

Open in app ↗

Sign up

Sign in

Medium

Search

Write



TDS Archive

# What is an OCR ??

A basic theoretical overview of the working of an Optical Character Recognition system.



Susmith Reddy

Follow

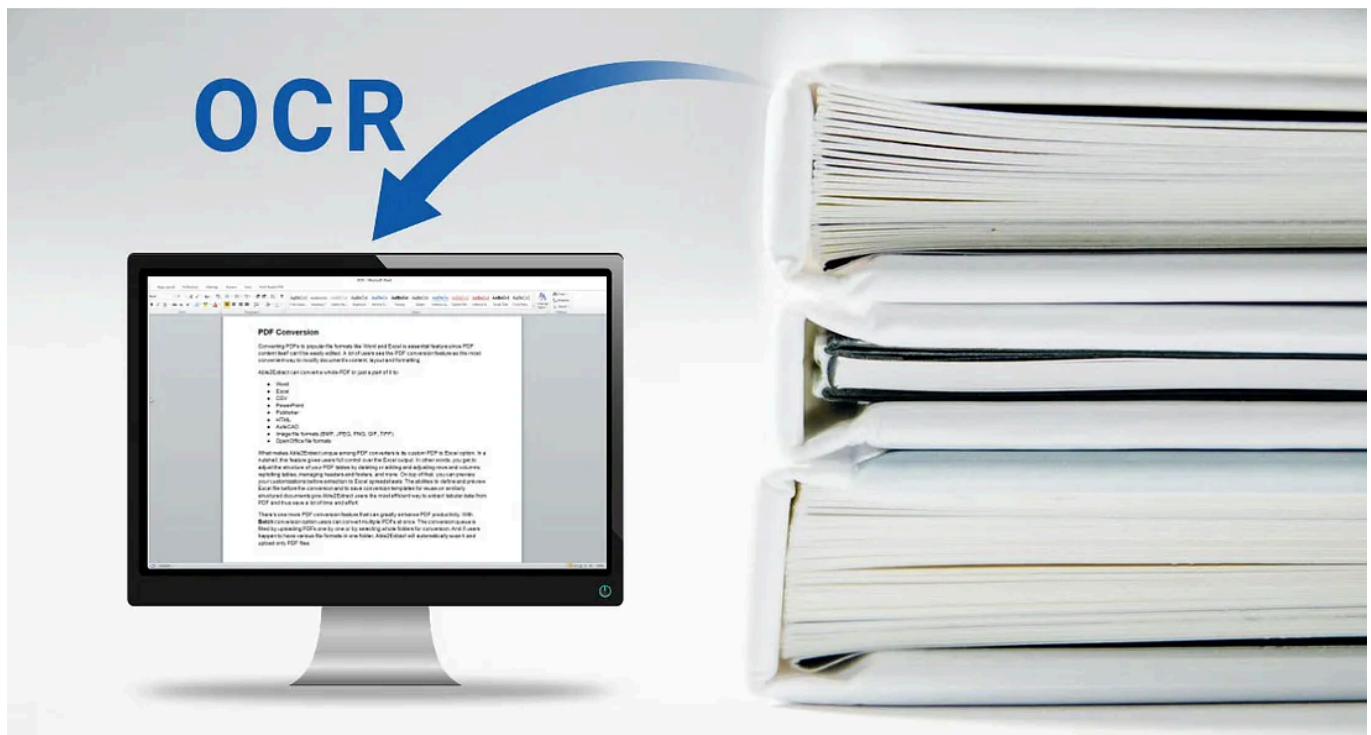
5 min read · Mar 25, 2019



190

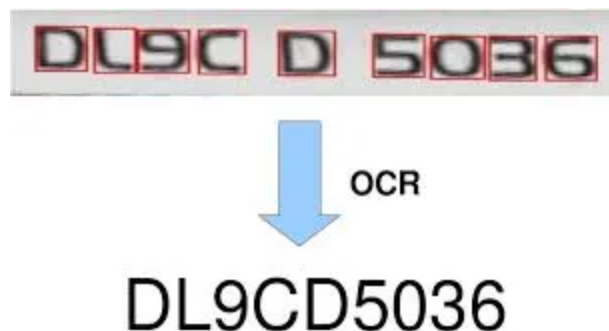


1

Source: [investintech.com](https://investintech.com)

The necessity of digitisation is rapidly increasing in the modern era. Due to the growth of information and communication technologies (ICT) and the wide availability of handheld devices, people often prefer digitized content over the printed materials including books and newspaper. Also, it is easier to organize digitized data and analyze them for various purposes with many advanced techniques like artificial intelligence etc. So to keep up with the present technological scenario, it is necessary to convert all the information present till now which is in the printed format to digitised format.

