TEXT RECOGNITION (OCR) USING DIGITAL IMAGE PROCESSING TECHNIQUES

PROJECT REPORT

Submitted in fulfilment for the JComponent of

Digital Image Processing (SWE1010)

CAL Course

In

M.Tech. – Software Engineering

By

K.KAVYA- 16MIS0084

Under the guidance of

Dr. Hemalatha S

SITE

School of Information Technology and Engineering

Winter Semester 2018-19

TABLE OF CONTENTS

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	3
1.	Problem statement	4
2.	System design	4
3.	Software Requirement Specifications	5
4.	Implementation Details	5
	4.1 source code	
5	RESULTS	6
	5.1 Screenshots	
6	Base reference paper (one or two)	7
7	References	8

ABSTRACT:

The objective of text recognition is to recognize the text from printed hardcopy document to preferred format. recognizing a text is a very easiest job for humans, but in a machine or electric device that does text recognition is a challenging task. the major steps involved in text recognition is pre-processing, segmentation, feature extraction, classification. to attain high speed in data processing it is essential to convert the analog data into digital data. The process of Text Recognition involves several steps including preprocessing, segmentation, feature extraction, classification, post processing. Preprocessing is for done the basic operation on input image like binarization which convert gray Scale image into Binary Image, noise reduction which remove the noisy nature from image. Segmentation stage for segment the given image into line by line and segment each character from segmented line. Future extraction calculates the characteristics of character. A classification contains the database and does the comparison. storage of hard copy of any document occupies large space and retrieving of information from that document is time consuming.

INTRODUCTION:

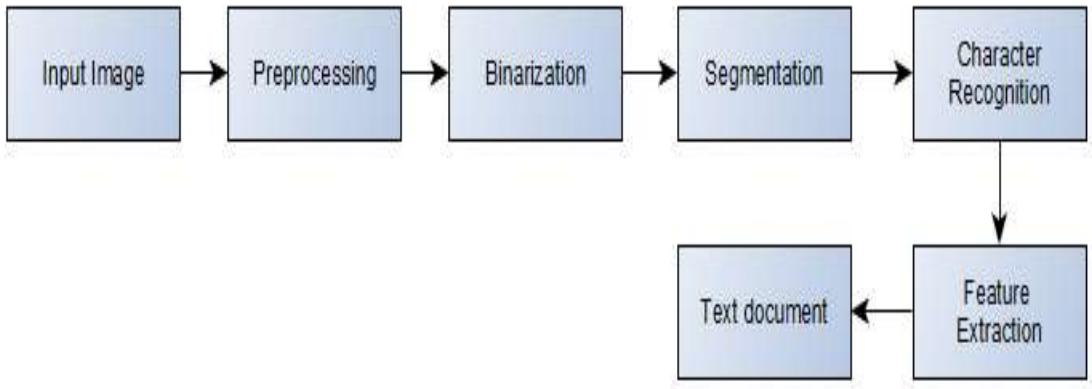
Nowadays all over digitization technology is used. Text Recognition usually abbreviated to OCR, involves a computer system designed to translate images of typewritten text (usually captured by a scanner) into machine editable text or to translate pictures of characters into a standard encoding scheme representing them. OCR began as a field of research in artificial intelligence and computational vision . 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. People wish to scan in a document and have the text of that document available in a .txt or .docx format. in this project will recognize text that are in images using an open source tool called **tesseract** and opencv. the method of extracting text from images is also called optical character recognition (**ocr**)

or sometimes simply text recognition. Optical character recognition system is an effective way in recognition of text. OCR affords the easiest way to recognize and convert the printed text on image into the editable text. It also increases the speed of data reclamation from the image.There are numerous applications in which text extraction is useful. Latest technology in the field of image processing express a great amount of interest in content retrieval from images and videos. Character recognition is an art of detecting segmenting and identifying characters from image. More precisely Character recognition is process of detecting and recognizing characters from input image and converts it into ASCII or other equivalent machine editable form [1], [2], [3].

