Scanning 101

George Soules
October 29, 2018

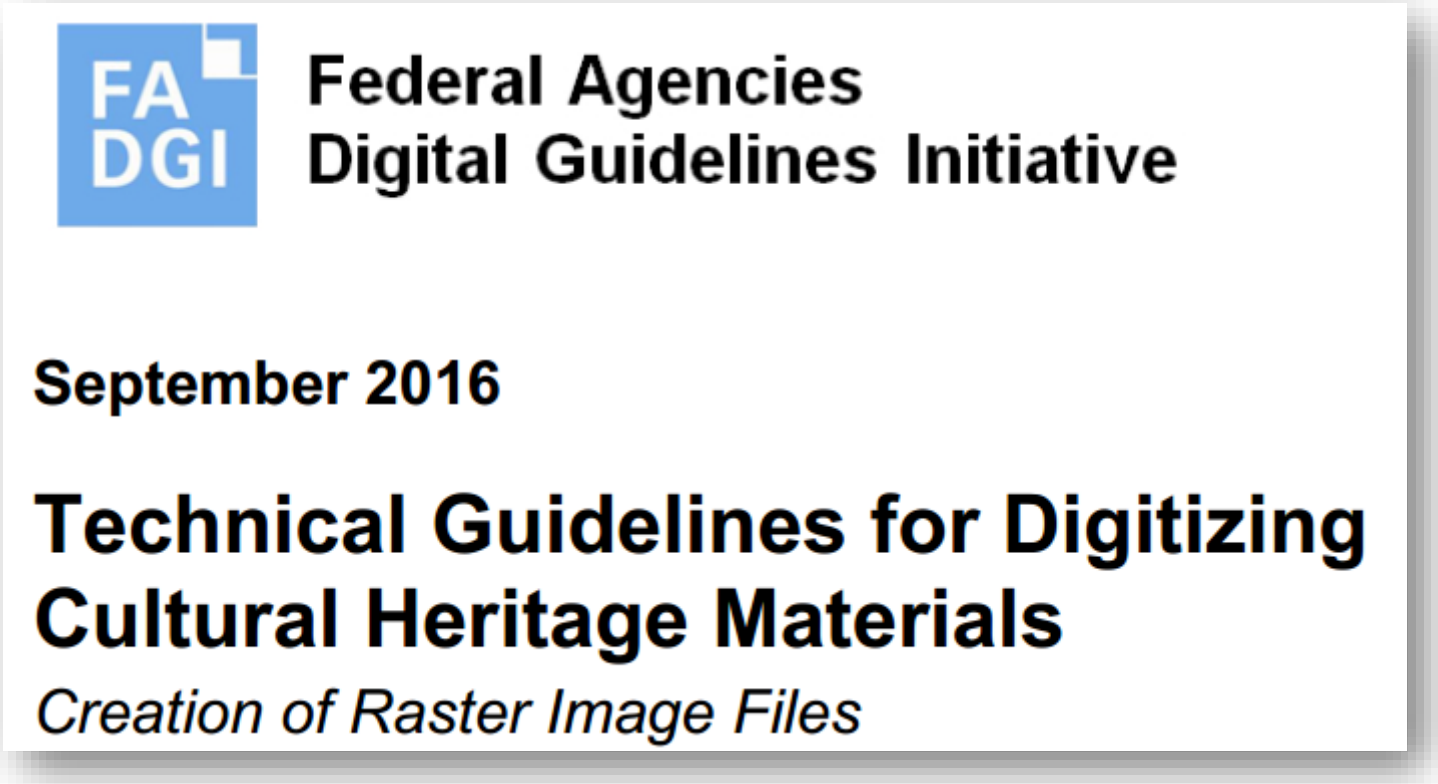
Goals of this workshop

- Understand the fundamentals of digitization technology (and why it's so confusing)
- Become aware of your organization's digitization weaknesses and what it would take to strengthen them
- Learn to produce scans that meet your digitization needs for 2D reflective items

What are your digitization needs?

- ★Thumbnail to identify the original item
- ★★Visual record with some informational value, but not suitable for printing or OCR
- ★★★Professional image good enough for most uses
- ★★★★State of the art image capture suitable for use as a Digital Surrogate

The FADGI Star System



digitizationguidelines.gov/guidelines/digitize-technical.html

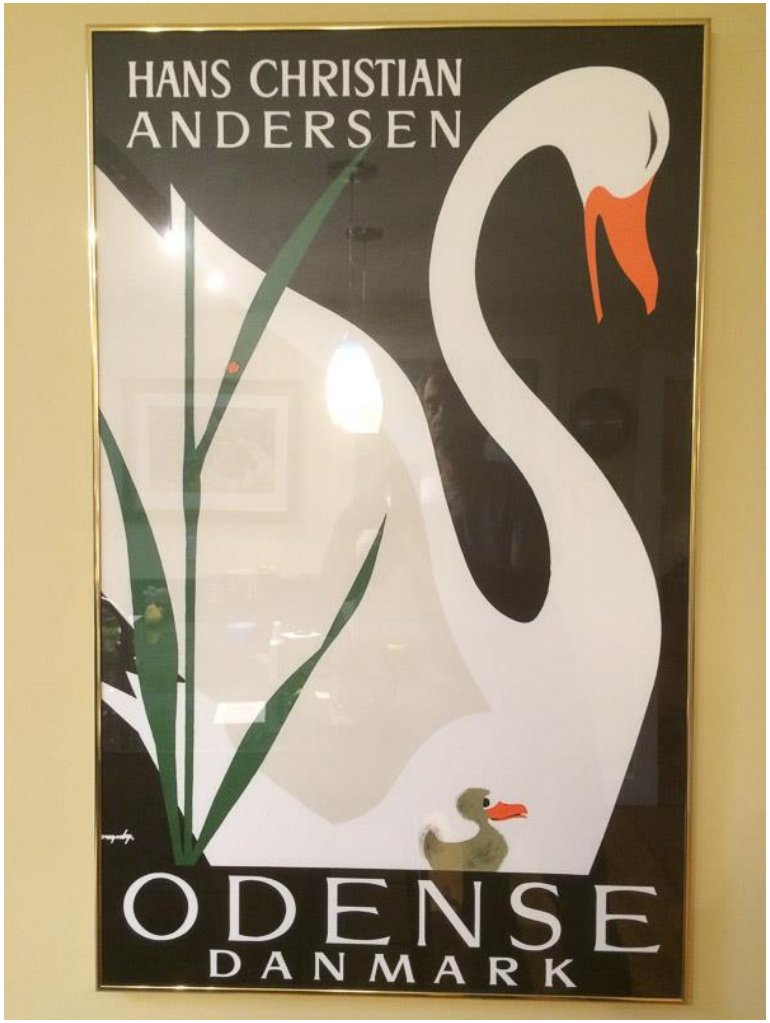
FADGI guidelines for photographs

"The *Guidelines* are intended to be informative, not prescriptive."

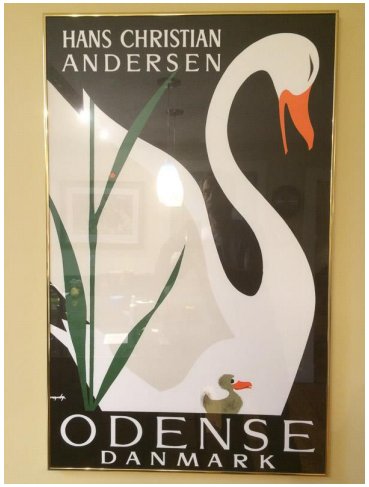
	1 Star	2 Star	3 Star	4 Star
Master File Format	TIFF	TIFF	TIFF	TIFF
Access File Formats	All	All	All	All
Resolution	100 ppi	200 ppi	400 ppi	600 ppi[1]
Bit Depth	8	8	8 or 16	16
Color Space	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRGBv2	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRGBv2	Adobe 1998 ProPhoto, ECIRGBv2	Adobe 1998 ProPhoto, ECIRGBv2
Color	Grayscale or Color	Grayscale or Color	Color	Color
Measurement Parameters				
Tone Response (OECF) (Luminance)	± 9 count levels ≤ 8	± 7 count levels ≤ 6	± 5 count levels ≤ 4	± 3 count levels ≤ 2
White Balance Error (Luminance)	± 8 counts ≤ 8	± 6 counts ≤ 6	± 4 count levels ≤ 4	± 3 count levels ≤ 2
Illuminance Non-Uniformity	<8%	<5%	<3%	<1%
Color Accuracy (Mean ΔE 2000)	<10	<6	<4	<2
Color Channel Misregistration	<1.2 pixel	<.80 pixel	<.50 pixel	<.33 pixel
MTF10 (10% SFR)	sampling efficiency > 60% and SFR response at half sampling frequency < 0.4	sampling efficiency > 70% and SFR response at half sampling frequency < 0.4	sampling efficiency > 80% and SFR response at half sampling frequency < 0.3	sampling efficiency > 90% and SFR response at half sampling frequency < 0.2
MTF50 (50% SFR)	50% of half sampling frequency: [25%,95%]	50% of half sampling frequency: [30%,85%]	50% of half sampling frequency: [35%,75%]	50% of half sampling frequency: [40%,65%]
Reproduction Scale Accuracy	<+/- 3% of AIM	<+/- 3% of AIM	<+/- 2% of AIM	<+/- 1% of AIM
Sharpening (Maximum MTF)	<1.3	<1.2	<1.1	<=1.0
Noise ΔL* St. Dev (Luminance)	>6 count levels < 4	>5 count levels < 3	>4 count levels < 2	>3 count levels < 1

digitizationguidelines.gov/guidelines/digitize-technical.html

★★ Visual record



★★★ Professional image



★★★★ Digital Surrogate

- A digital reproduction of an original material object
- Serves as a surrogate for the material object
- *As good as* the original to the extent that it can be used in place of the original
- Can be used to create a Material Surrogate
- You have to define "as good as" for your organization

dlib.org/dlib/march13/jacobs/03jacobs.html

Material Surrogate

Theory versus practicality

- Creating a true Digital Surrogate is a nearly impossible goal to achieve ★★★★+
- Creating a professional digital image is entirely possible, when you know how ★★★
- Anyone with an iPhone can produce 2-star digital images ★★
- What's acceptable for your organization?

Why are you scanning?

- Produce Digital Surrogates for preservation purposes ★★★★
- Create the best reproductions that our resources and budget will allow ★★★
- Provide information for research and educational purposes ★★½
- Show people what's in the collection to entice them to visit the building ★★
- Make a visual index of collection content ★

A practical approach

- Choose a star level on an item by item basis
- Aim for 4 stars for your most rare and valuable items
- Follow 3-star guidelines as practical for the rest
- Use 2 stars for low value items

Types of scanners

Sheetfed

Flatbed

Book

Film

Flatbed

Drum

Flatbed scanner sizes and prices

A4 / Letter – 8.5” x 11.7”		
Epson V600	6400 dpi	$170
Epson V850	6400 dpi	$884
A3 / Tabloid Size – 12.2” x 17.2”		
Plustek Optic Pro A320	1600 dpi	$435
Epson Expression 12000XL	2400 dpi	$3,422
A2 (18” x 24”)		
Contex IQ Flex	1200 dpi	$6,140
Large Format – 24” x 36”		
Kurabo ARCH-D	800 dpi	$50,380