Here comes OCR ....Our saviour 🦹 🦹 which helps us in performing the tedious work of digitising the information. OCR stands for *Optical Character Recognition*, whose primary job is to *recognise the printed text in an image*. Once we recognise the printed text with the help of OCR, we can use that information in various types.



Recognizing the text on Number Plate. **Source:** [scholorworks](#)

## What are you going to learn?

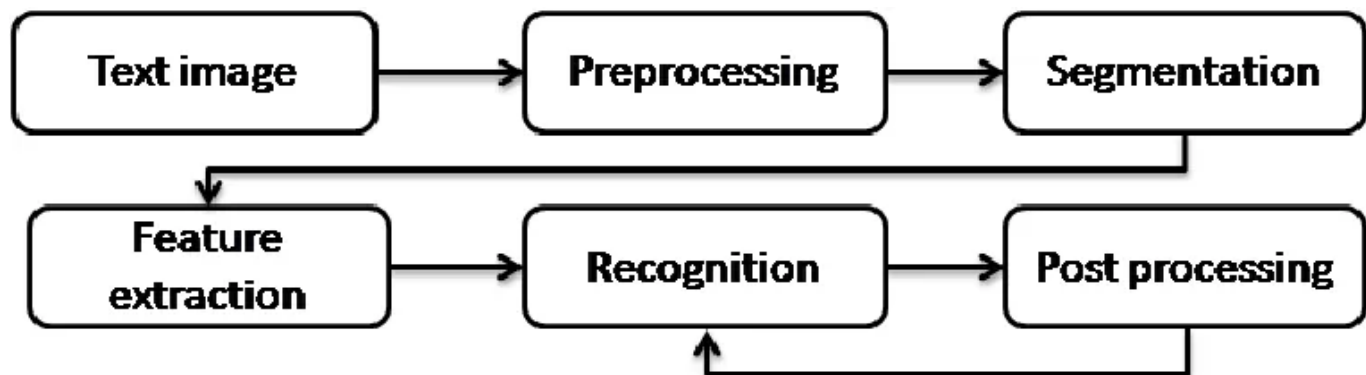
This is a 3-part series of articles that explains various concepts and phases of an OCR system. Let's have a look at what you are going to learn in each part

- *part-I* (this article), a high-level theoretical overview of the working of an OCR system

- **part-II**: different steps performed in the *Pre-processing stage* along with code samples
- **part-III**: different types of *Segmentation* that can be performed on a *pre-processed image*.

## Let's go...

Below image shows the different phases in the workflow of an OCR system.



Phases of OCR. **Source:** Reference [4]

Let's briefly discuss each phase shown in the above image:-

### A. Image acquisition

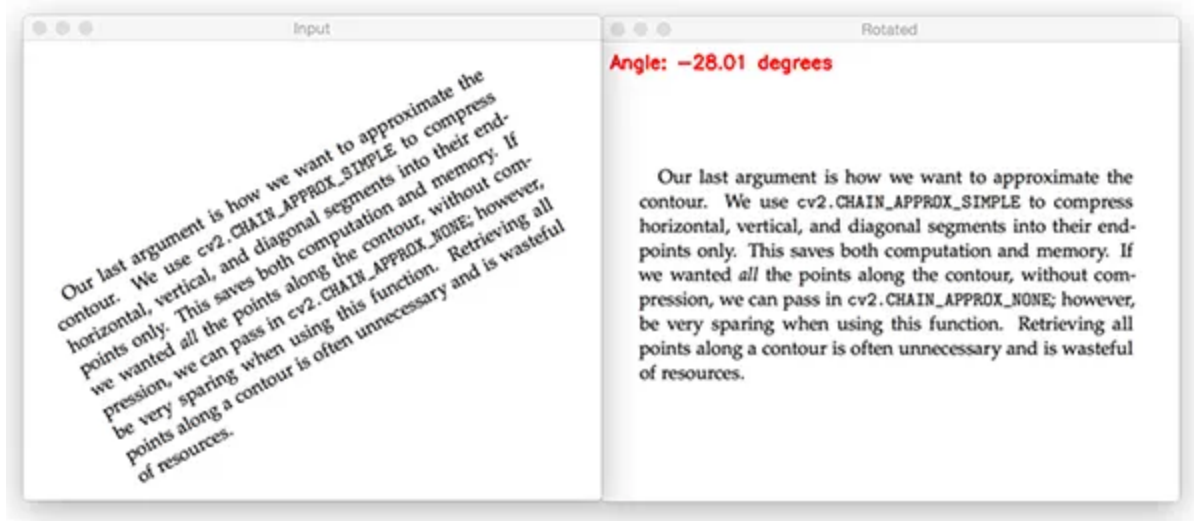
This involves scanning a document and storing it as an image on which recognition has to be performed.