PROBLEM STATEMENT:

OCR technology has been used to convert the text in scanned paper documents into ASCII symbols .However, current commercial OCR systems do not work well if text is printed against shaded or hatched backgrounds, often found in documents such as photographs, maps, monetary documents ,engineering drawings and commercial advertisements .Furthermore ,these documents are usually scanned in greyscale or color to preserve details of the graphics and pictures which often exist along with the text. For current OCR systems, these scanned images need to be binarized before actual character segmentation and recognition can be done. A typical OCR system does the Binarization to separate text from the backgrounds by global thresholding .Unfortunately, global thresholding is usually not possible for complicated images, as noted by many researches. Consequently, current OCR systems work poorly in these cases. One solution to the global thresholding problem is to use different thresholds for different local regions(adaptive thresholding)

SYSTEM DESIGN



3.SOFTWARE REQUIREMENT SPECIFICATIONS

PYTHON –ANACONDA NAVIGATOR

SPYDER:

Spyder is a powerful scientific environment written in Python, for Python, and designed by and for scientists, engineers and data analysts. It features a unique combination of the advanced editing, analysis, debugging and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection and beautiful visualization capabilities of a scientific package.n Furthermore, Spyder offers built-in integration with many popular scientific packages, including NumPy, SciPy, Pandas, IPython, QtConsole, Matplotlib, SymPy, and more.n Beyond its many built-in features, Spyder can be extended even further via third-party plugins.n Spyder can also be used as a PyQt5 extension library, allowing you to build upon its functionality and embed its components, such as the interactive console or advanced editor, in your own software.

4.IMPLEMENTATION DETAILS

4.1 SOURCE CODE

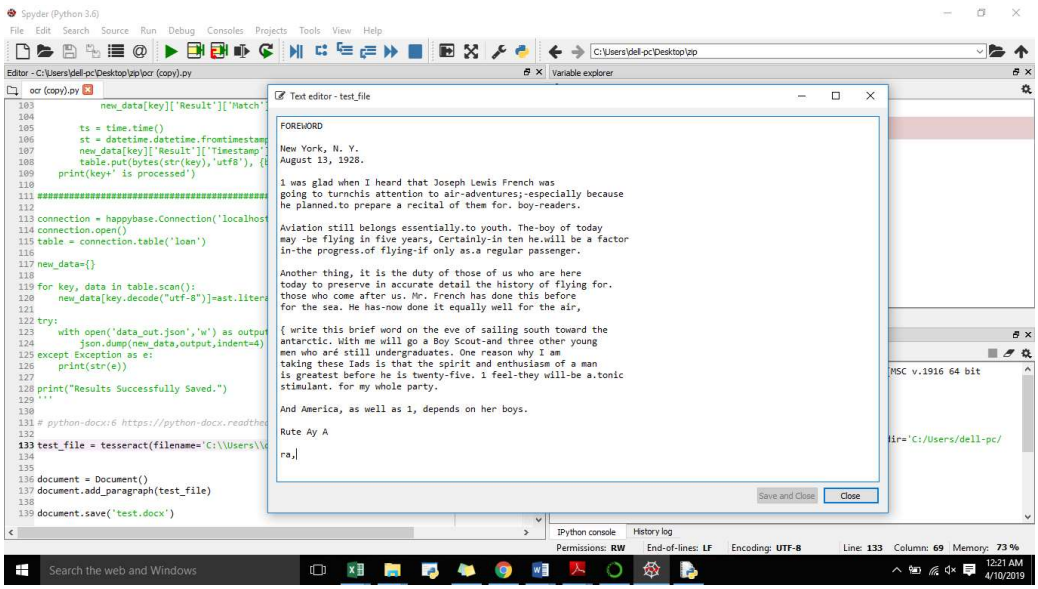
```
import pytesseract
import cv2
import os
import numpy as np
from docx import Document
from PIL import Image,ImageEnhance
def tesseract(filename):
im = Image.open(filename)
rgb_im = im.convert('RGB')
rgb_im.save("test.jpg", dpi=(300,300))
image_dpi = cv2.imread('test.jpg',0)
os.remove("test.jpg")


blur = cv2.bilateralFilter(image_dpi,15,75,75)
```

```
th3=cv2.adaptiveThreshold(blur,255,cv2.ADAPTIVE_THRESH_GAUSSIAN_C,cv
2.THRESH_BINARY,11,2)

ret3,th3cv2.threshold
(th3,0,255,cv2.ADAPTIVE_THRESH_GAUSSIAN_C+cv2.THRESH_OTSU)

imagem = cv2.bitwise_not(th3)

  kernel = np.ones((1,1),np.uint8)

  eroded_img = cv2.erode(imagem,kernel,iterations = 1)

    #canny=cv2.Canny(imagem,100,200)

     config = ('-l eng --oem 1 --psm 3')

 text = pytesseract.image_to_string(eroded_img, config=config)

 return text

test_file = tesseract(filename='C:\\Users\\dell-pc\\Desktop\\scan4.png')

document = Document()

document.add_paragraph(test_file)

document.save('test.docx')
```