Large format alternatives

- Stitching
- Camera scanning

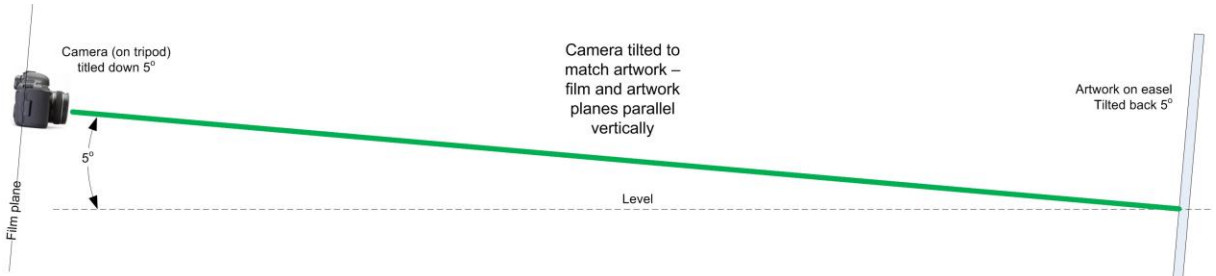
Stitching

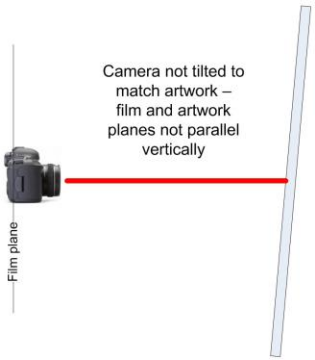
Archival Masters

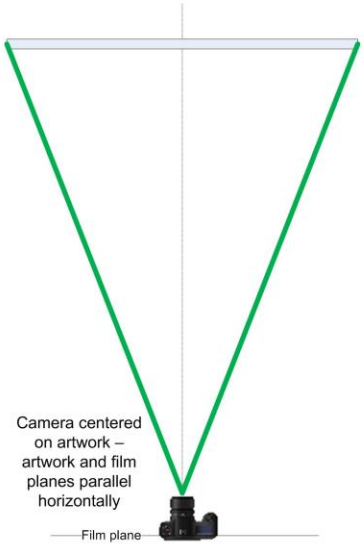
Production Master

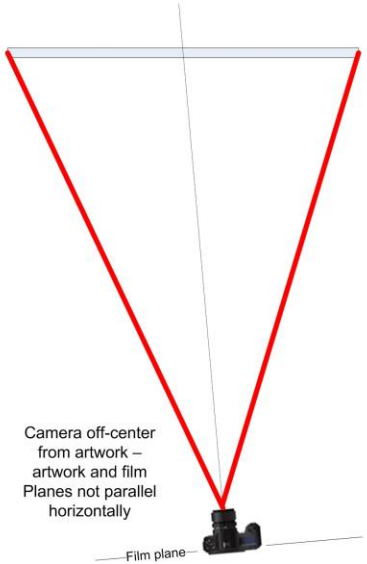
Derivative Images

Camera scanning

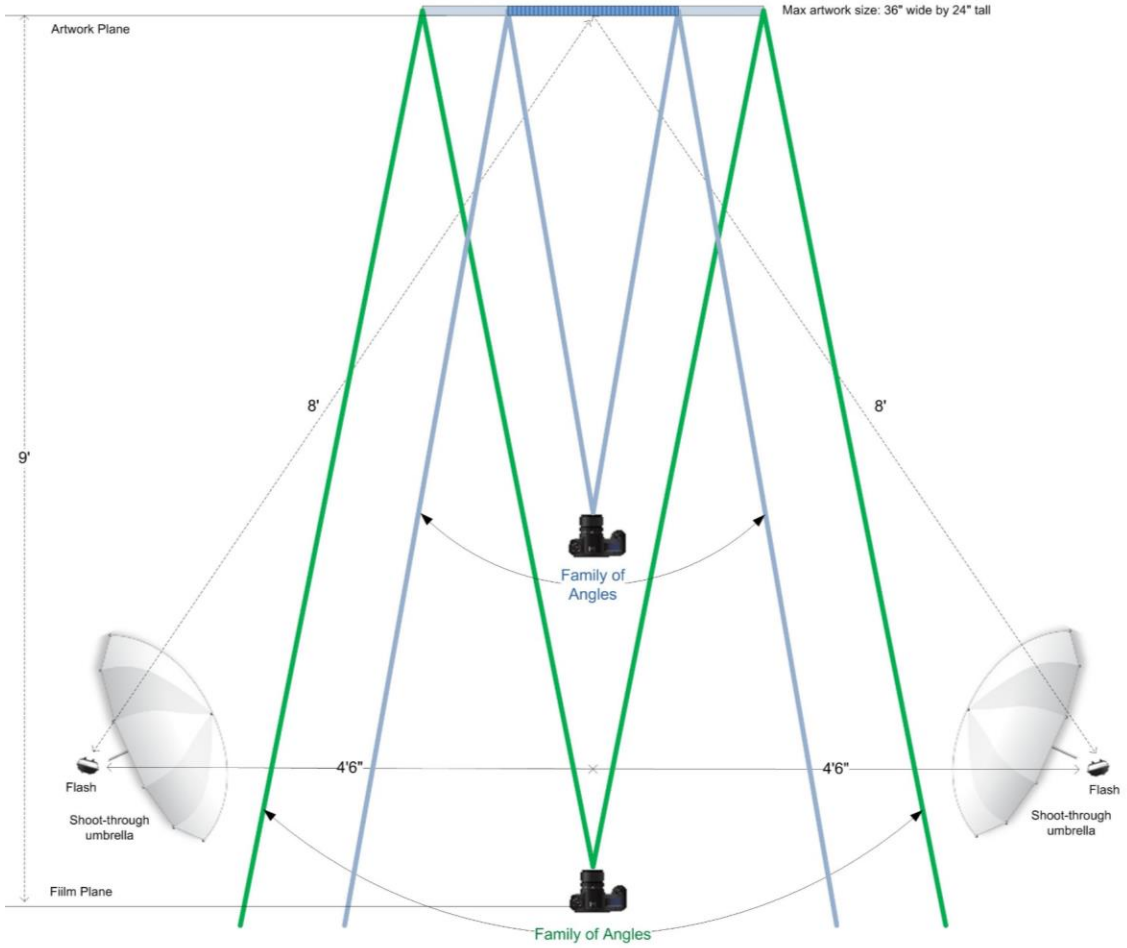




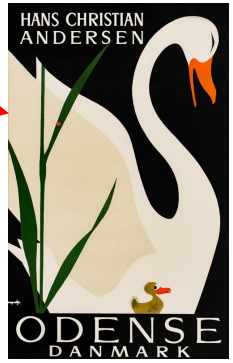




Camera scanning



Camera scanning



Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Choose which items to scan

- Available only briefly (items loaned to you)
- Rare, fragile, or valuable (preserve ASAP)
- Needed soon e.g. for an exhibit or publication
- Based on your Collection Management Policy

You don't have to scan everything. Choose the best from a set of related photos or a few representative pages from a multi-page document. You can also consider performing 1-star scanning on the rest.

Scanning approaches

- Process one item at a time end-to-end
  - Add item to Digital Archive without an image
  - Scan the item and use Digital Archive identifier in output file name
  - Upload the low res image(s) to the Digital Archive
- Add items to Digital Archive first, then scan and upload images later
- Scan items first, then add items and upload images to the Digital Archive later

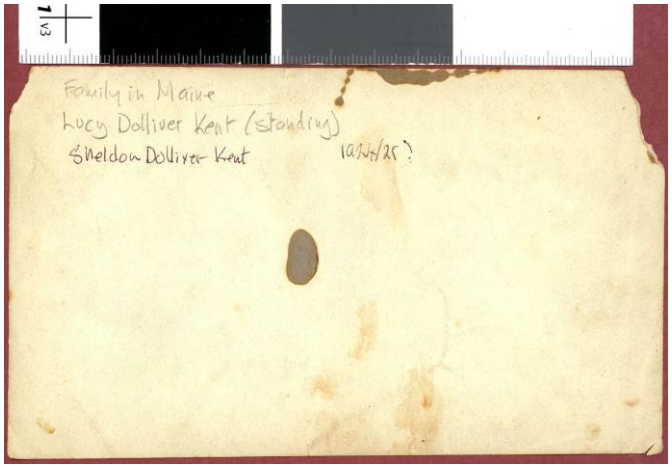
Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Decide parts of item to scan

- Front with small border visible all around
- Back if shows important information
- Mount or border *if* shows important information
- Interior pages – all or only relevant pages
- Scan less important information at lower resolution

600 ppi
28 MB



300 ppi
7 MB

Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Scanning environment

- Stable surface
- Dust-free environment
- Constant temperature
- Plenty of room for items to be scanned
- Good light for viewing computer screen

Prepare the item for scanning

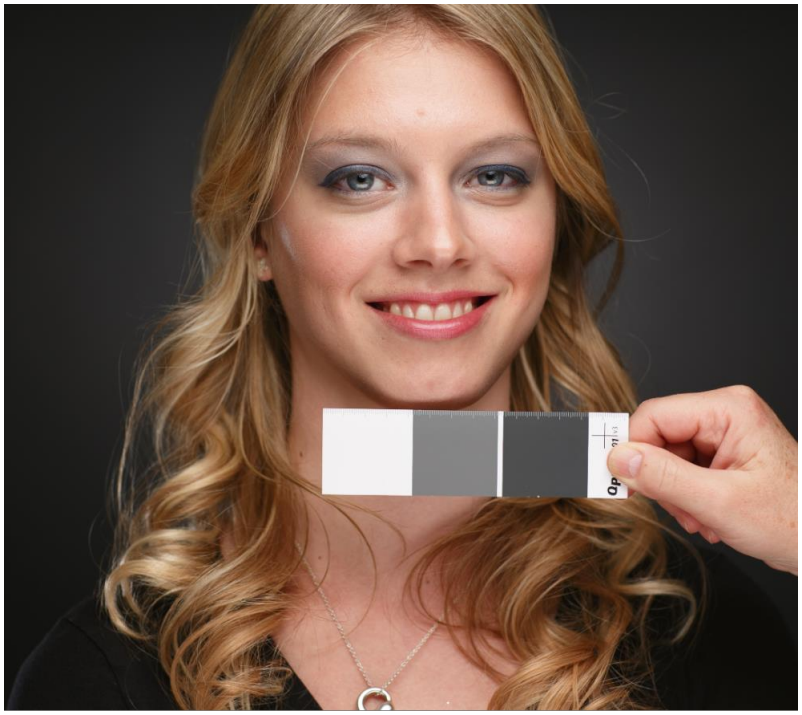
- Put on white gloves
- Clean scanner
- Disassemble item if necessary to lay flat
- Lay item on scanner *within scanning area* which may be smaller than glass platen
- Add white or colored backing if needed
- Include a reference target to use for color correction

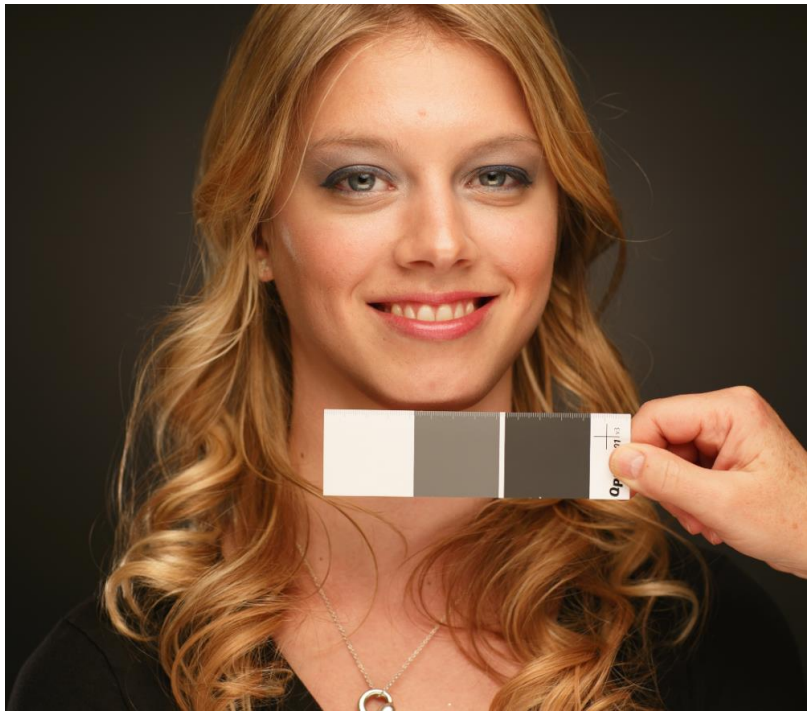
Color Correction

- Every scanner and camera captures color differently
- Every monitor, projector, and printer displays color differently
- We can't trust our eyes to determine correct color

Color Correction

A grayscale or color reference target allows us to correct color “by the numbers” to ensure that an image does not have an unwanted color cast.





Color Correction

- Colorimeter for monitor calibration
  - X-rite i1Display Pro $249
  - X-rite ColorMunki Display $169
  - Datacolor Spyder5PRO $149
- Reference Targets
  - QpCARD 101 neutral gray reference card $6
  - X-Rite ColorChecker Passport $149
  - Others
- Monitor that can be calibrated $500+
- Image editing software to adjust color based on the reference target

Scanning tools

- Lint-free wipes (PEC PADS)
- Lint-free cotton gloves (oil from fingers attracts dust and dirt and damages photos)
- Air blower
- Brush
- White or dark paper as backing when bleed-through or holes/tears

Bleed through

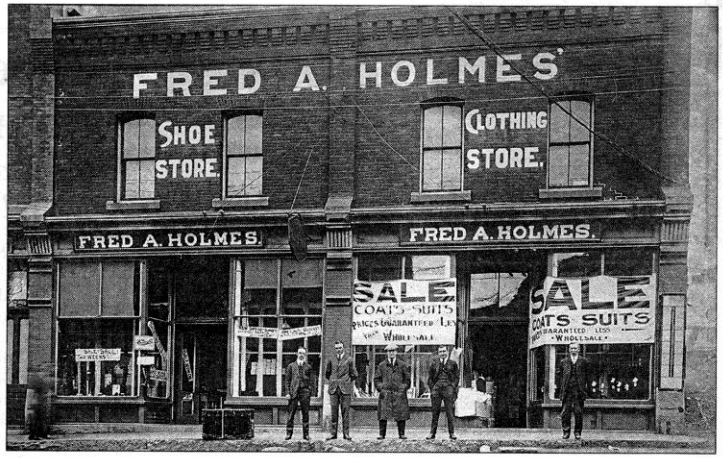
- Occurs when the back of a page, or the next page, shows through the scan
- Use white paper between one-sided pages to prevent next page from bleeding through
- Use black paper when print on back of page bleeds through

Bleed through

WWW.WORKINGWATERFRONT.COM · NOVEMBER 2018 9

RESTORING STOREFRONTS—

This circa-1920 photo shows Eastport storefronts in a building constructed in 1887 and designed by architect Henry Black of Boston and St. John, New Brunswick and now part of a National Register historic district. In late November renovation work will begin on the building and continue through April. Funds are being sought by Eastport's Tides Institute to offset costs.

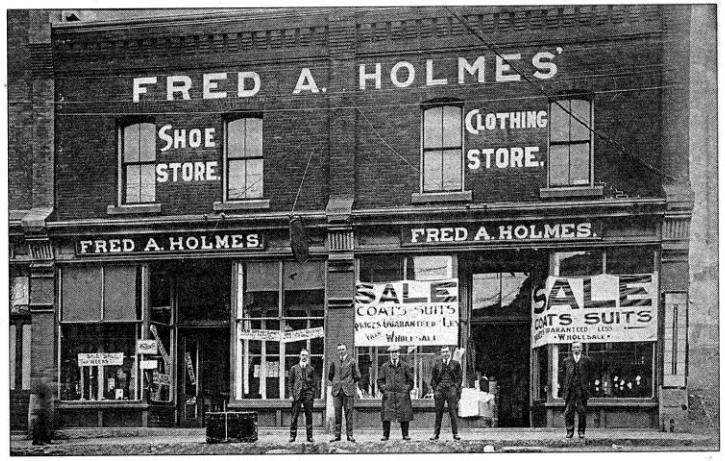


No backing

WWW.WORKINGWATERFRONT.COM · NOVEMBER 2018 9

RESTORING STOREFRONTS—

This circa-1920 photo shows Eastport storefronts in a building constructed in 1887 and designed by architect Henry Black of Boston and St. John, New Brunswick and now part of a National Register historic district. In late November renovation work will begin on the building and continue through April. Funds are being sought by Eastport's Tides Institute to offset costs.



