

Advanced Scanning Techniques

High Tech Center Training Unit

of the California Community Colleges at the
Foothill-De Anza Community College District

21050 McClellan Road
Cupertino, CA 95014
(408) 996-4636
(800) 411-8954

www.htctu.net



URL to our CC license:

<http://creativecommons.org/licenses/by-nd-nc/1.0/>

Creative Commons website:

<http://creativecommons.org>

Improving Your Scanning

Scanning is more of an art than a science. Most of the time auto settings with Black and White at 300 dpi will give you a very good scan, but some paper types may require adjusting the brightness or contrast or both in order to get a more accurate scan. Some small fonts may require increasing the dpi to 400. Some thin papers may require decreasing the dpi to 200 or even 150.

As with all arts, practice makes perfect, and it is always a wise idea to scan and OCR a couple of representative pages in order to determine the best scanner settings for a book.

It is also wise to keep track of the experiments you make and the settings that you find work well. After awhile, you will remember and know what settings need to be altered, but in the beginning, writing down your settings is a good idea.

Scanner Settings

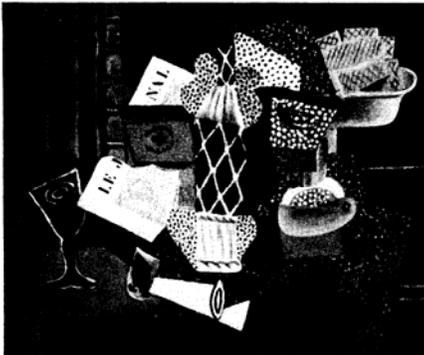
Mode

For most text, Black and White mode is preferable. There are occasions, however, when dealing with color or shaded backgrounds that Black and White-ED (which uses an "error diffusion" algorithm to simulate halftones, i.e., grays) might be a good option.

Black and White mode will produce small files and is often the best choice if you need to change the brightness/threshold setting, but it can sometimes be worth trying other options with books that have lots of different colored sections on a page.

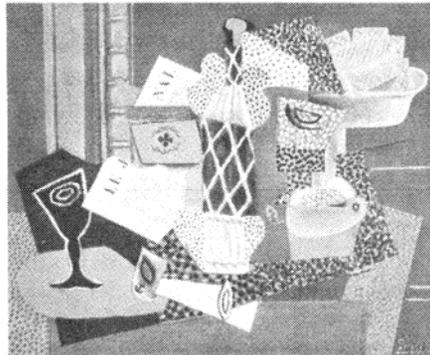
Note that the 9080C now offers "Advanced Text Enhancement" mode to help with very light documents or when text is printed on a dark background.

Image scanned in B/W—file size was 474 KB



34 LA BOUTEILLE DE MARASQUIN (1914)

Image scanned in B&W ED—file size was 474 KB



34 LA BOUTEILLE DE MARASQUIN (1914)

Image scanned in Grayscale—file size was 3,731 KB



34 LA BOUTEILLE DE MARASQUIN (1914)

Resolution

Your resolution will normally be set at 300 dpi. This resolution is considered optimal for text and is what OCR programs are geared to work with.

Small text may require 400 dpi

If you have thin paper, you may be getting "bleed through" from the back side. In such a case, drop your dpi to 150–200 to improve the scan.

Brightness

Brightness lightens or darkens all the pixels on the page. Sometimes with very glossy paper so much light bounces back from the page that you will need to reduce the brightness of the page so that you don't get areas of "white out" where the image disappears entirely.

You can think of brightness as bringing "balance" to the image—not too dark, not too light. Increasing the brightness lightens the image. Decreasing brightness darkens it.



Brightness is a measure of how dark or pale a scanned image will be.

Too dark: Letter shapes run together

Too light: Letter shapes are thin or broken

The value scale is 1–255. The default setting is 128.