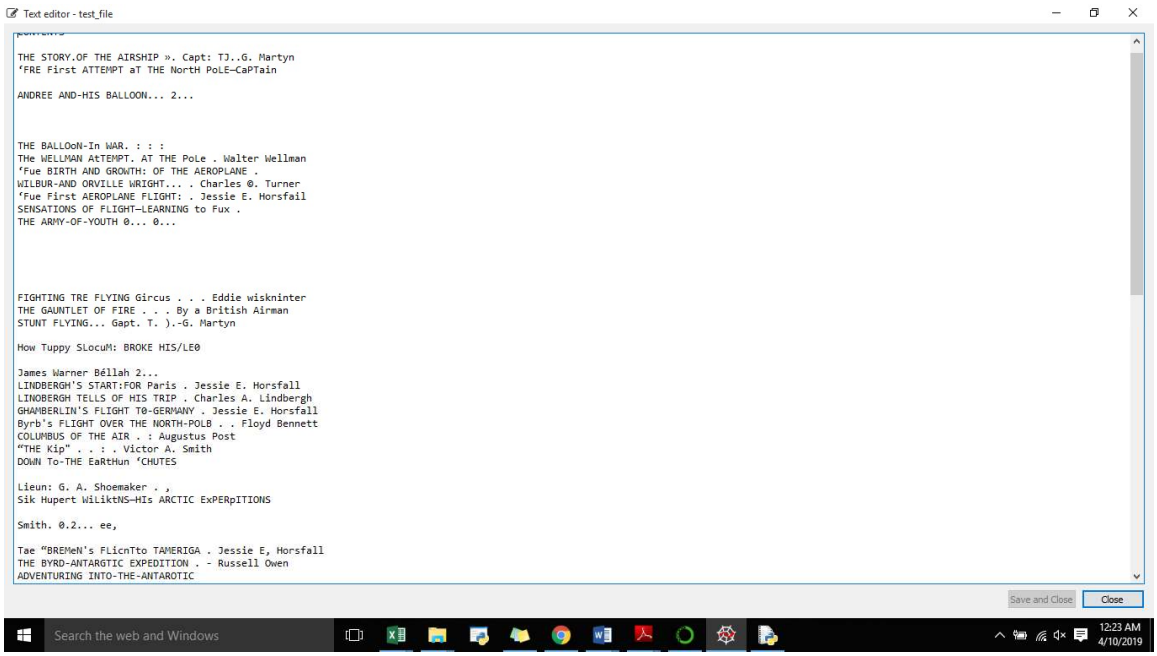
5.RESULTS

5.1 Screenshots

Scan1.jpg



Scan2.jpg



6.BASE REFERENCE PAPER

http://www.ijcea.com/wp-content/uploads/2018/02/NCDCM_2017_paper_55.pdf

7. REFERENCES

[1] Kai Ding, Zhibin Liu, Lianwen Jin, Xinghua Zhu, "A Comparative study of GABOR feature and gradient feature for handwritten 17hinese character recognition", International Conference on Wavelet Analysis and Pattern Recognition, pp. 1182-1186, Beijing, China, 2-4 Nov. 2007