Black backing

Scanning workflow

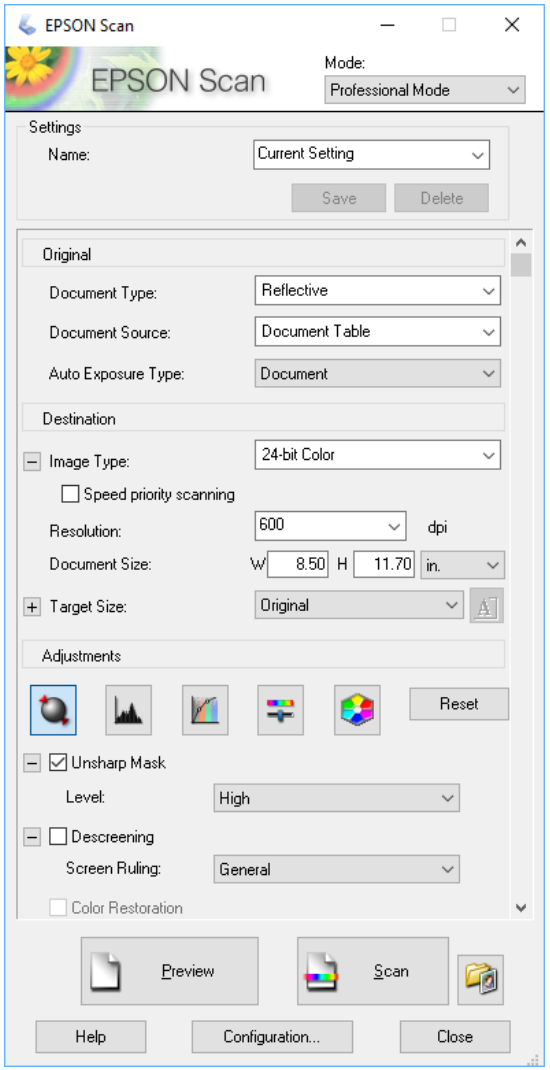
- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Software

- Scanning software
  - Epson Scan (free)
  - SilverFast $49 - $598 (consider if scanning film)
  - VueScan $50 - $100
  - Abbyy Finereader $200 (consider if OCR is very important)
  - Others
- Image editing software
  - Adobe Photoshop +
    Adobe Lightroom $10/month
  - Adobe Photoshop Elements $100 (limited support for 16 bit color)
  - Corel Paintshop Pro $55
  - Affinity Photo $50
  - Others

Choose scanner settings

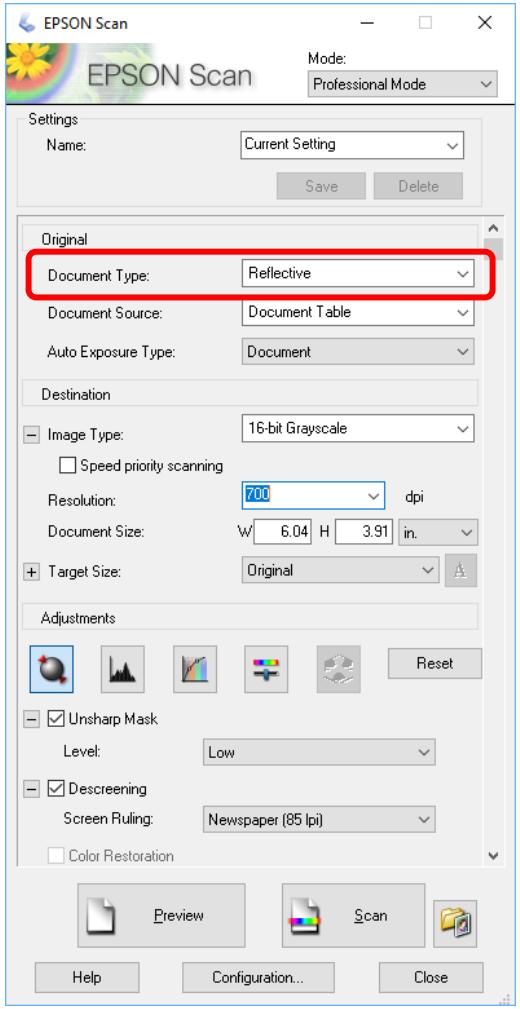
- Document type
- Image type
- Bit depth
- Resolution
- Sharpening
- Descreening
- Output file format
- File name



Document type

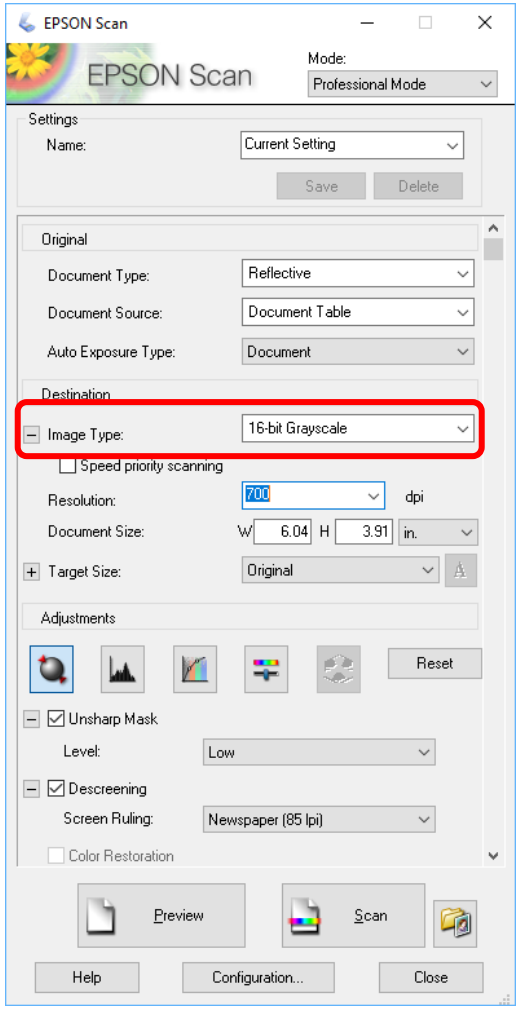
- Reflective
- Film
  - Roll or sheet film negatives (color or B&W)
  - 35mm slides
  - Glass plate negatives
  - Lantern slides

When scanning film, remove the cover over the top platen!



Bit depth

- A measure of pixel accuracy
- Every pixel is represented by a number of bits (a 1 or a 0)
- Bigger numbers – the bit depth – provide more accuracy (up to a point)
- Bit depths range from 1 - 48



A Pixel is the Smallest Element of a Digital Image

(kind of like an atom)

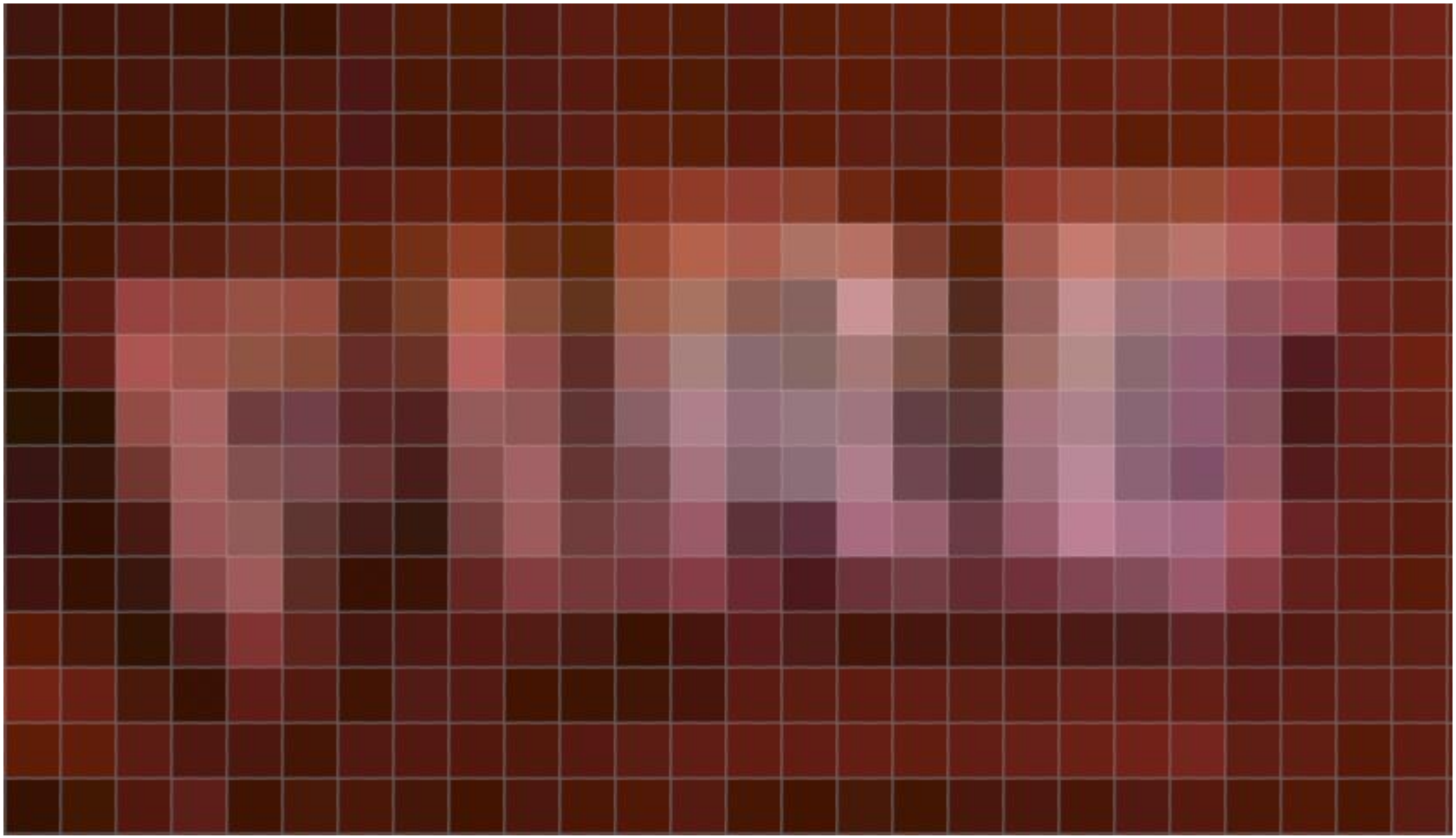
One is the loneliest number

Photograph of a fire alarm taken with a 1 pixel camera

Two can be as bad as one

(it's the loneliest number since the number one)

