

Advanced Scanning

High Tech Center Training Unit
21050 McClellan Rd. Cupertino, CA 95014
www.htctu.net



Foothill – De Anza Community College District
California Community Colleges

Advanced Scanning Techniques

10:00

Introduction and Expectations

Paper Overview

Color dropout
Glossy papers
Thin papers
Static build-up

Scanner Settings Overview

Mode
Resolution: DPI
Brightness
Contrast
Gamma
Color dropout
Reverse image

Demo Scans

11:45

Lunch

1:00

Practice Scans

Colored pages

2:45

Break

3:00

Practice Scans cont.

Head-to-head Scans

4:30

End

The auto settings of the scanners handle most projects quite well, but what do you do with those colored backgrounds, thin papers, glossy papers, and other scanning challenges? This one-day training will explore what lies beyond the auto settings: contrast, brightness, color, etc.

Training Objectives

Participants will be able to do the following:

- understand contrast and brightness settings
- recognize paper that may need to be scanned at less than 300 dpi
- adjust the scanner settings to drop out gray backgrounds
- recognize scanner problems on individual pages

Intended for Alternate Media Specialists.



Advanced Scanning Techniques



Workflow

- Separate pages one by one
 - Gives a feel for book structure and paper
 - Pick a few representative pages (6 max)
 - Use these as your test scan set
 - Scan and OCR with auto settings
 - Fewer than 1 error per page, settings probably fine
 - 2 errors per page or more, adjust settings
-



Skew

- Skew is slant (i.e., the page is not straight)
 - Snug the feed guides!
 - Use deskew settings.
-



Just the Page

- To avoid large black areas around the edge of the page and to avoid measuring, have the scanner determine page size.
-



Mode

- Black and white
 - Best for text
 - Smallest file size
 - Black and white ED (error diffusion)
 - Better for graphics
 - Slightly larger files
 - Color
 - Sometimes necessary
 - Large files
-



Resolution

- DPI—dots per inch
 - Standard for text is 300 dpi
 - Small text may require 400 dpi
 - Thin paper may require 150-200 dpi
 - Really large text may require 200 dpi

 - Infty Reader for math requires 600 dpi
-



Brightness

- Overall darkness or lightness of page
 - Balance
 - Not too dark, not too light
 - Scale 1-255
 - Lower numbers decrease brightness
 - Higher numbers increase brightness
 - High = light
-



Adjusting Brightness

- Default is 128
 - Too dark
 - Letter shapes run together
 - Too light
 - Letter shapes are thin or broken
 - Newsprint type papers may need increased brightness
-



Contrast

- Difference between light and dark on page
 - Scale is 1-13
 - Higher number increases contrast
 - Darks darker, lights lighter
 - Lower number decreases contrast
 - Moves toward middle
 - Darks get lighter, lights get darker
-



Threshold

- In black and white mode
 - Sometimes also just uses brightness scale
 - Sets where gray will be seen
 - Increased threshold adds more white
 - More grays seen as white
 - Decreased threshold adds more black
 - More grays seen as black
-



Color Dropout

- “Erases” color
 - Red, blue, or green
 - Drop out colored highlighters
 - Orange highlighter especially
 - Yellowish pages
 - Drop out red (improves contrast)
 - Watch out for dropping out text, as well
 - Be aware of color that has white text on it
-



Despeckle

- “Erases” speckles
 - Helps with small stray black dots
 - Beware of erasing periods and umlauts
-



Erase Notch

- Can help when scanning documents that have been three-hole punched
-



Gamma...it's complicated...

- Adjusts the middle tones
 - Usually more useful for scanning graphics than text
 - Can be altered to bring out more detail in shadows in photos
 - Changes the calibration of the hardware
 - Try everything else first!
-



Settings Summary

- Brightness = overall tone
 - Contrast = difference in highs and lows
 - Gamma = adjustment in midtones

 - Threshold = on or off switch for grays
 - Grays seen as white or black
 - May appear as just the “brightness” bar
-



Static

- Dryer sheets
 - Add humidity
 - Run humidifier
 - Mist paper lightly with spray bottle
-



Keep It Clean!

- Clean your scanner often!

 - Bits of glue and paper fall into scanner
 - Clean glass

 - Print builds up on rollers
 - Clean rollers and feed guides
-

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>

Table of Contents

Improving Your Scanning.....	1
Scanner Settings.....	1
Mode.....	1
Resolution.....	2
Brightness.....	2
Contrast.....	3
Threshold (Or where did my other bar go?).....	4
Color Dropout.....	4
Gamma.....	6
Despeckle/Remove Dot.....	7
Erase Notch.....	7
Reverse Image.....	7
Static.....	7
Specific Tips.....	8
<i>Activity Summary Sheet.....</i>	8

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

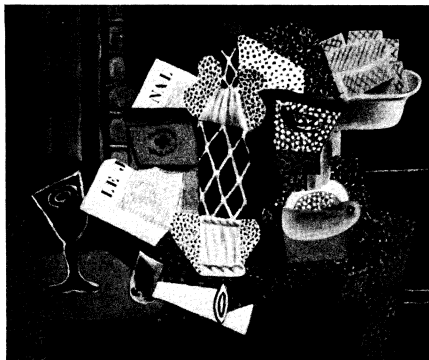
Mode

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.

