http://www.paperfile.net/freeocr/help.htm:

FreeOCR Guide

Free OCR Software Help Guide

FreeOCR (paperfile.net)

Contents

FreeOCR V3 User Guide

This guide is intended to get you started using FreeOCR properly and hopefully get excellent results using the software

What is FreeOCR?

Basically FreeOCR is OCR software which stands for Optical Character Recognition which means that the software can read characters/words from an image to produce text output. If you scan a letter for example you get an image file which is just a bunch of dots (pixels) and your computer has no knowledge if the image contains text or a picture of your cat! but by using OCR software your computer will attempt to read to pixels and use pattern matching algorithms to output editable text.

Some people think that OCR software just converts an image to a Word processor document and although this is partially true the process is a lot more complicated than that and FreeOCR will only output plain text so you have to spell-check and format the output text yourself and although there are some high end OCR packages that will do this for you (with variable results) they carry very high price tags $500 or more so if you only need OCR occasionally then FreeOCR would be a better bet and will do 90% of the work for you.

OCR engines are never 100% accurate and manufacturers quote percentages like 99.8% but these quotes are always for very high quality clean scans in the real world the results are much lower however FreeOCR has a rate of between 98-99% accuracy which means for every 100 characters recognised there will be 1 or 2 errors and these can often be fixed using a spell-check of your word processor.

The Basics

Ok, if you are the sort of person that does not like to read help guides then just follow the 3 points below to get good results with FreeOCR

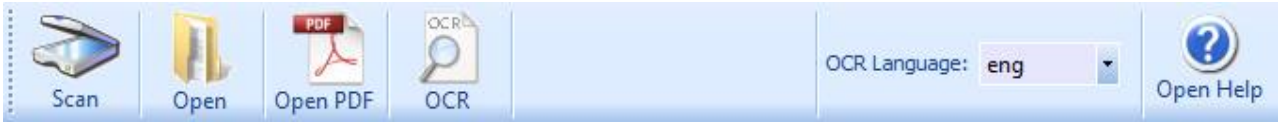
1) Scan your originals at **300dpi greyscale** - higher resolutions such as 600dpi will NOT give better results and will probably give worse FreeOCR needs between 200-300 dpi and greyscale will give slightly better results than Black & White or Color

2) Rotate the scan – FreeOCR cannot read images that are upside down or rotated by 90 Deg. So use the rotate buttons

3) Select the text area to recognise by drawing a box around it – this will often give better results than OCR'ing the whole page unless you have a very clean scan.

To use FreeOCR you should be comfortable using Windows software and know how to copy & paste text between programs.

Description of Main Buttons



Scan

Lets you scan an image from any Twain or Wia compatible scanner, remember to set the scanning resolution to 300 dots per inch

Open

Opens/loads an image from file

Open PDF

This will open a PDF file, to do this FreeOCR converts each page to a bitmap, note that only scanned PDF's are supported ie. PDF's that contain an image.

OCR

starts the OCR process

OCR Language

If you have installed additional laguages then you can select them here

Open Help

Opens the online help guide - This Guide infact!

Image Buttons

Navigate page - if you have a multipage Tiff or PDF you can change the page with these.

Fit all of Image in window

Fit Width

Zoom buttons

Rotate Buttons

Selection tool - lets you crop an image or copy the selection to the clipboard

Text Buttons

Clear Text window

Save Text - save as txt file

Remove Line Breaks - Handy if you need to fit the text to a page

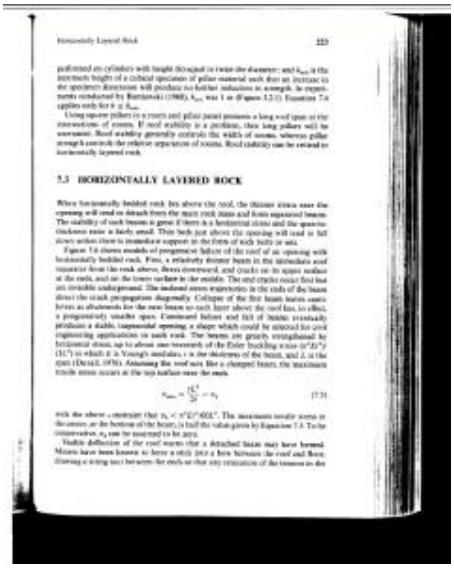
Copy Text - copies the contents of the text window to the clipboard