[2] Pranob K Charles, V.Harish, M.Swathi, CH. Deepthi, "A Review on the Various Techniques used for Optical Character Recognition",International Journal of Engineering Research and Applications, Vol. 2, Issue 1, pp. 659-662, Jan-Feb 2012

[3] Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering, Vol.2, Issue 2, pp. 504-58, May 2012

[4]Rashid, S. F., Shafait, F., & Breuel, T. M. (2012, March). Scanning neural network for text line recognition. In *2012 10th IAPR International Workshop on Document Analysis Systems*(pp. 105-109). IEEE.[4]

[5]. Wick C, Reul C, Puppe F. Improving OCR Accuracy on Early Printed Books using Deep Convolutional Networks. arXiv preprint arXiv:1802.10033. 2018 Feb 27.

[6] Breuel, T. M. (2017, November). High performance text recognition using a hybrid convolutional-lstm implementation. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 11-16). IEEE.

[7] Cheng, Z., Bai, F., Xu, Y., Zheng, G., Pu, S., & Zhou, S. (2017). Focusing attention: Towards accurate text recognition in natural images. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5076-5084).

[8] Sankaran, N., & Jawahar, C. V. (2012, November). Recognition of printed Devanagari text using BLSTM Neural Network. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)* (pp. 322-325). IEEE.[8]

[9] Ray, A., Rajeswar, S., & Chaudhury, S. (2015, January). Text recognition using deep blstm networks. In *2015 eighth international conference on advances in pattern recognition (ICAPR)* (pp. 1-6). IEEE.[9]

[10]. Su, B., & Lu, S. (2017). Accurate recognition of words in scenes without character segmentation using recurrent neural network. *Pattern Recognition*, *63*, 397-405.

International Journal of Computer Engineering and Applications,
Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321-3469

TEXT RECOGNITION USING DIGITAL IMAGE PROCESSING TECHNIQUES

S.Banu Chitra[1], Dr R.S.Vetrivel[2]

1. *Research Scholar, Mother Teresa Women's University, Kodaikanal.*
2. *Principal, NPR Arts & Science College,Natham*

ABSTRACT

The objective of Text Recognition is to recognize the text from printed hardcopy document to preferred format. Recognizing a text is a very easiest job for humans, but in a machine or electric device that does text recognition is a challenging task. The major steps involved in text Recognition is pre-processing, segmentation, feature extraction, classification. To attain high speed in data processing it is essential to convert the analog data into digital data. Storage of hard copy of any document occupies large space and retrieving of information from that document is time consuming. Optical character recognition system is an effective way in recognition of text. OCR affords the easiest way to recognize and convert the printed text on image into the editable text. It also increases the speed of data reclamation from the image

Keywords: *Character Recognition System (OCR), Digital Image Processing, Image Segmentation, Feature Extraction.*

[1] INTRODUCTION

The role of text detection is to extract the relevant text data from a collection of images. The recent technology is controlled to extracting text against clean backgrounds. Thus, there is a need for a system to extract text from the backgrounds. There are numerous applications in which text extraction is useful. Latest technology in the field of image processing express a great amount of

interest in content retrieval from images and videos. Character recognition is an art of detecting segmenting and identifying characters from image. More precisely Character recognition is process of detecting and recognizing characters from input image and converts it into ASCII or other equivalent machine editable form [1], [2], [3].

[2] HISTORY OF OCR

The objective of Optical Character Recognition (OCR) is a process of converting scanned document into text document so it can be easily edited. Recognition engine of the OCR system interpret the scanned images and turn images of handwritten or printed characters into ASCII data (Machine readable characters). A document is first scanned by an optical scanner, which produces an image form of it that is not editable. Optical character recognition involves. Translation of this text image into editable character codes such as ASCII [4].

2.1 First generation OCR

First generation OCR systems appeared in the period from 1960 to 1965. IBM 1418 was the first commercialized OC, which was designed to read a special IBM font407. The recognition method was template matching, which compares the character image with a library of prototype images for each character of each font [5]. The number of fonts were limited by the pattern recognition method applied, template matching, which compares the character image with a library of prototype images for each character of each font.