### B. Pre-Processing

We cannot input an image directly for the OCR system. Some pre-processing has to be done on the image so that it becomes moderately easy for OCR to recognise the information in the image. It is like getting dressed up according to the occasion.

## Pre-processing of image includes:

1. *Skew Correction*: Image obtained from the previous stage may not be correctly oriented, It may be aligned at any angle. So we need to perform skew correction to make sure that the image forwarded to subsequent stages is correctly oriented.



Skew Correction. **Source:** [pymimagesearch.com](https://pymimagesearch.com) by [Adrian Rosebrock](#)

2. *Binarization*: It means converting a *Coloured image* to *Binary image* (containing only black & white colours). Usually, in practice, this conversion of *Coloured image* to *Binary image* is done by an intermediate *GrayScale image*.

Coloured image → Gray Scale image → Binary image

This can be done using different methods like

- Adaptive Thresholding
- Otsu's Binarization
- Local Maxima and Minima Method

## Get Susmith Reddy's stories in your inbox

Join Medium for free to get updates from this writer.

Enter your email

Subscribe

---

I prefer *Adaptive Thresholding*, as it binarizes the image with a certain threshold which is calculated dynamically depending upon its locality in the image.

3. *Noise Removal*: Noise (small dots or foreground components) may be introduced easily into an image while scanning it during *Image Acquisition* because of various reasons like low clarity camera, a shadow on image etc. This noise should be removed so that the image will be clean and uniform.

4. *Thinning and Skeletonization*: Different images have text with different width of strokes. This variability is very high in the case of handwritten words. Skeletonization is a technique, using which we can make all strokes to have a uniform width (Maybe 1 pixel wide or few pixels wide)



(a) Original Image. (b) Converted to Grayscale. (c) Binarized image. (d) Thinning and Skeletonization are done. (e) Noise Removed. **Source: Reference [3]**

### C. Segmentation

Once a *Clean* image is obtained, after the Preprocessing stage, the next stage is the *Segmentation*. It is a technique of breaking the whole image into subparts to further process them.

Three types of *Segmentation* can be done in OCR:-

- Line level Segmentation
- Word level Segmentation
- Character level Segmentation

### D. Feature Extraction

In this phase, we extract some *Unique Features* of the segmented subcomponents obtained from the previous stage. There are many

techniques through which we can extract features like its shape, strokes etc. But nowadays, for feature extraction, we mostly rely on Machine learning models which uses a stack of CNN, RNN(Recurrent Neural Networks), LSTM(Long Short term memory) layers.

### ***E.Classification***

This is the decision-making stage of an OCR system . Classification uses the features extracted in the feature extraction stage to identify the text segment. Algorithms like SVM can be used for classification.

Apart from these phases, To improve the performance of OCR we can also perform an optional *Post Processing* step after the Classification stage.

### ***F. Post-processing***

The most probable errors that would occur in the OCR system is due to the wrong prediction in the *Classification* stage (this might be due to the poor feature extraction, a lot of noise in the picture etc). In the majority of the cases, these prediction errors will result in small spelling mistakes because of the wrong prediction of one or two letters in a word (like the word 'ball' is predicted as 'boll'). So these type of spelling mistakes can be corrected using language models, Word2Vec models (like CBOW and skip-gram) etc.

Wohooo.....now you know the basic workflow of an OCR system.

## **Further Readings**

- **Part-II** — Pre-Processing in OCR
- **Part-III** — Segmentation in OCR

Happy Learning !!!!

Any doubts, Suggestions, Corrections are Welcome. 😊

## References:

- [1] P. Dhande and R. Kharat, “Recognition of cursive English handwritten characters,” *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, 2017, pp. 199–203, doi: 10.1109/ICOEI.2017.8300915.
- [2] Hamad, Karez & Kaya, Mehmet. (2016). A Detailed Analysis of Optical Character Recognition Technology. *International Journal of Applied Mathematics, Electronics and Computers*. 4. 244–244. 10.18100/ijamec.270374.
- [3] Choudhary, Amit & Rishi, Rahul & Savita, Ahlawat. (2013). A New Character Segmentation Approach for Off-Line Cursive Handwritten Words. *Procedia Computer Science*. 17. 88–95. 10.1016/j.procs.2013.05.013.
- [4] Boulid, Youssef & Souhar, Abdelghani & Elkettani, Youssfi. (2017). Handwritten Character Recognition Based on the Specificity and the Singularity of the Arabic Language. *International Journal of Interactive Multimedia and Artificial Intelligence*. 4. 45–53. 10.9781/ijimai.2017.446.

Machine Learning

Ocr

Segmentation

Image Processing

Recognition