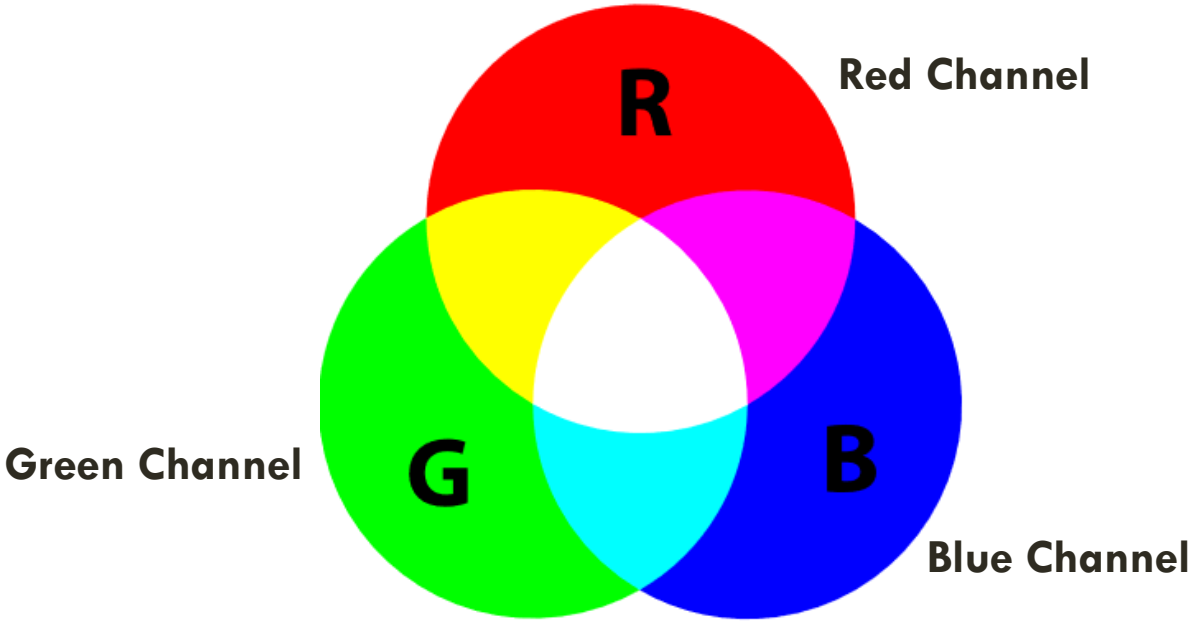
Put enough pixels together…



. . . and a picture starts to form

11,778 x 5,860 pixels = 69,019,080 pixels

Each pixel has three color components called channels



Each channel has 256 tonal values from black to white in shades of gray

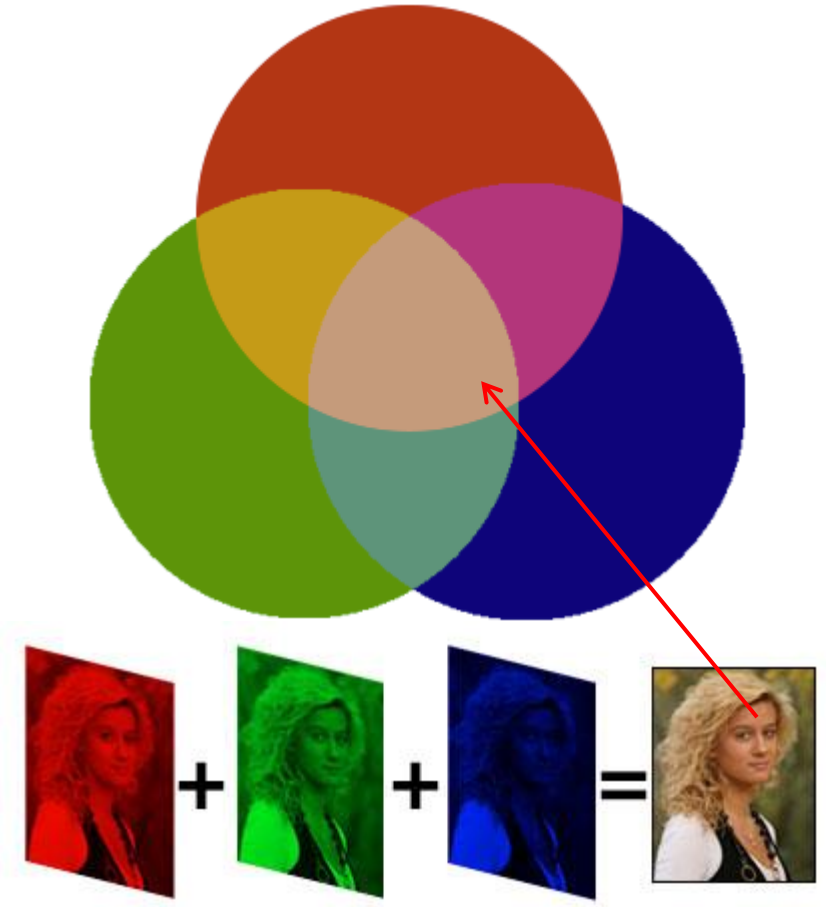
256 × **256** × **256** = 16,777,216 colors

This is for 24-bit color. 48-bit color provides 281 trillion colors!

Scientists think the human eye can discern 10 million colors

docs.gimp.org/en/glossary.html

The individual channel values determine a specific color

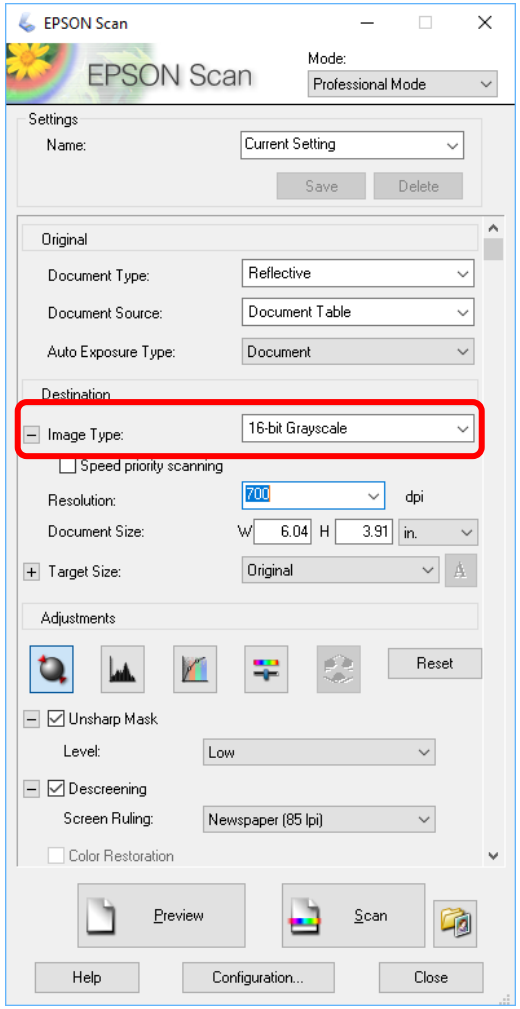


Combine the channels to get a full color image

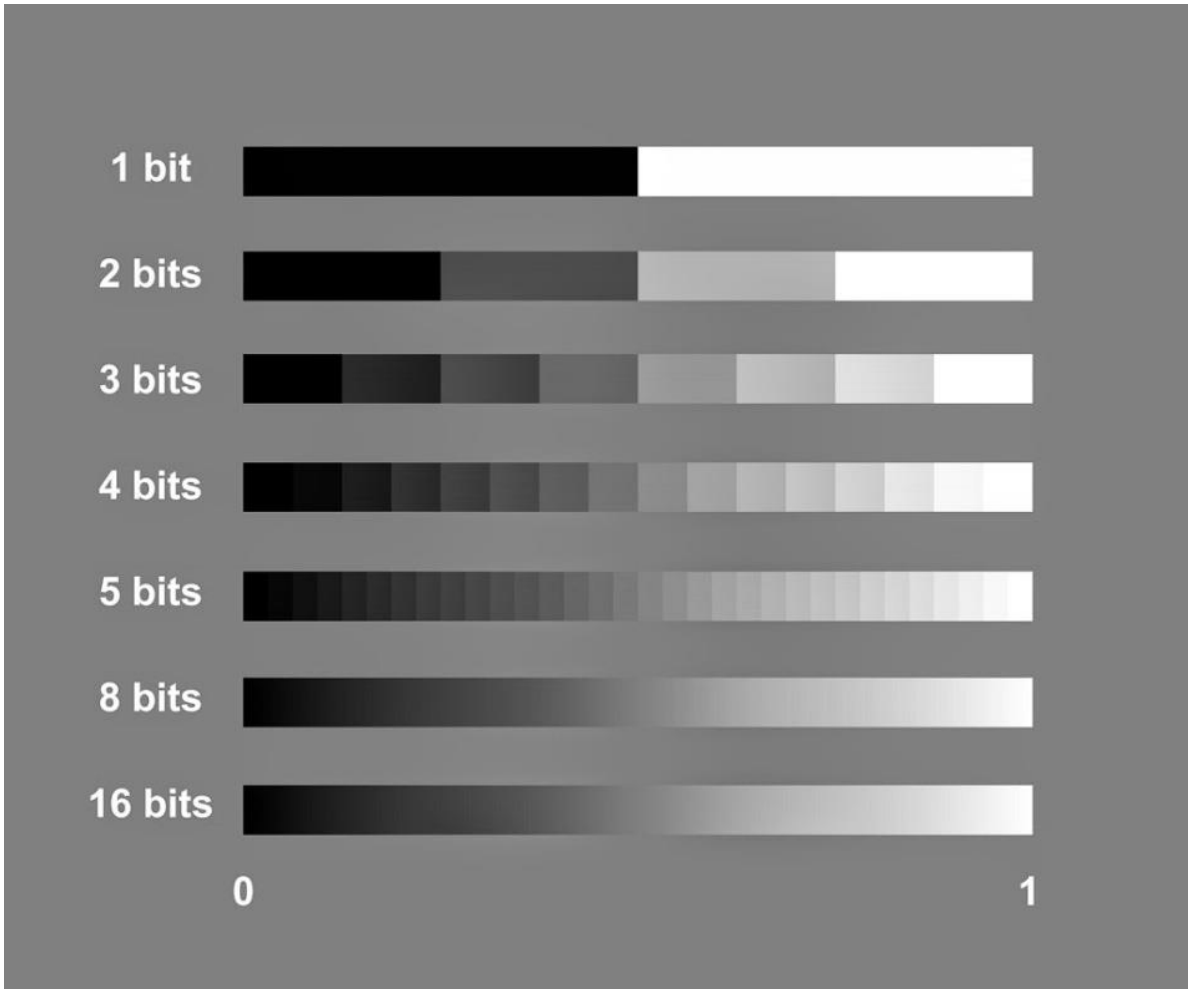
docs.gimp.org/en/glossary.html
adsell.com/scanning101.html

Bit depth

- 24-bit color means 8 bits for each RGB color channel (16.8 million colors) 48-bit mean 16 bits for each channel (281 trillion)
- 8-bit Grayscale is good enough, but choose16-bit
- 1-bit only suitable for line art

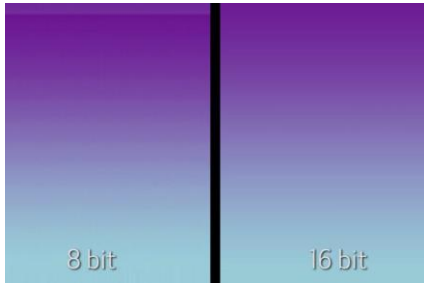


Bit depth – Gray scale



azooptics.com/Article.aspx?ArticleID=1151

Bit depth – Color



8 bits per channel (bpc) = 24 bits per pixel (bpp)
16 bit per channel = 48 bits per pixel
Grayscale has only 1 channel and so bpc = bpp

thephoblographer.com/2017/07/13/does-it-even-matter-8-bit-vs-16-bit-color-depth/
diyphotography.net/8-bit-vs-16-bit-color-depth-use-matters/

Bit depth

48 bit color

24 bit color

File Sizes

Queen Mary 16 bit grayscale 600 dpi 001.tif	26,423 KB
Queen Mary 8 bit grayscale 600 dpi 002.tif	13,212 KB
Queen Mary 1 bit grayscale 600 dpi 003.tif	1,652 KB
Queen Mary 48 bit color 600 dpi 004.tif	79,222 KB
Queen Mary 24 bit color 600 dpi 005.tif	39,629 KB

16 bit grayscale

8 bit grayscale

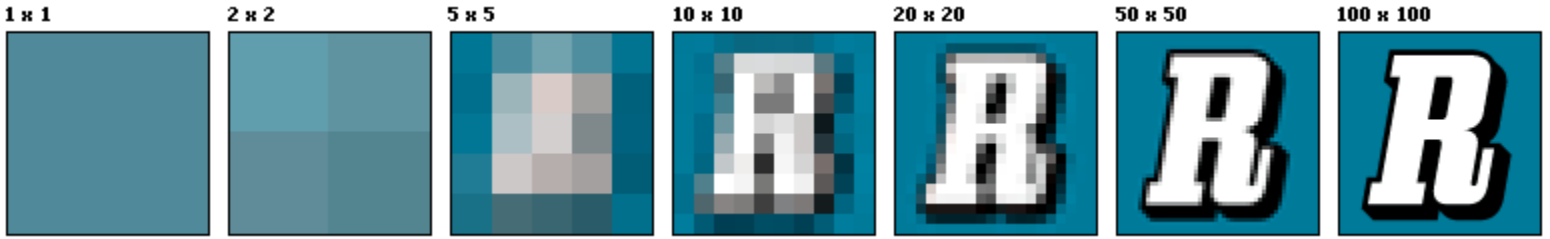
1 bit Black & White

Resolution

- Scanning resolution – samples per inch (spi)
  - Doubling e.g. from 300 to 600 *quadruples* the number of pixels produced (and file size)
- Tonal resolution – bits per pixel (bpp)
- Image Resolution – pixels per inch (ppi)
- Print Resolution – dots per inch (dpi)

Pixel = Pix + Element

A digital photograph is formed from a grid of picture elements arranged in rows and columns



A megapixel (MP) is a million pixels

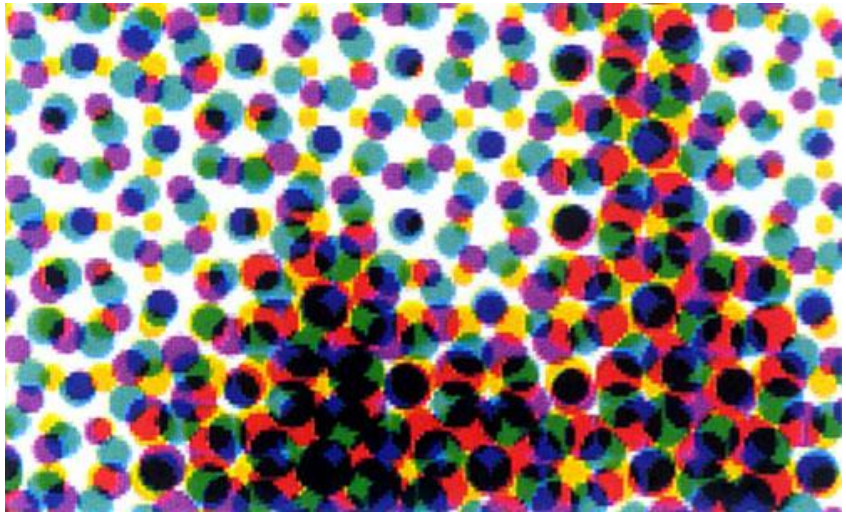
Confusing and Often Misused

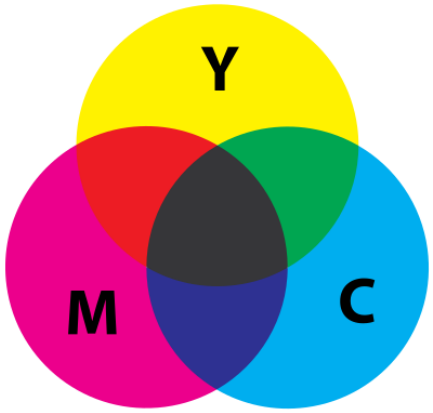
Pixel Terminology

DPI vs. PPI

DPI = Dots Per Inch

Applies to paper prints, *not* digital files





CMYK - Subtractive Color

davebritzius.com/photoshop-101-lesson-4-colour-printing-and-cmyk
toptenreviews.com/computers/printers/best-wide-format-inkjet-printers/