2.2 Second generation OCR

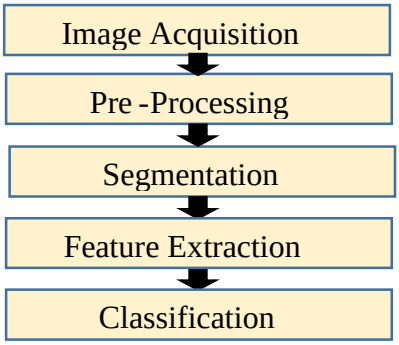
Second generation OCR systems appeared in the middle of the 1960's and early 1970's. This generation machines were able to recognize regular machine printed characters and also had hand-printed character recognition capabilities. When hand-printed characters were considered, the character set was constrained to numerals and a few letters and symbols. Such machines appeared in the middle of 1960s to early 1970s [5].

2.3 Third generation OCR

Third generation of OCR systems, appeared in the middle of the 1970's. In this generation, the challenge was documents of poor quality and large printed and hand-written character sets. Low cost and high performance were also important objectives.

[3] PROCESS OF OCR

The following steps are performed by character recognition system. i.e. i) Image acquisition, ii) Pre-processing, iii) Segmentation, iv) Feature extraction and v) Classification.



3.1 Image Acquisition

Images of OCR system may be attained by scanning document or by capturing photograph of document. This is also known as digitization process [6].

3.2 Pre-processing

Pre-processing operations is used to enhance an image and make it suitable for segmentation process. Quality of handwritten documents depends on several factors including quality of paper, aging of documents, quality of pen, colour of ink etc. Due to external disturbance, any noise is defined as any degradation in the image. Some examples of noises are Gaussian noise, salt and pepper noise. Using filtering technique, these noises can be removed. After the preprocessing is completed, the noise free image is passed to the segmentation phase.

3.3 Segmentation

The process of dividing an image into multiple parts is called Image segmentation. It is one of the decision processes in a system for optical character recognition (OCR). Its decision that a pattern isolated from the image is that of a character. The document is processed in a hierarchical manner. At first level lines are segmented using row histogram, in each row, words are extracted using column histogram and finally characters are extracted from words. Accuracy of final result is highly depends on accuracy of segmentation [7].

3.4 Feature extraction

Feature extraction is to play the essential role in any character recognition application. It can be measured as finding a set of parameters that define the shape of the character as precisely and uniquely as possible. Feature extraction techniques like Linear Discriminant Analysis (LDA), Principle Component Analysis (PCA),Independent
Component Analysis (ICA), Chain Code (CC), Scale Invariant
Feature Extraction (SIFT),Gradient based features, and Histogram. The methods that create a new features based on transformations, or combination of original features are called feature extraction algorithms [8].

3.5 Classification

Detection is the process of identification and **classification** is the categorization of the **object** based on a previously defined classes or types .The classification stage is the decision making part of a recognition system and it uses the features extracted in the previous stage [9]. Classifiers compare the input feature with stored pattern and find out the best matching class for input. The classification rules are written for other characters. Classifiers compare the input feature with stored pattern and find out the best matching class for input. In simple terms, it is this part of the OCR which finally recognizes individual characters and outputs them in machine editable form [8]. **3.6 Post processing**

This is the last step of this process, and it is used to improve the accuracy of recognition.

[4] TEXT INFORMATION EXTRACTION (TIE)

This system receives an input of a still image or a sequence of images. The input images may be a color or gray scale and it's compressed or uncompressed. And the text in the images may or may not move. TIE system process consists of five steps. (i.e.) i) text detection, ii) text localization, iii) text tracking, iv) segmentation or binarization [10], and v) character recognition.

4.1 Text Detection

This phase receives the input of images or video frames and its recognizes the text regions in the given input
image

4.2 Text Localization:

Localization of scene text is a difficult and important task and it merges the regions of text to express the text and define the tight bounds around the text objects.

4.3 Text Tracking:

This phase is applied to video data only. For the readability purpose, text embedded in the video appears in more than thirty consecutive frames. Text tracking phase exploits this temporal occurrences of the same text object in multiple consecutive frames. It can be used to rectify the results of text detection and localization stage [12]. It is also used to speed up the text extraction process by not applying the binarization [11] and recognition step to every detected object.

