

Forum:[Hack Forums](#)

Thread: [Pytesseract OCR shipping labels](#)

LatestGreatestPro - February 21, 2024, 2:45 pm

Hello all,

I recently joined this forum and upgraded and I love it here!

Anyways, my issue is i have shipping labels from UPS for my business and I want to use a script in python to extract the name of the recipient on the shipping label, the reference numbers, and the tracking number. Issue is when I run this script:

```
Code import cv2
import pytesseract
# Load image
img = cv2.imread("example_image.jpg")
# Convert image to grayscale
gray = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
# Apply threshold to convert to binary image
threshold_img = cv2.threshold(gray, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU)[1]
# Pass the image through pytesseract
text = pytesseract.image_to_string(threshold_img)
# Print the extracted text
print(text)
```

It's inaccurate as shipping labels have different font sizes and the text can be quite close to each others (horizontally)

Hexicidal - March 1, 2024, 2:26 pm

You want a script that can scan the shipping label, and then extract the recognized string name correct?

LatestGreatestPro - March 3, 2024, 9:42 am

I figured it out thanks ! Had to use simple rectangles with openCV and then resize it a throw that into pytesseract image to string function !