35 Research Paper in National & International Journal and 5 papers National & International Conference.