4.4 Text Binarization:

This phase is used to segment the text object from the background in the bounded text objects. The output of text binarization is the binary image, where text pixels and background pixels appear in two different binary levels [12].

4.5 Character Recognition:

The last module of text extraction process is the character recognition. This module converts the binary text object into the ASCII text [12].

[5] FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES USED IN THE OCR SYSTEMS

Veena Bansal and R.M.K Sinha [13] presented a complete OCR for printed Hindi text written in Devanagari script. The system used following features: Coverage of the region of the core strip, Vertical bar feature, Horizontal zero crossings, Number of positions of the vertex points, Moments, Structural descriptors of the characters for classification, Tree classifiers are used. Overall accuracy obtained at the character level is 93%. Sinha and Mahabala [14] designed a syntactic pattern analysis system for Devanagari script recognition. The system stores structural descriptors for each symbol of the script. They achieved 90% accuracy. Reena, Lipika and Chaudhury [15] have tried to exploit information about similarity between numerals, Style invariant features and stylistic variations. They presented a approach for recognition of handwritten Devnagari numerals using multiple neural classifiers. Sandhya Arora [16] have used Intersection features with Neural Network for Devanagari script and achieved 89.12% accuracy.

Singh and Budhiraja [17] presented an OCR system for handwritten isolated Gurumukhi script using Zoning, Projection histogram, Distance profile features, and Background directional features and used Support Vector Machines (SVM) for classification and thus obtained 95.04% of overall accuracy. Further Geeta and Rani [18] represented an OCR system for Gurumukhi numerals using Zone Distance features and SVM classifier and achieved 99.73% accuracy. G. S. Lehal and Chandan Singh [19] directed their efforts towards development of OCR system for Gurumukhi. They used Local features

(concave/convex parts, number of endpoints, branches, joints) and Global features (number of holes, projection profiles, connectivity etc.). For classification hybrid classification technique, binary decision tree and nearest neighbour was used. They achieved a recognition rate of 91.6%. Dharamveer Sharma and Puneet Jhajj [20] used zoning feature with hybrid classification technique using KNN and SVM classifier and achieved 72.7% accuracy. A very influential attempt made by the Jalal, Feroz and Choudhuri [21] for Bangla script. They represent neural network classifier by using Bounded rectangle calculation, Chain code generation, Slope distribution generation features. They achieved 96% system accuracy.

Chaudhuri and Paul [22] represent an OCR system to recognize Bangla and Devanagari using stroke and shaded portion feature with tree classifier. U. Bhattacharya, M. Shridhar, and S.K. Paruil [23] implemented Neural network classifier for isolated Bangla characters with chain code features and achieved 92.14% accuracy on testing sets and 94.65% on training sets. Negi and Chakravarthy [24] represent an OCR system with 92% performance using template matching, fringe distance for Telegu script. Another attempt was made by Patvardhan and Lakshmi [25] for Telegu script. They used neural classifier by using directional features and they achieved 92% accuracy. Arun K Pujari, and C Dhanunjaya Naidu [26] implemented an adaptive character recognizer for Telugu scripts using Multi resolution Analysis. They represented DNN (Dynamic Neural Network) using Wavelet analysis and achieved 93.46 % success rate. In south India, Kannada and Telugu have similar scripts. R Sanjeev and R D Sudhakar [27] represent an OCR system for printed Kannada Script using two stage Multi-Network (Neural Network) classification technique employing wavelet feature and achieved 91% accuracy at character level. M Sagar, Shobha and Ramakanth [28] designed a syntactical analysis system using Ternary Tree based classification for isolated Kannada characters. They have given more emphasis on Postprocessing step, using dictionary based approach to increase the OCR accuracy. T V Ashwin and P S Sastry [29] represents a font and sizeindependent OCR system for printed Kannada documents using support vector machines (SVM). B Chaudhuri U Pal and Mitra [30] gave a prototype OCR system for Oriya script. They use Directional features and Global Features and classified them using Decision tree classifier and achieved 96.03% accuracy at character level. Junaid, Umar, and Muhammad Umair [31] attempted to make an OCR system for isolated Urdu characters using NN classifier using structural features like width, height and checksum of the character. Their prototype gained the accuracy of 97.43%. Another good attempt was made by Jhuwair and Abdul [32] for Urdu script. They achieved the 97.12% recognition rate using Sliding window and Humoment feature using KNN classifier.