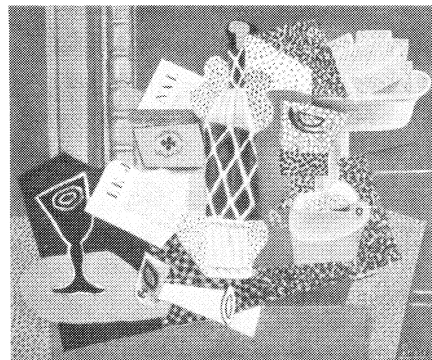
Note that the 9080C now offers "Advanced Text Enhancement" mode to help with very light documents or when text is printed on a dark background.

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)

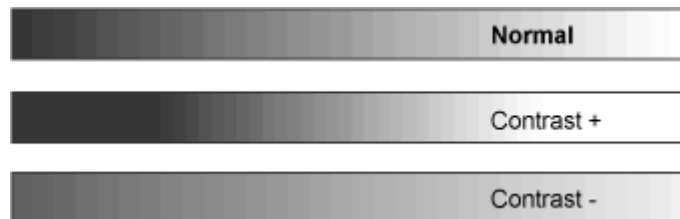
According to OmniPage Help, “Brightness plays an important role in OCR accuracy. After loading an image, check its appearance. If characters are thick and touching, lighten the brightness. If characters are thin and broken, darken it....The diagram [below] illustrates an optimum brightness.”



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.

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]
[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 may be marks that 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. This setting can also help in removing the holes from scans of documents that were three-hole punched.

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 may even find it helpful to mist the paper with a sprayer very slightly.

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.

Specific Tips

Condition	Setting
Black pen	do not compliment thin line, use despeckle (remove dot); possibly increase brightness; use eraser tool in Abbyy FineReader
Blue pen	drop out blue, despeckle (remove dot), possibly increase brightness
Glossy paper	may need to reduce brightness or contrast
Gray boxes	scan in B&W and increase brightness, despeckle (remove dot)
Multiple highlighters	drop out red (many highlight colors will be light enough to be ignored; it is the red-tinted ones that are the most problematic, followed by the blue-tinted ones)
Newsprint	increase brightness, despeckle (remove dot)
Orange highlighter	drop out red
Punched paper	use erase notch
Small text	increase resolution to 400 DPI
Thin paper	reduce resolution to 200 DPI or less; may also need to adjust brightness to remove the shadow that bleeds through from the back side of page
White text on dark background	usually fine as is, but make sure not to drop out the background color the text is on
Yellowish paper	drop out red

Scan the test books, and list below the settings for each situation and what you learned.

Test Books

Test Book Title	Issue	Mode	DPI	Brightness	Contrast	Filters
<i>Social Superstitions</i>	white					
<i>The Winter Room</i>	green					
<i>American Heritage Dictionary</i>	gray					
<i>Carpal Tunnel Syndrome</i>	orange					
<i>Racing Vacation</i>	yellow					
<i>Nursing Drug Guide</i>	red					
<i>CSS Web Design</i>	gray boxes on nonglossy					
<i>New Psychology of Women</i>	gray boxes on glossy					
<i>Farmers Almanac</i>	newsprint					
<i>College catalog</i>	newsprint					

Activity Summary Sheet

Scan the test books, and list below the settings for each situation and what you learned.

Test Book	Mode	DPI	Brightness	Contrast	Filters	Tips
Blue pen						
Colored text on colored background						
Glossy paper						
Gray boxes on glossy paper						
Gray boxes on matte paper						
Highlighter: blue						
Highlighter: green						
Highlighter: orange						
Highlighter: yellow						
Multi-colored books						
Newsprint						
Thin paper						
Yellowish paper						

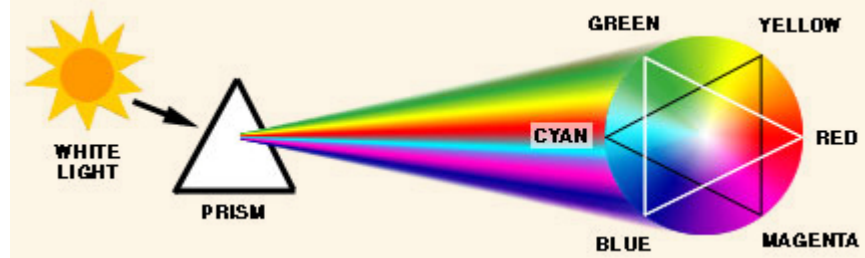
Scanner Settings

Situation (paper, color, etc.)	Mode	DPI	Brightness	Contrast	Dropout*	Other Filters Applied

* In the RGB color system, red, green, and blue are the three basic colors used to create all colors: green + blue = cyan; red + blue = magenta; green + red = yellow

RGB World - Understanding Color

www.rgbworld.com/color.html#add



What is Color?