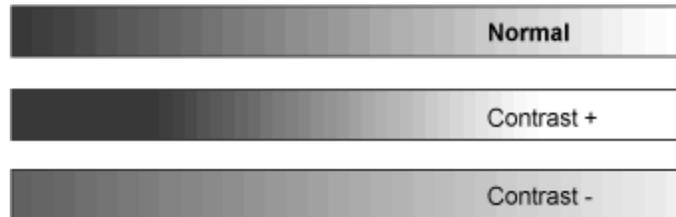
Lower numbers: Darker (decrease in brightness)

Higher numbers: Lighter (increase in brightness)

Contrast

Contrast is a measure of how much difference there is between the light parts and dark parts of an image. Changing the contrast alters the range of lights and darks.

Increasing the contrast will make the lights lighter and the darks darker. Decreasing the contrast will lighten the darks and darken the lights.



The value scale is 1–13. The default setting is 7.

Larger contrast value (higher number): Increases the contrast

Smaller contrast value (lower number): Decreases the contrast

Threshold (Or where did my other bar go?)

Setting the mode to black and white will gray out the contrast bar and leave only the brightness bar. Although labeled "brightness," this bar now serves a slightly different function than it does when scanning in grayscale.

When scanning in black and white, the machine has to make a decision about all the grays in the image. Since only black or white are choices, the scanner has to decide, "Should I call this gray black or call it white?" That cut-off point between black and white is known as the threshold.

When scanning in black and white mode, the brightness bar now functions as that cut-off point.

Increasing the threshold will add more white to the image. Decreasing the threshold will add more black to the image.

To improve the scan when the textbook uses gray boxes around text, try increasing the threshold (brightness). Essentially, you are telling the scanner to consider the gray in those boxes as white.

Use care, however, that you do not increase the threshold to the point that you are losing some of the main body of the text. On the 5080C, you can use the compliment thin line option to fill in lines that are reduced too much by increasing the brightness.

Color Dropout

If you are scanning a color book that has boxes or screens behind text, you can have the scanner dropout a color. (Note that the 9080C now allows you to drop out color on only one side of the page.)

Also note that most papers are slightly colored and not pure white. Dropping out the paper color can improve the scan. For yellowish papers, drop our red. For olive or greenish papers, drop out green.

One of the most useful results of this feature is the ability to drop out marks such as highlighter (drop out color of highlighter) and blue ballpoint pen (drop out blue). Note that some highlighters show up less than others. Green, for instance, often does not show up on the scan, while orange, pink, and yellow can be more problematic.

Below is a page that has orange highlighter on it, scanned without color drop out.

HUMAN NATURE

::: 5 :::

developed except by evolution, and there is now overwhelming evidence that there is no other way for evolution to work except by competitive reproduction. [REDACTED]

[REDACTED] The ability to reproduce is what makes living things different from rocks. Besides, there is nothing inconsistent with free will or even chastity in this view of life. Human beings, I believe, thrive according to their ability to take initiatives and exercise individual talent. But free will was not created for fun; there was a reason that evolution handed our ancestors the ability to take initiatives, and the reason was that [REDACTED]

[REDACTED] than human beings who do not reproduce. Therefore, free will itself is any good only to the extent that it contributes to eventual reproduction.

Look at it another way: If a student is brilliant but terrible in examinations—if, say, she simply collapses with nervousness at the very thought of an exam—then her brilliance will count for nothing in a course that is tested by a single examination at the end of the term. Likewise, if an animal is brilliant at survival, has an efficient metabolism, resists all diseases, learns faster than its competitors, and lives to a ripe old age, but is infertile, then its superior genes are simply not available to its descendants. [REDACTED]

[REDACTED] None of your direct ancestors died childless. Consequently, if we are to understand how human nature evolved, the very core of our inquiry must be reproduction, for reproductive success is the examination that all human genes must pass if they are not to be squeezed out by natural selection. Hence I am going to argue that there are very few features of the human psyche and nature that can be understood without reference to reproduction. I begin with sexuality itself. Reproduction is not synonymous with sex; there are many asexual ways to reproduce. But reproducing sexually must improve an individual's reproductive success or else sex would not persist. I end with intelligence, the most human of all features. It is increasingly hard to