[6] APPLICATIONS

With the help of OCR, people no longer need to manually retype important documents when entering them into electronic databases. Instead, OCR extracts relevant information and enters it automatically. The result is accurate, efficient information processing in less time. The uses of OCR vary across different fields.Some of them have been explained below.

6.1 Banking

One widely known OCR application is in banking, where OCR is used to process checks without human involvement. A check can be inserted into a machine, the writing on it is scanned instantly, and the correct amount of money is transferred. This technology has nearly been perfected for printed checks, and is fairly accurate for handwritten checks as well, though it occasionally requires manual confirmation.

Overall, this reduces wait times in many banks.

*6.*2 Legal

OCR in the legal industry, for the significant movement to digitize paper documents. In order to save space and eliminate the need 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.

6.3 Healthcare

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.

6.4 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 [33].

6.5 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 [33].

6.6 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 [33].

6.7 OCR in Other Industries

OCR is widely used in many other fields, including education, finance, and government agencies. OCR has made countless texts available online, saving money for students and allowing knowledge to be shared. Invoice imaging applications are used in many businesses to keep track of financial records and prevent a backlog of payments from piling up. In government agencies and independent organizations, OCR simplifies data collection and analysis, among other processes.

[7] CONCLUSION

Digital image processing techniques, provides a wide application variety in feature extraction and classification. This paper provides a broad study of text extraction techniques and algorithms and it’s represents a study of feature extraction methods with different classifiers implemented in OCR

systems. In future, the proposed system mainly focuses on developing an algorithm for exact and fast text extraction from an image.

REFERENCES

[1] Kai Ding, Zhibin Liu, Lianwen Jin, Xinghua Zhu, "A Comparative study of GABOR feature and gradient feature for handwritten 17hinese character recognition", International Conference on Wavelet Analysis and Pattern Recognition, pp. 1182-1186, Beijing, China, 2-4 Nov. 2007

[2] Pranob K Charles, V.Harish, M.Swathi, CH. Deepthi, "A Review on the Various Techniques used for Optical Character Recognition",International Journal of Engineering Research and Applications, Vol. 2, Issue 1, pp. 659-662, Jan-Feb 2012

[3] Om Prakash Sharma, M. K. Ghose, Krishna Bikram Shah, "An Improved Zone Based Hybrid Feature Extraction Model for Handwritten Alphabets Recognition Using Euler Number", International Journal of Soft Computing and Engineering, Vol.2, Issue 2, pp. 504-58, May 2012

[4] "Optical Character Recognition Techniques: A Survey" Sukhpreet Singh M.tech Student, Dept. of Computer Engineering, YCOE Talwandi Sabo BP. India. [5] OCR System: A Literature Survey

[6] "A Literature Review on Hand Written Character Recognition" by Mansi shah & Gordhan B Jethava Department of Computer Science & Engineering Parul Institute of Technology, Gujarat, India. Information Technology Department Parul Institute of Engg. & Technology, Gujarat, India.

[7] Lim, Jae S.," Two-Dimensional Signal and Image Processing", Englewood Cliffs, NJ, Prentice Hall, 1990, pp. 469-476.

[8] http://shodhganga.inflibnet.ac.in:8080/jspui/bitstream/10603/4 166/10/1 0_chapter%202.pdf

[9] http://arxiv.org/ftp/arxiv/papers/1103/1103.0365.pdf

[10]Datong Chen, Juergen Luettin, Kim Shearer, "A Survey of Text Detection and Recognition in Images andVideos". Institute Dalle Molle d'Intelligence ArtificiellePerceptive Research Report, August 2000, Page no. 00-38.