Color is all around us. It is a sensation that adds excitement and emotion to our lives. Everything from the cloths we wear, to the pictures we paint revolves around color. Without color; the world (especially RGB World) would be a much less beautiful place. Color can also be used to describe emotions; we can be red hot, feeling blue, or be green with envy.

In order to understand color we need a brief overview of light. Without light, there would be no color, and hence no RGB World. Thank God for light!

Light is made up of energy waves which are grouped together in what is called a spectrum. Light that appears white to us, such as light from the sun, is actually composed of many colors. The wavelengths of light are not colored, but produce the sensation of color.

Visible light - The wavelengths our eyes can detect is only a small portion of the electromagnetic energy spectrum. We call this the visible light spectrum. At one end of the visible spectrum are the short wavelengths of light we perceive as blue. At the other end of the visible spectrum are the longer wavelengths of light we perceive as red. All the other colors we can see in nature are found somewhere along the spectrum between blue and red. Beyond the limits at each end of the visible spectrum are the short wavelengths of ultraviolet light and X-rays and the long wavelengths of infrared radiation and radio waves, which are not visible to the human eye.

Primary Colors

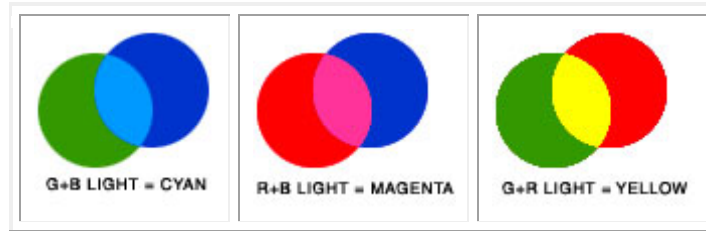


If the visible portion of the light spectrum is divided into thirds, the predominant colors are red, green and blue. These three colors are considered the primary colors of the visible light spectrum.

Primary colors can be arranged in a circle, commonly referred to as a color wheel. Red, green and blue (RGB) form a triangle on the color wheel. In between the primary colors are the secondary colors, cyan, magenta and yellow (CMY), which form another triangle.

The media and methods used to reproduce color include color paintings, printing presses, color film, color monitors, color printers, etc. There are only two basic ways, however, of reproducing color... additive and subtractive.

Additive Color System (RGB)



The additive color system involves light emitted directly from a source, before an object reflects the light. The additive reproduction process mixes various amounts of red, green and blue light to produce other colors. Combining one of these additive primary colors with another produces the additive secondary colors cyan, magenta, yellow. Combining all three primary colors produces white.

Television and computer monitors create color using the primary colors of light. Each pixel on a monitor screen starts out as black. When the red, green and blue phosphors of a pixel are illuminated simultaneously, that pixel becomes white. This phenomenon is called additive color.

To illustrate additive color, imagine three spotlights, one red, one green and one blue focused from the back of an ice arena on skaters in an ice show. Where the blue and green spotlights overlap, the color cyan is produced; where the blue and red spotlights overlap, the color magenta is produced; where the red and green spotlights overlap the color yellow is produced. When added together, red, green and blue lights produce what we perceive as white light.



As mentioned before, television screens and computer monitors are examples of systems that use additive color. Thousands of red, green and blue phosphor dots make up the images on video monitors. The phosphor dots emit light when activated electronically, and it is the combination of different intensities of red, green and blue phosphor dots that produces all the colors on a video monitor. Because the dots are so small and close together, we do not see them individually, but see the colors formed by the mixture of light. Colors often vary from one monitor to another. This is not new information to anyone who has visited an electronics store with various brands of televisions on display. Also, colors on monitors change over time. Currently, there are no color standards for the phosphors used in manufacturing monitors for the graphics arts industry.

All image capture devices utilize the additive color system to gather the information needed to reproduce a color image. These devices include digital cameras, flatbed scanners, drum scanners, and video cameras.

To summarize: Additive color involves the use of colored lights. It starts with darkness and mixes red, green and blue light together to produce other colors. When combined, the additive primary colors produce the appearance of white.

Microsoft Document Imaging

To install the Microsoft Office Document Imaging program, follow these steps:

1. Click **Start**, click **Run**, type **appwiz.cpl**, and then click **OK**.
2. In the **Currently installed programs** list, click the 2007 Office version that you have installed.
3. Click **Change**.
4. Click **Add or Remove features**, and then click **Continue**.
5. Expand **Office Tools**.
6. Click **Microsoft Office Document Imaging**, and then click **Run all from My Computer**.
7. Click **Continue**.