Export to Word send the text directly to MS Word (2000 & later)

Please note that as well as these buttons you can freely select text from the window and copy and paste into any Windows application

Example 1: OCR'ing a simple document

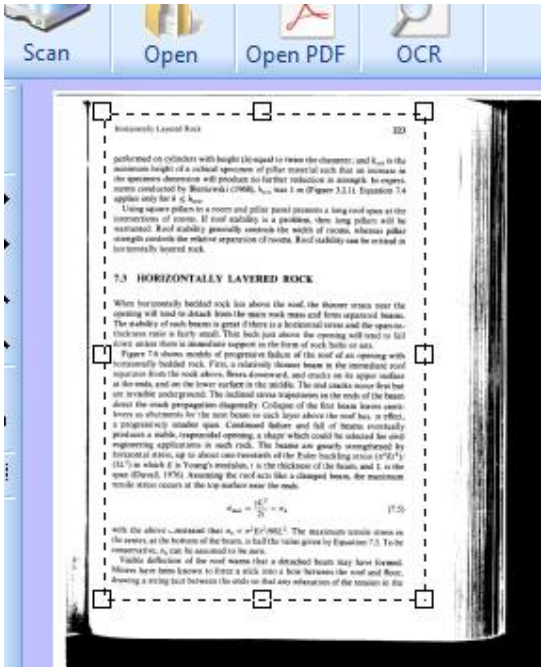
This is a page from a book scanned at 300dpi, you can download the image HERE if you want to try (if the image just opens in your browser then you may need to right-click and choose 'save link as' or 'save target as'



Ok, the image is quite good but this scan has very dirty borders and if we just press the OCR button we will get results similar to those below:

```
Q it ( { . »
Horizontally Layered Rock 223 { V
( (
= . i2,.
performed on cylinders with height (h) equal to twice the diameter; and hm, is the ’ ji ° ~ Q
( minimum height of a cubical specimen of pillar material such that an increase in - { p(
the specimen dimension will produce no further reduction in strength. In experi- ( Xi ( B
ments conducted by Bieniawski (1968), hm, was l m (Figure 3.2.1). Equation 7.4 (T § g
applies only for h 5 hm,. it g (
Using square pillars in a room and pillar panel presents a long roof span at the i ; j
intersections of rooms. If roof stability is a problem, then long pillars will be , ,
warranted. Roof stability generally controls the width of rooms, whereas pillar ( J
strength controls the relative separation of rooms. Roof stability can be critical in E , t
horizontally layered rock. , ·i (
. ‘ S .
```

The text recognition is fine but because of the dirty borders the OCR engine has tried to read text to the edges of the page which has added some extra characters, this would be time consuming to clean up so on the next scan let's just draw a box around the text.



I have drawn a box while holding down the left mouse button

Now pressing on the OCR button FreeOCR just OCR's the selected area and we get the results below

I-lorizontally Layered Rock 223 performed on cylinders with height (h) equal to twice the diameter; and hm, is the minimum height of a cubical specimen of pillar material such that an increase in the specimen dimension will produce no further reduction in strength. In experiments conducted by Bieniawski (1968), hm, was I m (Figure 3.2.1). Equation 7.4 applies only for h 5 hm,. Using square pillars in a room and pillar panel presents a long roof span at the intersections of rooms. If roof stability is a problem, then long pillars will be warranted. Roof stability generally controls the width of rooms, whereas pillar strength controls the relative separation of rooms. Roof stability can be critical in horizontally layered rock

Just that small extra step has increased the results dramatically.

Now we have the text we can copy it from the text window into Word or any Windows program.

That was pretty basic but on the next page we will OCR and reconstruct a document containing 2 columns and a photo

Example 2: OCR'ing a 2 column document with photo

This is a PDF with a fake UFO report you can download the file HERE if you want to try (you may need to right-click and choose 'save link as' or 'save target as'


UFO Report 10266

Just pressing the OCR button would be useless on this example because although we would get random characters when OCR'ing the photo it would also mix up the 2 columns into one so for this example using the selection tool is a must!