PPI = Pixels Per Inch

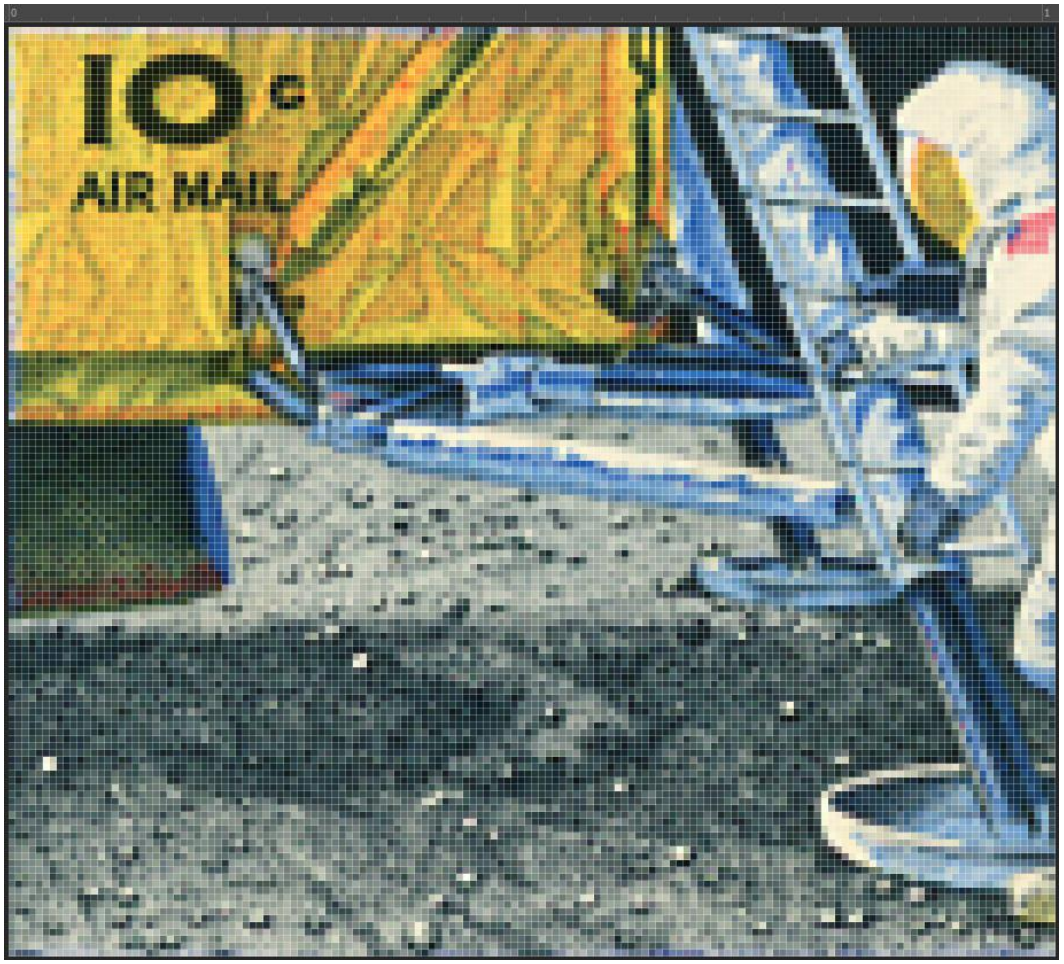
1”

50 ppi

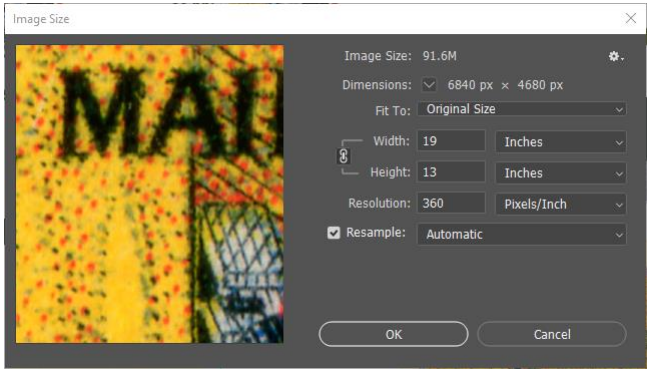
PPI = Pixels Per Inch

1”

150 ppi



Resolution

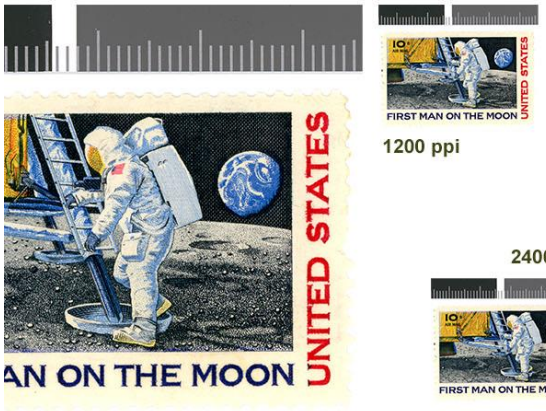


360 ppi
3 cm x 5 cm

50 ppi

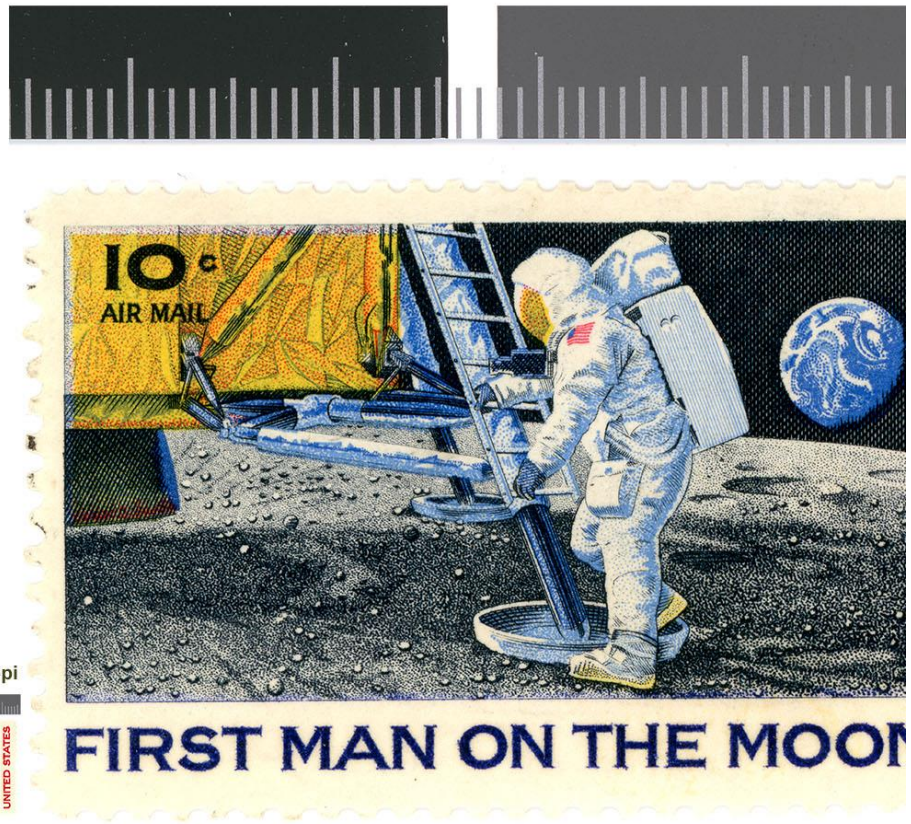
150 ppi

600 ppi



1200 ppi

2400 ppi



Pixilation

Digital file

Printed

Resolution

300 ppi
800%
6 MB

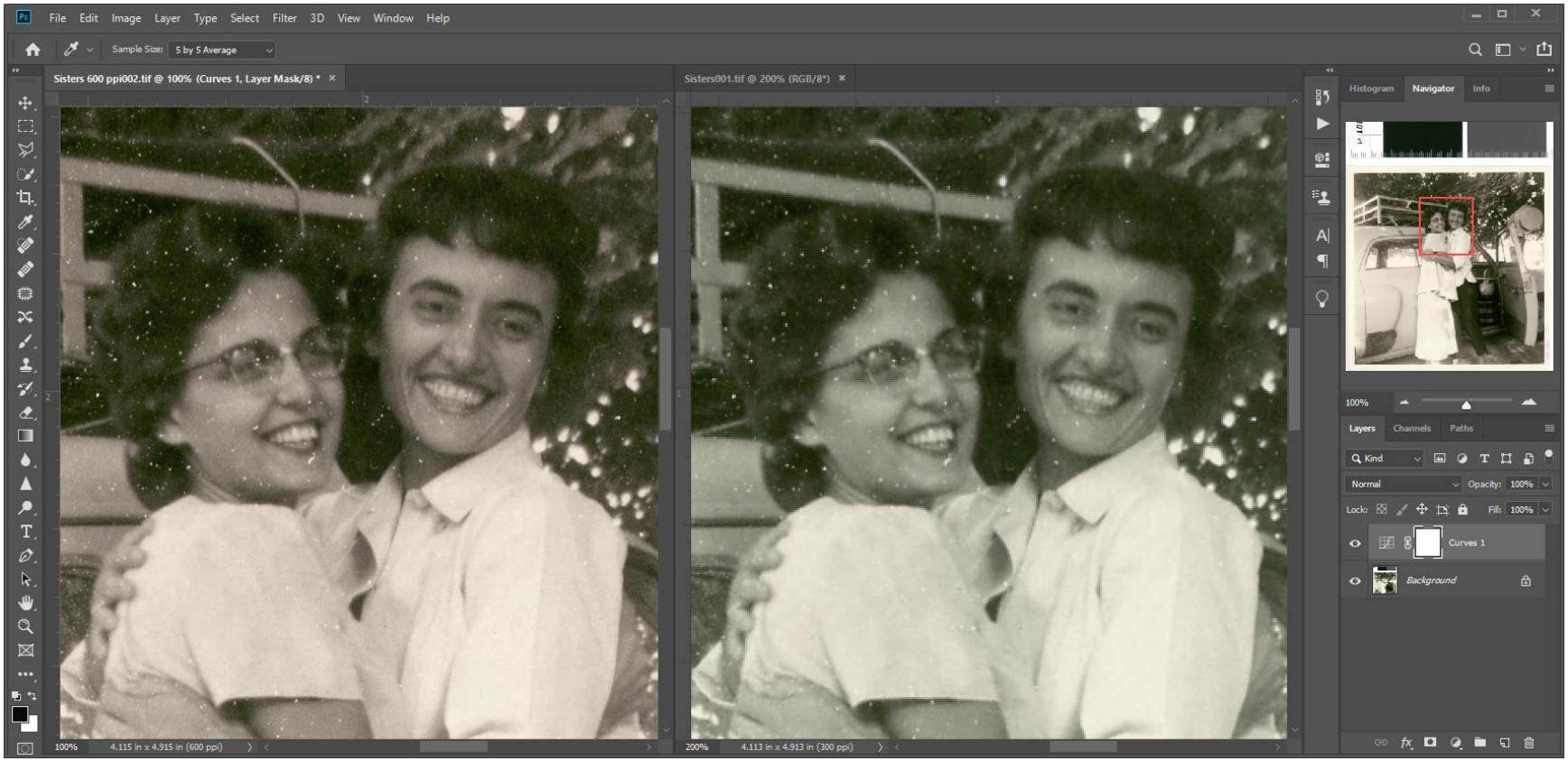
600 ppi
400%
25 MB

1200 ppi
200%
101 MB

2400 ppi
100%
405 MB

swhplibrary.net/digitalarchive/items/show/11449

Low resolution original

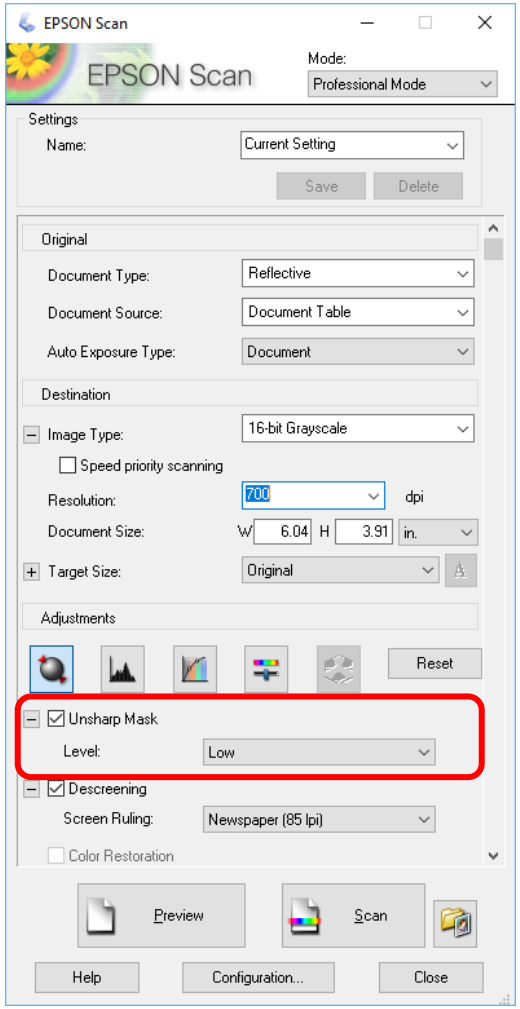


600 ppi
(color corrected)