Below is the same page scanned with the option of dropping out red.

developed except by evolution, and there is now overwhelming evidence that there is no other way for evolution to work except by competitive reproduction. Those strains that reproduce persist; those that do not reproduce die out. The ability to reproduce is what makes living things different from rocks. Besides, there is nothing inconsistent with free will or even chastity in this view of life. Human beings, I believe, thrive according to their ability to take initiatives and exercise individual talent. But free will was not created for fun; there was a reason that evolution handed our ancestors the ability to take initiatives, and the reason was that free will and initiative are means to satisfy ambition, to compete with fellow human beings, to deal with life's emergencies, and so eventually to be in a better position to reproduce and rear children than human beings who do not reproduce. Therefore, free will itself is any good only to the extent that it contributes to eventual reproduction.

Look at it another way: If a student is brilliant but terrible in examinations—if, say, she simply collapses with nervousness at the very thought of an exam—then her brilliance will count for nothing in a course that is tested by a single examination at the end of the term. Likewise, if an animal is brilliant at survival, has an efficient metabolism, resists all diseases, learns faster than its competitors, and lives to a ripe old age, but is infertile, then its superior genes are simply not available to its descendants. Everything can be inherited except sterility. None of your direct ancestors died childless. Consequently, if we are to understand how human nature evolved, the very core of our inquiry must be reproduction, for reproductive success is the examination that all human genes must pass if they are not to be squeezed out by natural selection. Hence I am going to argue that there are very few features of the human psyche and nature that can be understood without reference to reproduction. I begin with sexuality itself. Reproduction is not synonymous with sex; there are many asexual ways to reproduce. But reproducing sexually must improve an individual's reproductive success or else sex would not persist. I end with intelligence, the most human of all features. It is increasingly hard to

Gamma

Whereas contrast affects the end-points of the darks and lights, gamma alters the midrange tones.

Increasing the gamma will darken the midtones. Decreasing the gamma will lighten the midtones.

Contrast this effect with adjustments to brightness, which changes the darkness or lightness for **all** tones, or to contrast, which increases or decreases the **range** of lights and darks.

The default factor setting is 1. Lower numbers will lighten the midrange grays; higher numbers will darken the midrange grays.

The settings range from 0.2 to 5 and can be set in 0.1 increments or adjusted with the mouse by clicking and dragging on the line.

It is a good idea to first adjust brightness and contrast then to work with the gamma as necessary.

Despeckle/Remove Dot

If there are a lot of stray marks on the page, try using the despeckle or remove dot feature to help alleviate some of the "noise." Please be aware, however, that if you are scanning a foreign language document, some of those little dots are marks are supposed to be there.

(Note that on the 9080C, you may need to use the "soft" filter to get this effect.)

Erase Notch

If characters look a bit jagged, erase notch may help to smooth them.

Reverse Image

Reversing the image causes black to be seen as white and white as black. Use this setting when most of the page (or at least the portions you least want to reenter) is light on dark.

Note that the entire page is affected so small sections of reverse text should be ignored. The OCR programs can read light/white text on a dark background.

The 9080C allows you to enhance a color and this setting can sometimes be used to darken a light background for better contrast against white text.

Static

Sometimes with glossy paper, static electricity holds the pages together and causes double feeds.

Get dryer sheets from the store. Tear off a strip and cut it along one edge so that you have fringe. Tape the sheet above the paper tray so that the fringe brushes across the top of the paper as it is pulled through the feeder. Also, tape a similar sheet to the back of the feeder so that it lays over the paper but remains in place as the paper is pulled from beneath the fringe.

If you live in a dry climate, keeping a humidifier in the scanning room can help with paper that builds up static. Static electricity is worse in dry air.

You can also try setting the dpi to 400 in order to slow down the paper-feed rate. Sometimes a slower paper rate will reduce static.