---

layout: post

comments: true

title: Invoice Reader

description: Used to read an Invoice and convert it to readable text format.

---

{{Invoice Reader}}

================

<p class="meta">

Yash Sinha &

Sonal Sidhu Sharma

</p>

---

\*\*TOOLS USED:\*\*

Language Used For Analysis: Python

\*\*Libraries Used:\*\*

1.NumPy

2. PIL(Python Imaging Library) offers several standard procedures for image manipulation. These include:

* per-pixel manipulations,
* masking and transparency handling,
* image filtering, such as blurring, contouring, smoothing, or edge finding,
* image enhancing, such as sharpening, adjusting brightness, contrast or color,
* adding text to images and much more.

3. OpenCV-Python is a Python wrapper for the original OpenCV C++ implementation. OpenCV-Python makes use of **Numpy**, which is a highly optimized library for numerical operations with a MATLAB-style **syntax**. All the OpenCV array structures are converted to and from **Numpy** arrays.

**4. Python**-**tesseract** is an optical character recognition (OCR) tool for **python**. That is, it will recognize and “read” the text embedded in images. **Python**-**tesseract** is a wrapper for Google's **Tesseract**-OCR Engine.

### Implementation:

All the above mentioned libraries were imported and image was imported from PIL. Started with defining a function which take the path of the image as its input. OpenCV library is used to read and convert the image to gray scale. A Kernel was created which serves as a feature to the dilation and erosion. Then “Dilation and Erosion” were applied to remove noises from the image and the new image was saved by the name of removed\_noise. Then a threshold was applied to the image with only black and white and this change was saved by the name of thres. Then comes the job of pytesseract to get the text from the images. The previous steps of noise removal was necessary as pytesseract doesn’t work well with noisy images. Pytesseract is then used to convert the image words to string and the string obtained is then saved to a text file by the name of new\_days.

The Function work is completed and now the function is ready to use.

---

### PS

Feedback/Comments/suggestions, as well as contributions on this blog are

welcome.

<p class="meta"\>

Published on 02/July/2019

</p>