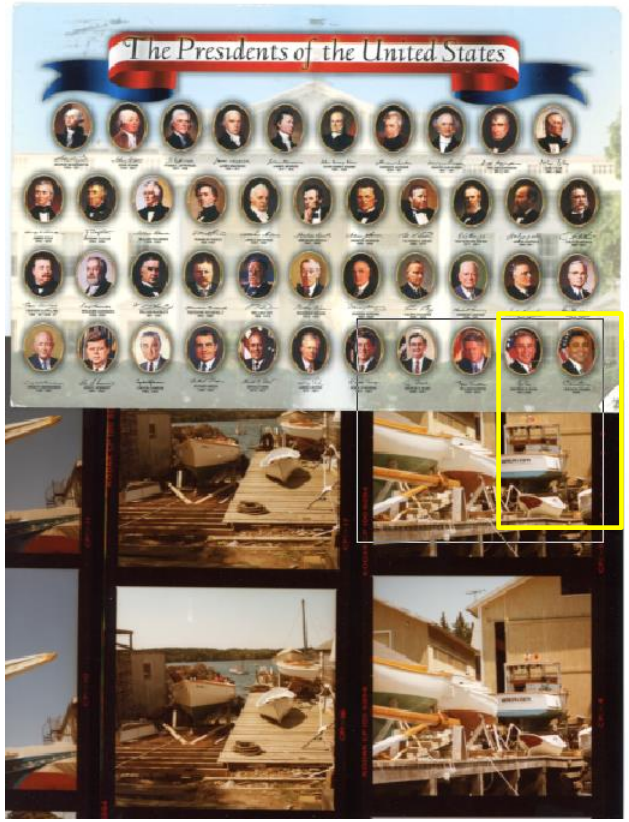
300 ppi

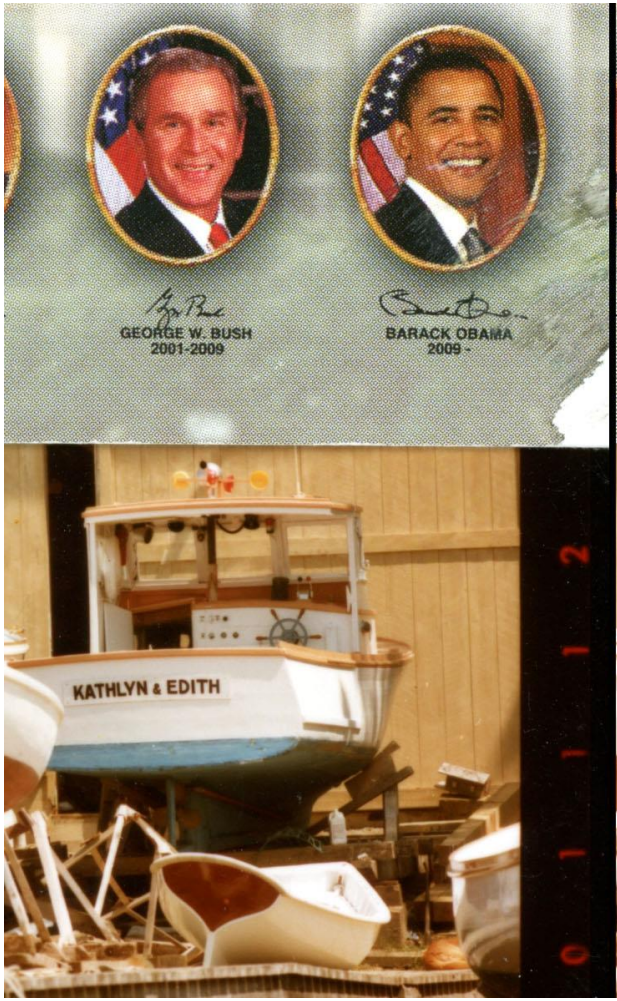
Sharpening

- Most images require some level of sharpening to produce an accurate digital rendition
- For 3-star, okay to use *medium* Unsharp Mask in scanner settings
- For 4-star, perform sharpening on Production Master

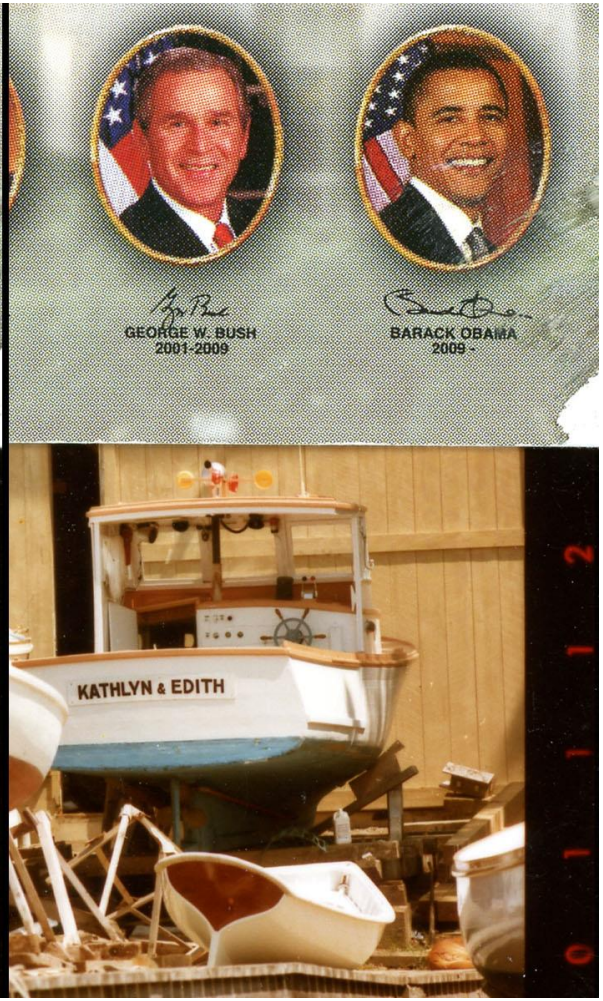


Sharpening





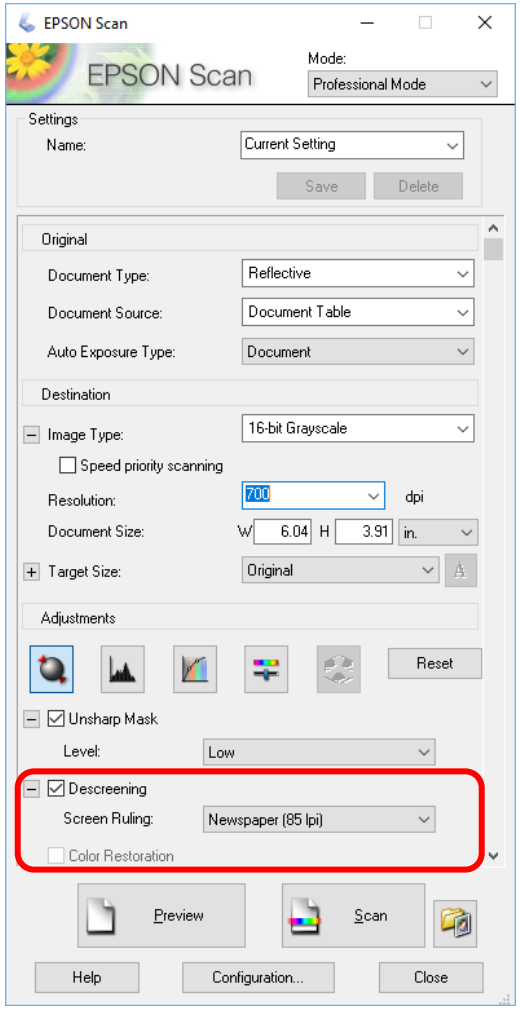
No sharpening



Medium sharpening

Descreening

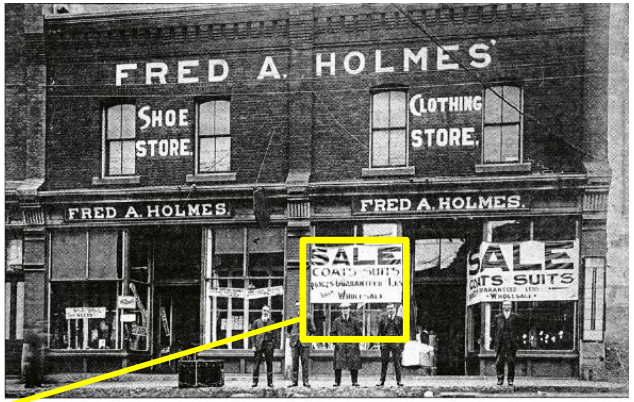
- For 3-star, leave it unchecked, otherwise Archival Master won't contain halftone dots
- Perform descreening when creating Production Master
- Salttva Descreen plug-in for Photoshop $17 - $70
  (trial works indefinitiely for images < 2000 px)



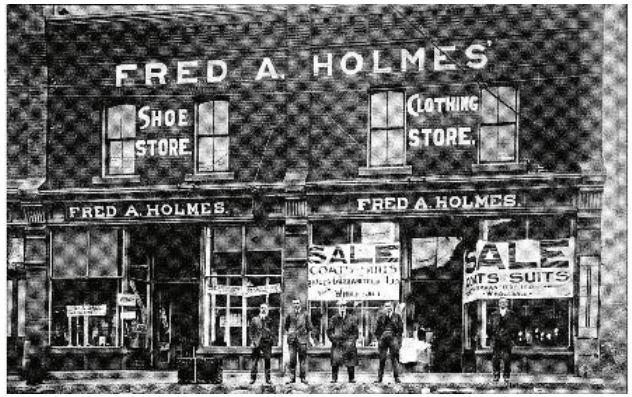
descreen.net/eng/soft/descreen/descreen.htm