I am going to use MS Word to recreate the document but you could use Open Office or publisher, in fact any program that supports columns and photos.

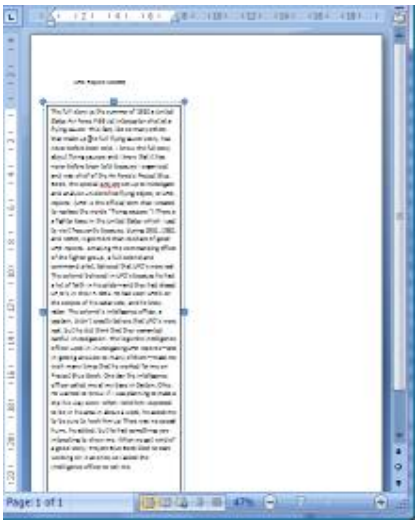
1) now to start off I am just going to OCR the page title (yes that is a bit lazy as I could have just typed the title in)


UFO Report 10266

2) Then press the Export to Word button which will open up a new document with the title in it, if you are not using Word then just open a new document and copy and paste the text.

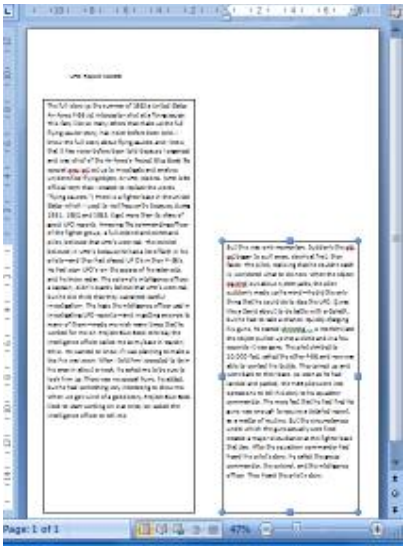
3) Next I have selected and OCR'ed the first column also I pressed the 'remove line breaks button' then in Word created a new Text Box (Insert menu on Word) and copy and pasted the text into it as shown below.

FreeOCR

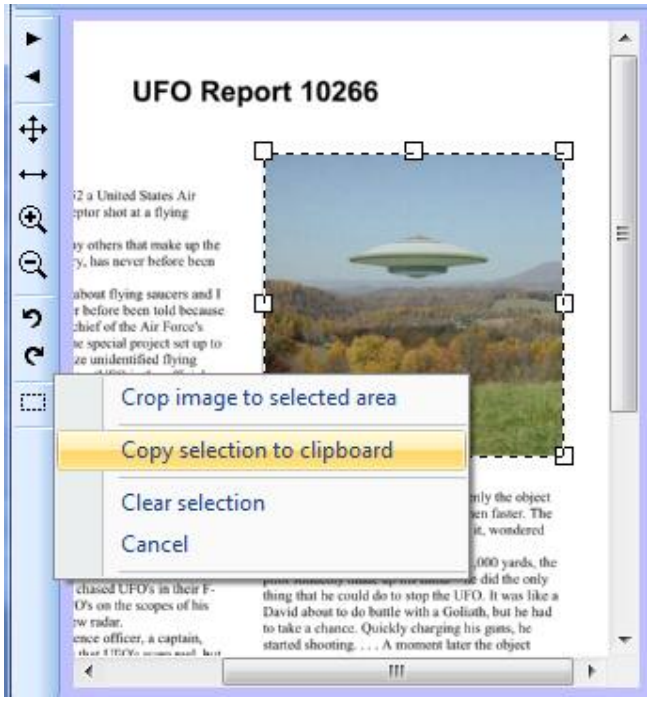

MS Word

4) Now clear the text window and do the same for the second column

FreeOCR


MS Word

5) Now draw a box around the photo to select the image, then press the selection button and choose
'Copy selection to clipboard' this will place the selected image onto the Windows clipboard so now we can just paste the image into Word (Tip: paste the image into a text box then you can position it easily)



FreeOCR


MS Word

Now we have all of the text and images in Word you can play around positioning the text boxes and adjusting the font sizes to suit and don't forget to do a final proof read and spell check.

As you can see there is a manual element of work envolved to get the best out of FreeOCR but I am sure you will find it easier than recreating a document from scratch.