[11]Datong Chen, Juergen Luettin, Kim Shearer, "A Survey of Text Detection and Recognition in Images andVideos". Institute Dalle Molle d'Intelligence ArtificiellePerceptive Research Report, August 2000, Page no. 00-38

[12] Chowdhury Md Mizan, Tridib Chakraborty* and Suparna Karmakar," Text Recognition using Image Processing" International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May – June 2017

[13] Veena Bansla and R M K Sinha, "A Complete OCR for printed Hindi Text in Devanagari Script", IEEE 800 - 804 2001.

[14] Sinha. M. K., Mahabala., "Machine Recognition of Devnagari Script", IEEE T. SYST. MAN Cyb., vol.. 9,pp.435449,1979.

[15] Reena Bajaj, Lipika Dey and Santanu Chaudhury, "Devnagari numeral recognition by combining decision of multiple connectionist classifiers", Vol. 27, Part 1, February 2002, pp. 59–72.

[16] Sandhya Arora, "A Two Stage Classification Approach for Handwritten Devanagari Characters", IEEE 399 - 403 vol 2. [17] Pritpal Singh and Sumit Budhiraja, "Feature Extraction and Classification Techniques in O.C.R. Systems for Handwritten Gurmukhi Script". International Journal of Engineering Research and Applications (IJERA), Vol.1, ISSUE 4,pp.1736-1739.

[18] Gita Sinha Rajneesh Rani Renu Dhir , "Handwritten Gurmukhi Numeral Recognition using Zone-based Hybrid Feature Extraction Techniques", International Journal of Computer Applications(09758887), Volume 47- No. 21 June 2012.

[19] G. S. Lehal and Chandan Singh, "Feature Extraction and

Classification for OCR of Gurmukhi Script". [20] Dharamveer Sharma, Puneet Jhajj, "Recognition of Isolated Handwritten
Charactersin Gurmukhi S Volume 4– No.8, August 2010cript",
International Journal of Computer Applications (0975 – 8887) [21] Jalal Uddin Mahtnud, Mohammed Feroz Raihan and
Chowdhury Mofizur Rahman, "A Complete OCR System for Continuous Bengali Characters",IEEE 1372 - 1376 Vol. Oct. 2003.

[22] B. B. Chaudhuri and U. Pal, "An OCR System to Read Two Indian Language Scripts: Bangla and Devnagari (Hindi)", IEEE 1011 - 1015 vol.2, Aug 1997. [23] U. Bhattacharya1, M. Shridhar, and S.K. Parui1, "On
Recognition of Handwritten Bangla Characters". [24] Atul
Negi, Chakravarthy Bhagvati, B. Krishna, "An
OCR system for Telugu", IEEE 1110 – 1114 -2001. [25] Vasantha Lakshmi, C. Patvardhan, C. , "A high accuracy OCR system for printed Telugu text", IEEE 725 - 729 Vol.2 , Oct2003.

[26] Arun K Pujari, C Dhanunjay Naidu, "An Adaptive Character Recognizer for Telugu Scripts using Multiresolution Analysis and Associative Memory".

[27] R Sanjeev Kunte, R D Sudhaker Samuel, "An OCR
System for Printed Kannada Text Using Two - Stage Multinetwork Classification Approach Employing Wavelet Features", IEEE 349 – 353,Dec-2007. [28] Sagar, B.M. ,
Shobha, G. , Kumar, P.R. , "Complete Kannada Optical Character Recognition with syntactical analysis of the script", IEEE 1 – 4 , Dec. 2008.

[29] T V Ashwin and P S Sastry, "A font and sizeindependent OCR system for printed Kannada documents using support vector machines", Vol. 27, Part 1, February 2002, pp. 35–58.

[30] Chaudhuri, B.B, Pal, U. , Mitra, M. , "Automatic recognition of printed Oriya script", IEEE 795 – 799, 2001. [31] Tariq, J, Nauman, U, Naru, M.U , "Softconverter: A novel approach to construct OCR for printed Urdu isolated characters", IEEE V3495 - V3-498 ,April 2010. [32] Sardar, S, Wahab, A, "Optical character recognition system for Urdu",IEEE 1 - 5 , June 2010. [33]"Survey of OCR Applications" by Amarjot Singh, ketan bacchuwar, Akshay bhasin.