Descreening

Halftone dots



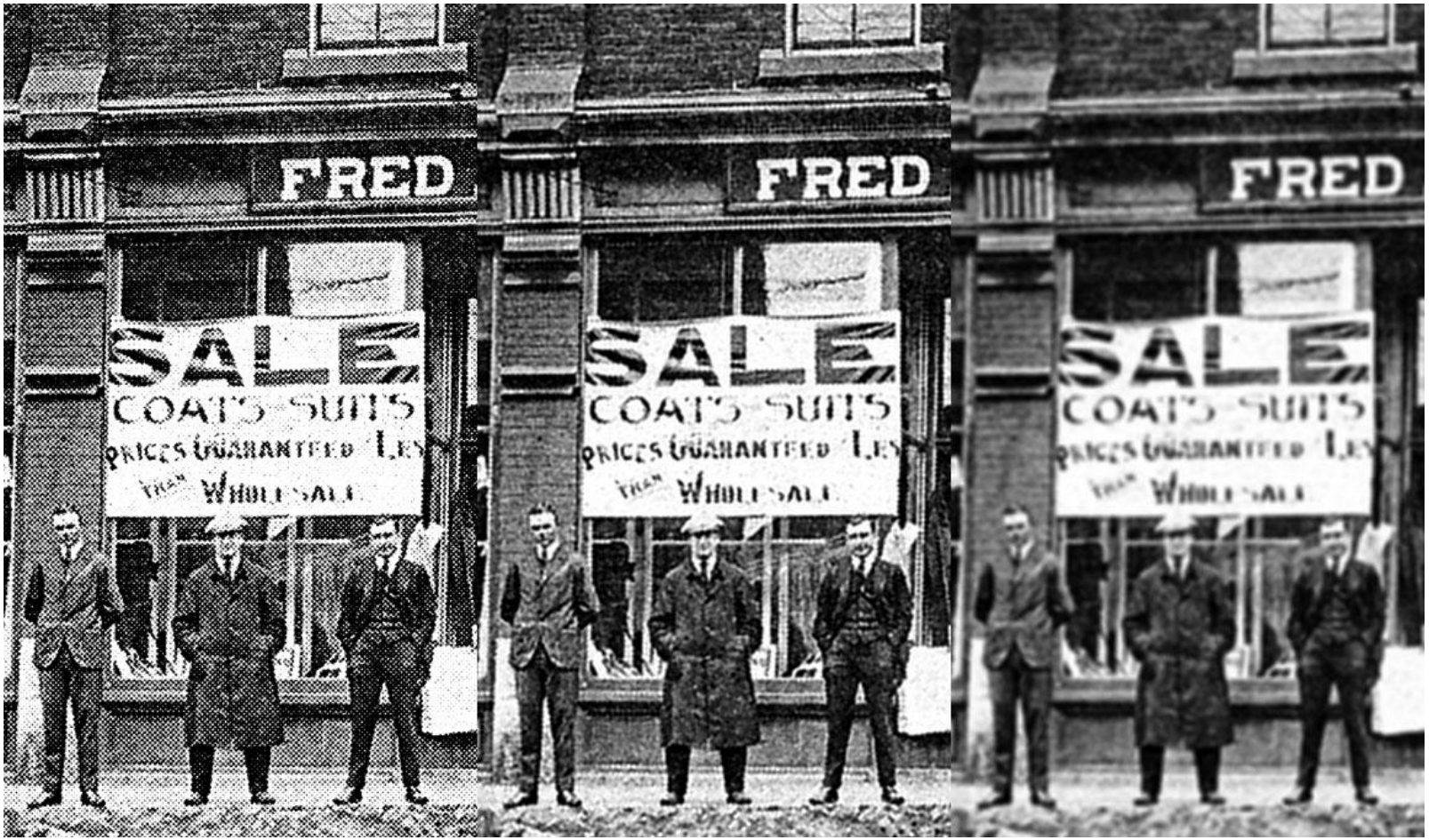
Newspaper image



Moiré-pattern

Common on laser printers

Descreening



No descreening | Sattva Descreen | Epson Scan Descreen

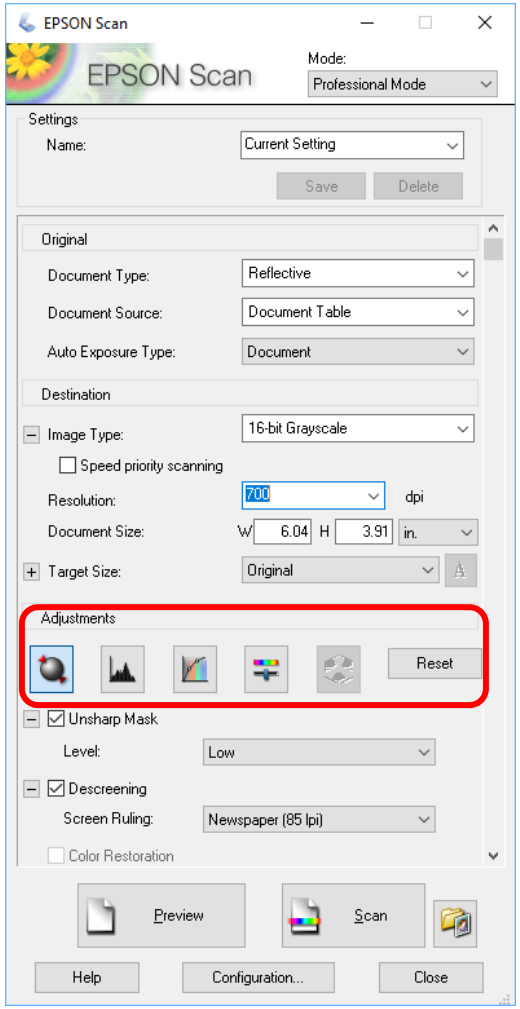
Newspaper 85 lpi setting

Image scanned at 300 ppi, 16-bit grayscale

Image adjustments

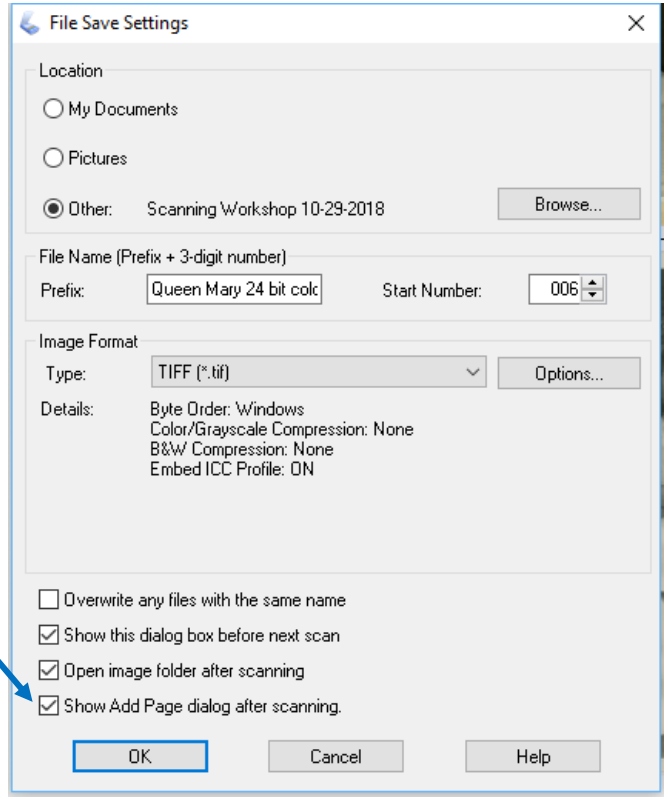
- Contrast
- Tone correction
- Color

Usually best done by editing the Production Master file so that these changes don't get "baked in" to the Archive Master



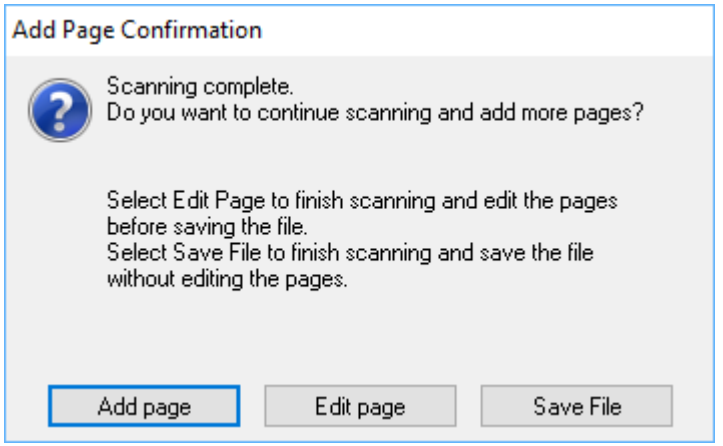
Output file formats

- TIFF (Tagged Image File Format)
- PDF (Portable Document Format)
- ~~JPEG~~ (Joint Photographic Experts Group)
  - *Never choose JPEG as output format*



PDF Output

- Can scan multiple pages into the same PDF file
- Do OCR after PDF is created
- Can use other scanner software for OCR e.g. Abbyy Finereader, Foxit
- 300 ppi recommended for OCR



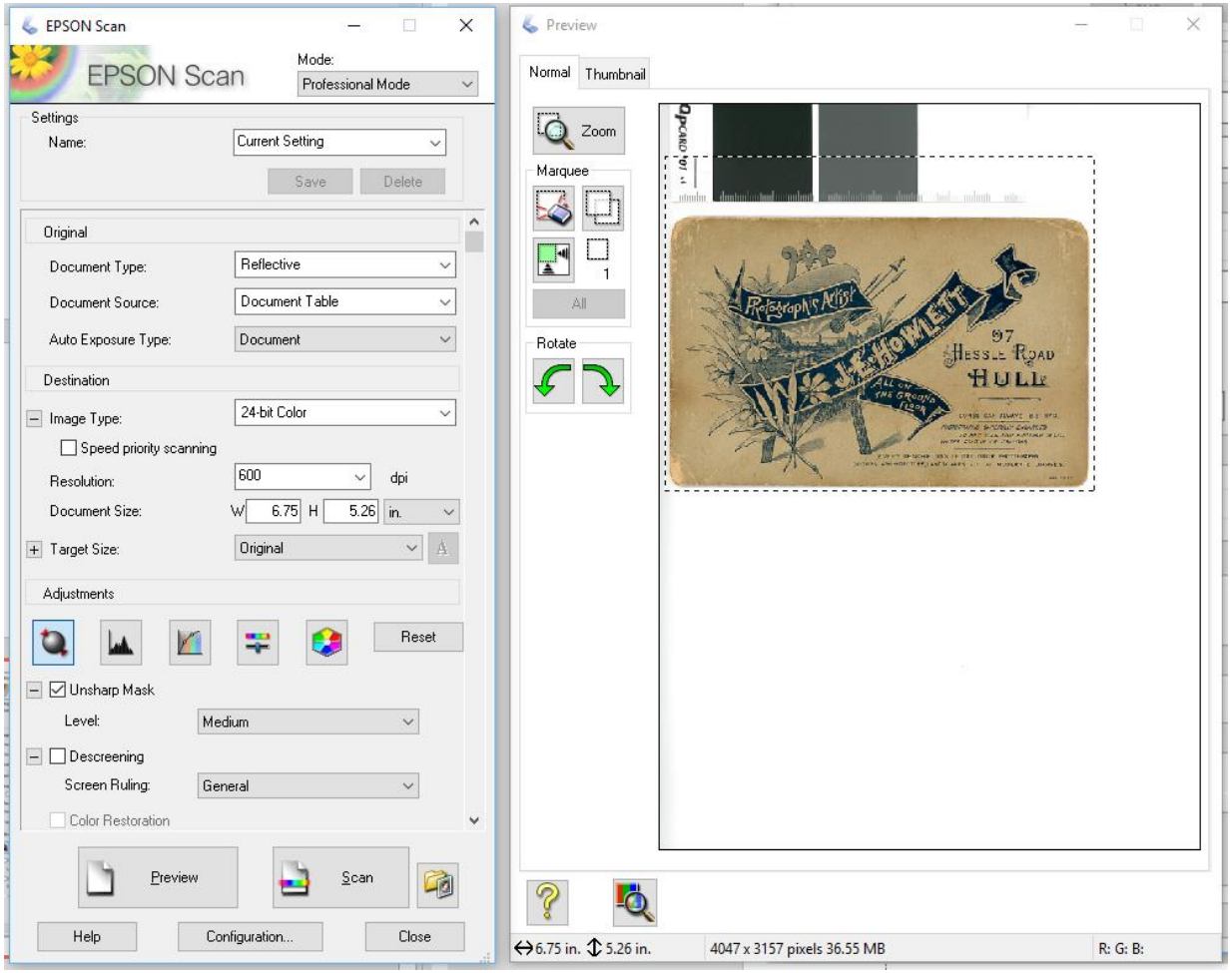
Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Scan the item

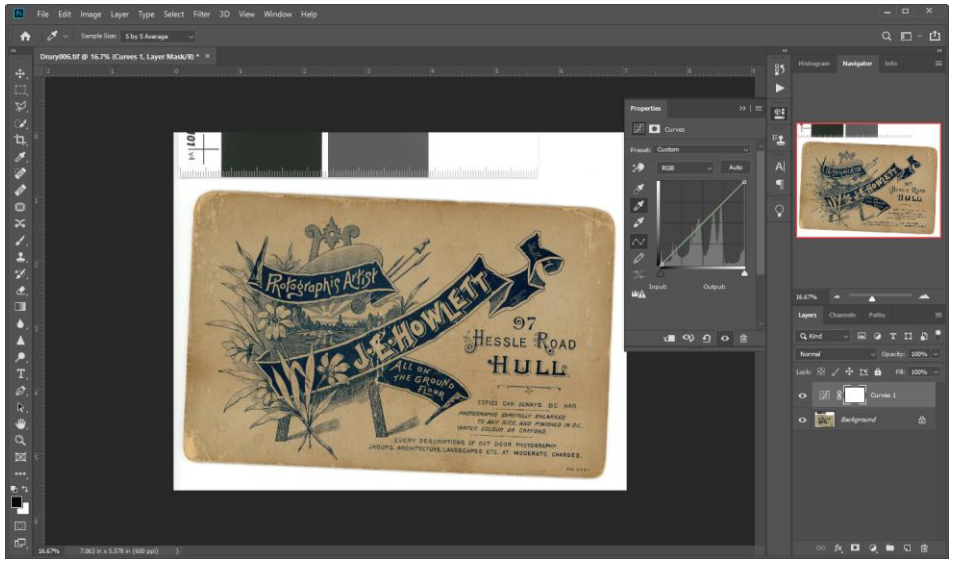
- Perform a preview scan
- Select the area to be scanned

Prescan



Scan the item

- Perform the actual scan
- Save the scan as the Archival Master (AM)
- Use curves layer to correct color



Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

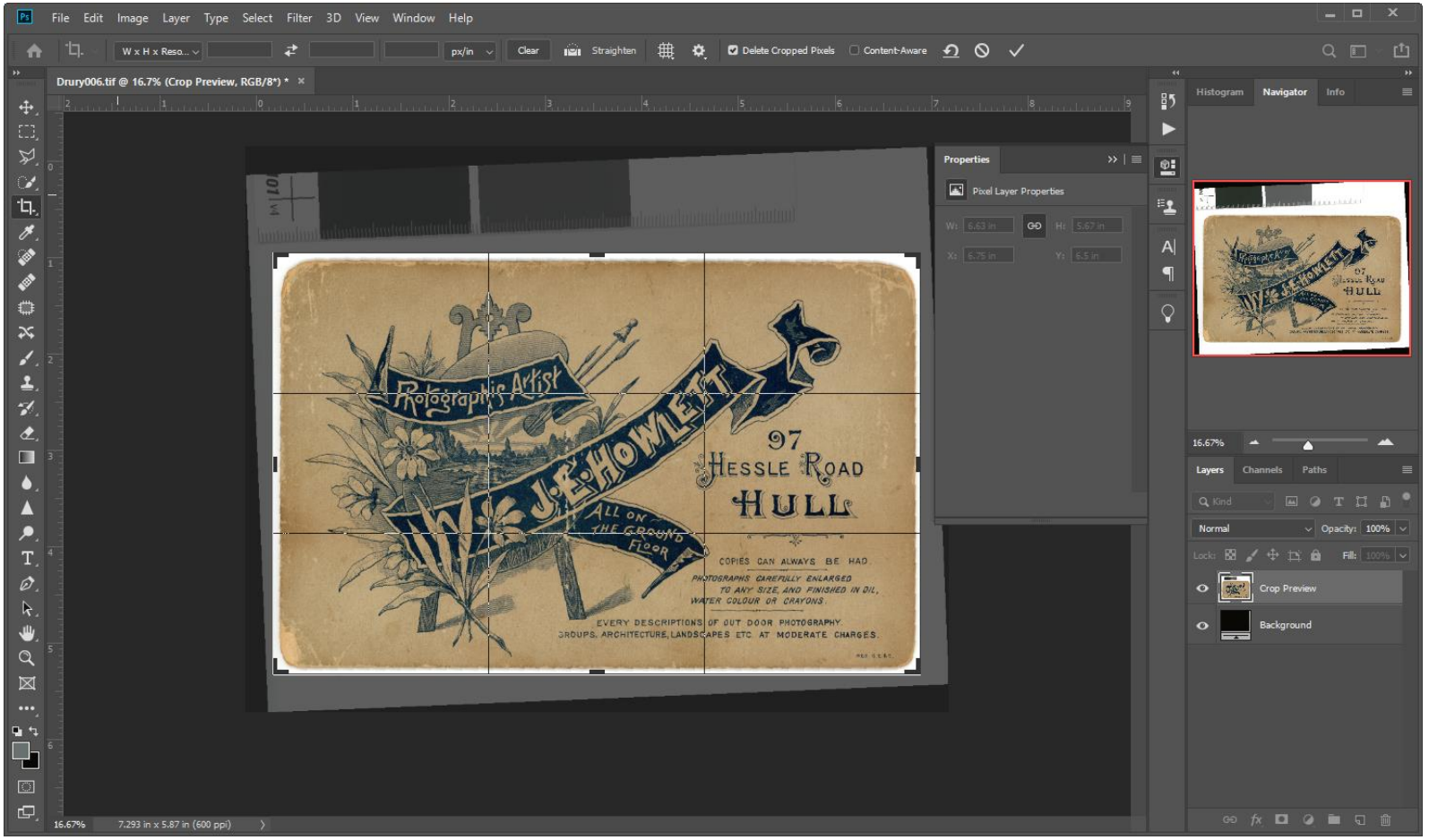
Create derivative images

- Production Master (PM)
  - Crop
  - Straighten
  - Contrast and tonal corrections

Image editing software

- Essential features
  - Crop and rotate by degrees
  - Resize
  - Save as JPEG
  - Sharpening
  - Tonal correction e.g. curves layer
- Programs
  - Adobe Photoshop and Lightroom $10/month
  - Corel Paintshop Pro 2019 $55
  - Affinity Photo $50
  - Adobe Photoshop Elements 2019 $100 (limited support for 16 bit color)

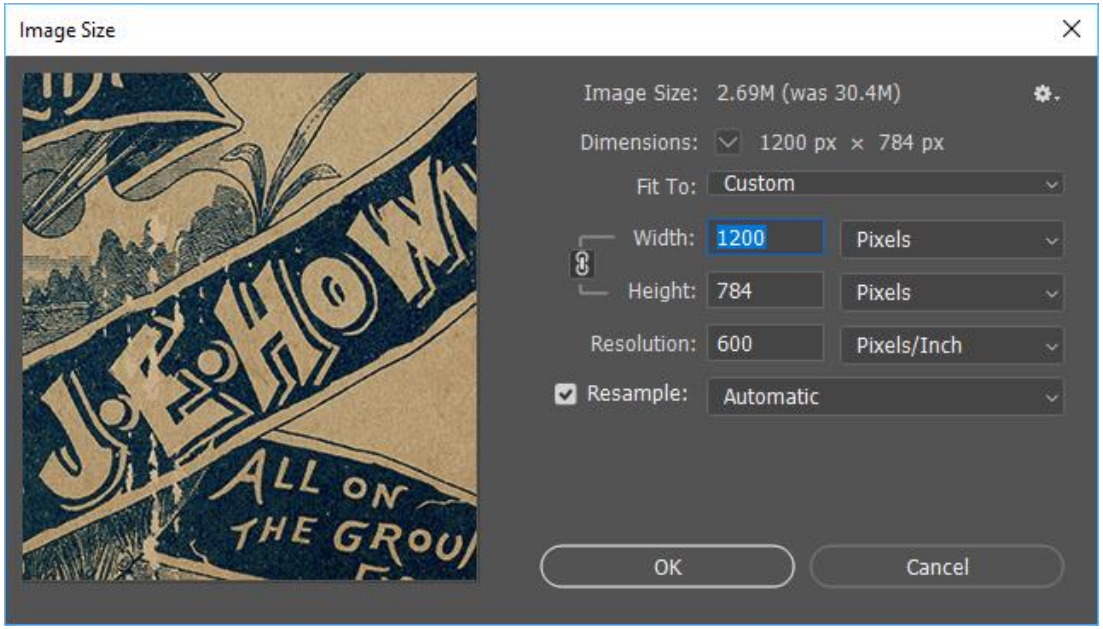
Crop and straighten

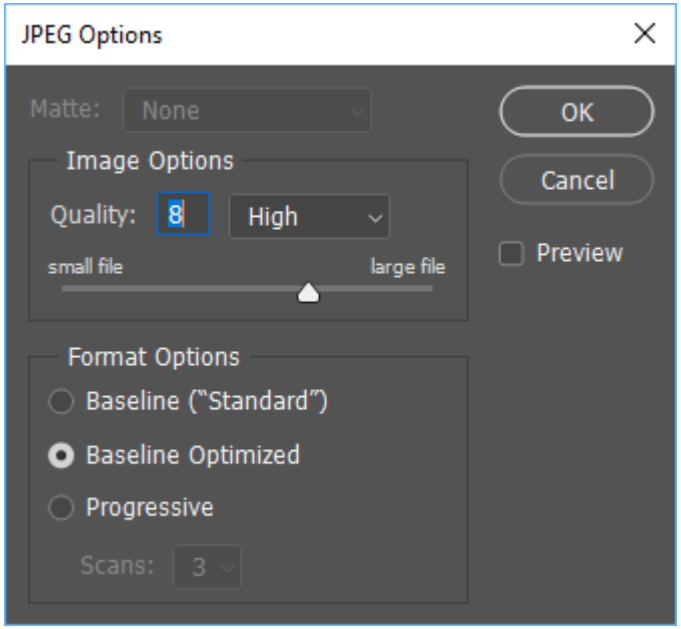


Create derivative images

- Low resolution (low res) for Digital Archive
  - 1200 px on the long edge
  - JPEG with high (not best) quality (8 in Photoshop, 75 in Lightroom)
- Other variations e.g. retouched or B&W
- Tiles for zoomable images

Save low res JPEG



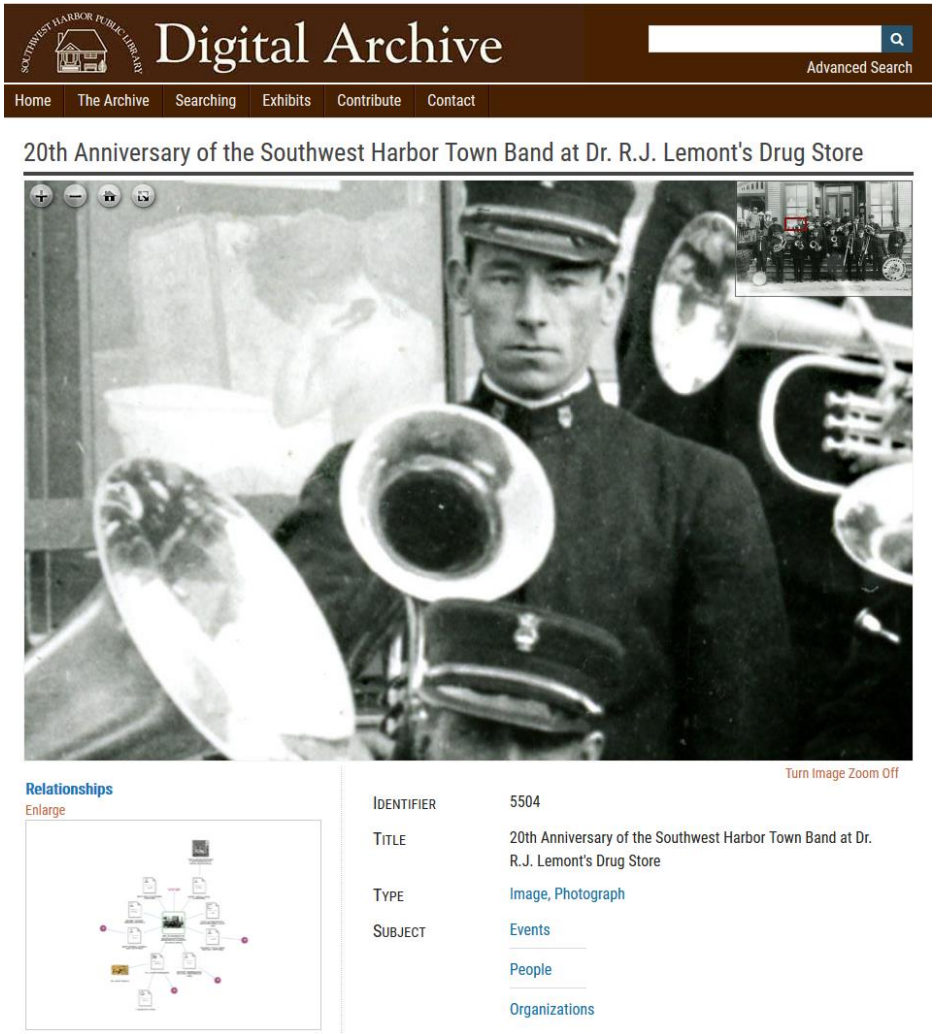


Low res JPEG

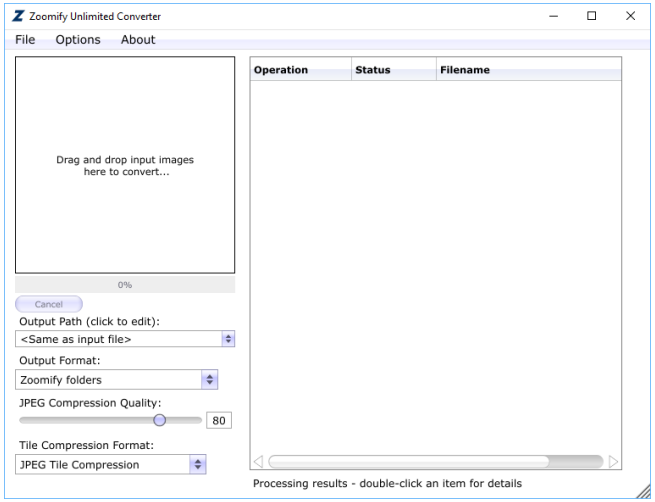
Looks good online

Looks bad printed at a large size

Zoomable images



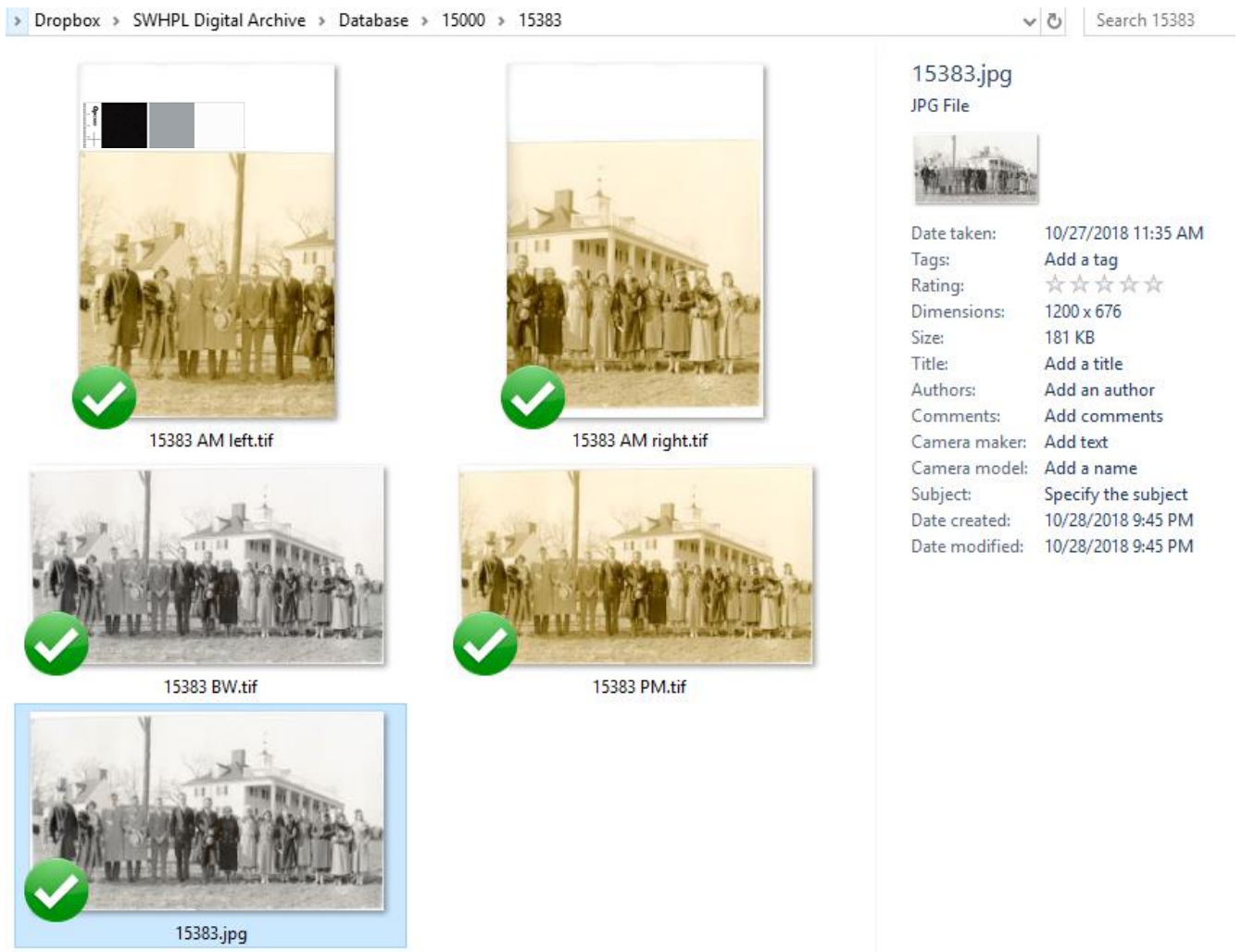
Zoomify tiles



File management

- Store all files for item in the same folder
  - Archival Master(s)
  - Production Master
  - Derivative images
- Make multiple backups
  - At least one hard drive kept off-site
  - Cloud storage e.g. Dropbox
- Digital Preservation
  - A formal endeavor to ensure that digital information of continuing value remains accessible and usable

Digital preservation



Scanning workflow

- Choose which items to scan
- Identify parts of the item to be scanned
- Prepare the item for scanning
- Choose scanner settings for the specific item
- Scan the item to create an Archival Master (AM)
- Create derivative images
- Add image and metadata to the Digital Archive

Add item to the Digital Archive

- Create or edit the item
- Record its metadata
  - Record dimensions and any other information about the item that you could not derive later without having the original item
- Upload low res image