Photo and Document Scanning at the Meridian East Idaho Family Search Center
Stewart Wyatt - March 2022

This Presentation was provided in an Inservice meeting and provided an overview of the existing scanning equipment.

Equipment mentioned: Lexmark MX622 Monochrome Laser Multi Function Printer
Kodak negative scanner
Wolverine 8mm and Super8 Film Scanner

Instructional presentations were offered on **photo scanning**:
Epson FastFoto FF-680W
HP Scanjet G3110 Flatbed Scanner

HP Scanjet G3110 Flatbed Scanner instruction on **document scanning** to create PDF documents and to use OCR (Optical Character Recognition) software to recover and editable text file from a document scan.

Goal of this presentation:

Introduce the scanning equipment in the Meridian Idaho East Family History Center

Focus on scanners that support FamilySearch Memories Uploads
(Pictures and PDF documents)

Describe the scanner operation to illustrate their capabilities.

After this presentation, you should have a good idea of which scanner to choose for a particular task and what the it can do .

We have Users' Guides printed in the library and available at
https://meridianidahoeastfhc.com/ Center Resources -> Photo and Film Digitizing

FamilySearch Document Requirements:

For documents: .pdf file formats up to 15MB. Other document formats such as .doc and .docx files will need to be converted to .pdf before they can be uploaded.

For Photos: .jpg, .png, .tif, and .bmp file formats up to 15MB.

Lexmark MX622 Monochrome Laser Multi Function Printer.

It has a scanner with an automatic document feeder (ADF).

Scan to Family Search also USB Drive option

Kodak negative scanner.
Can also scan 35 mm slides

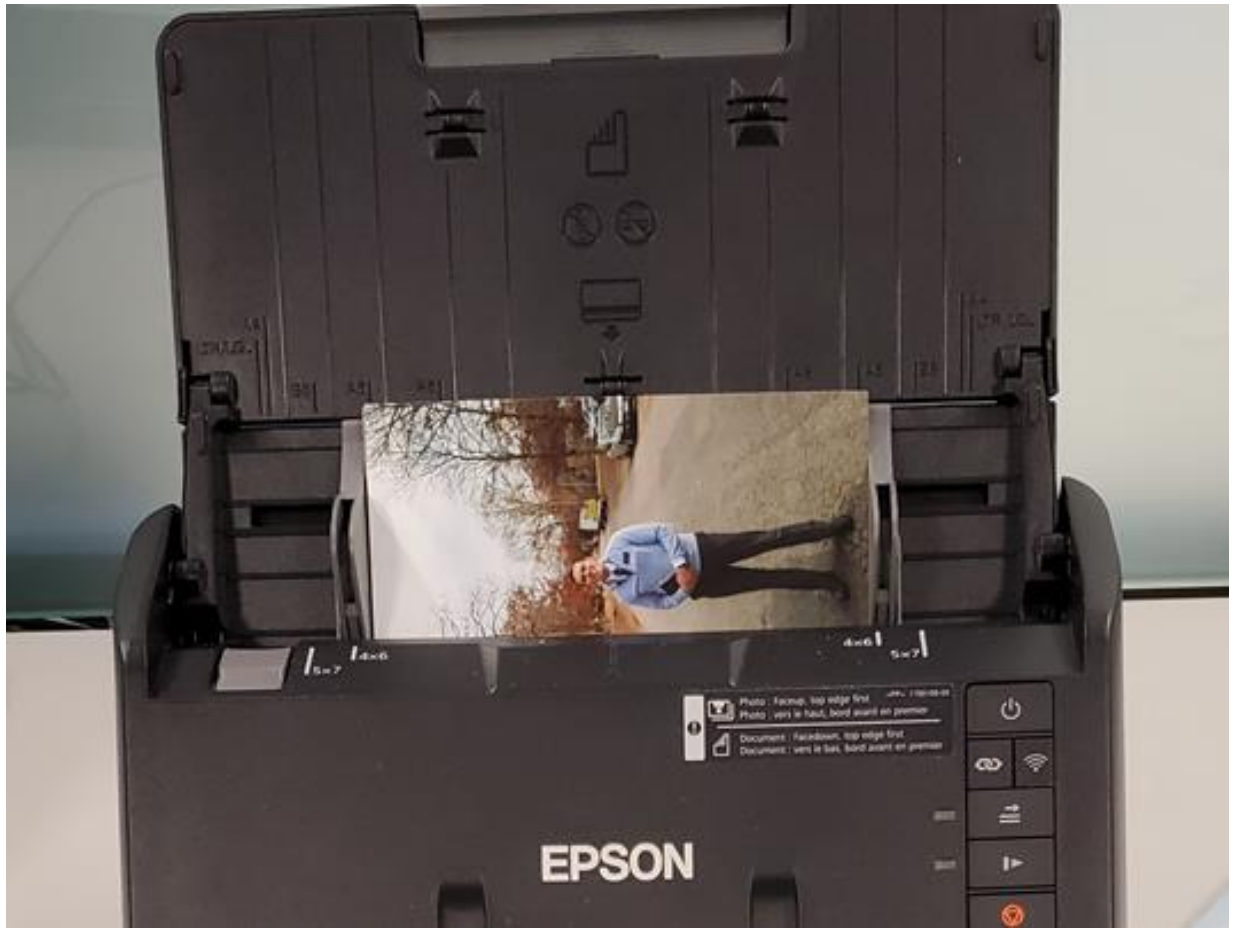
Wolverine 8mm and Super8 Film Scanner

Scan home movies to a digital mp4 format
We have a user's guide for this device.

Photo Scanners: Two Options

Epson FastFoto FF-680W
ADF – Automatic Document Feeder
Photos pass through the scan path while the scan bar remains fixed

HP Scanjet G3110 Flatbed Scanner.
Photos are scanned on the flatbed glass.
The photos are stationary while the scan bar moves past the photos


EPSON


EPSON
FF-680W

Epson FastFoto FF-680W Photo Scanner

ADF – Automatic Document Feeder. Can scan up to 30 photographs of the same width at a time. Alternately you can manually feed photos of varying widths.

The scan path includes a photo scan bar on the top and a document scan bar underneath. The photographs will be bent during the scan process. Consequently this scanner should only be used to scan standard flexible photos. This scanner is very fast and effective at scanning these types of photos.

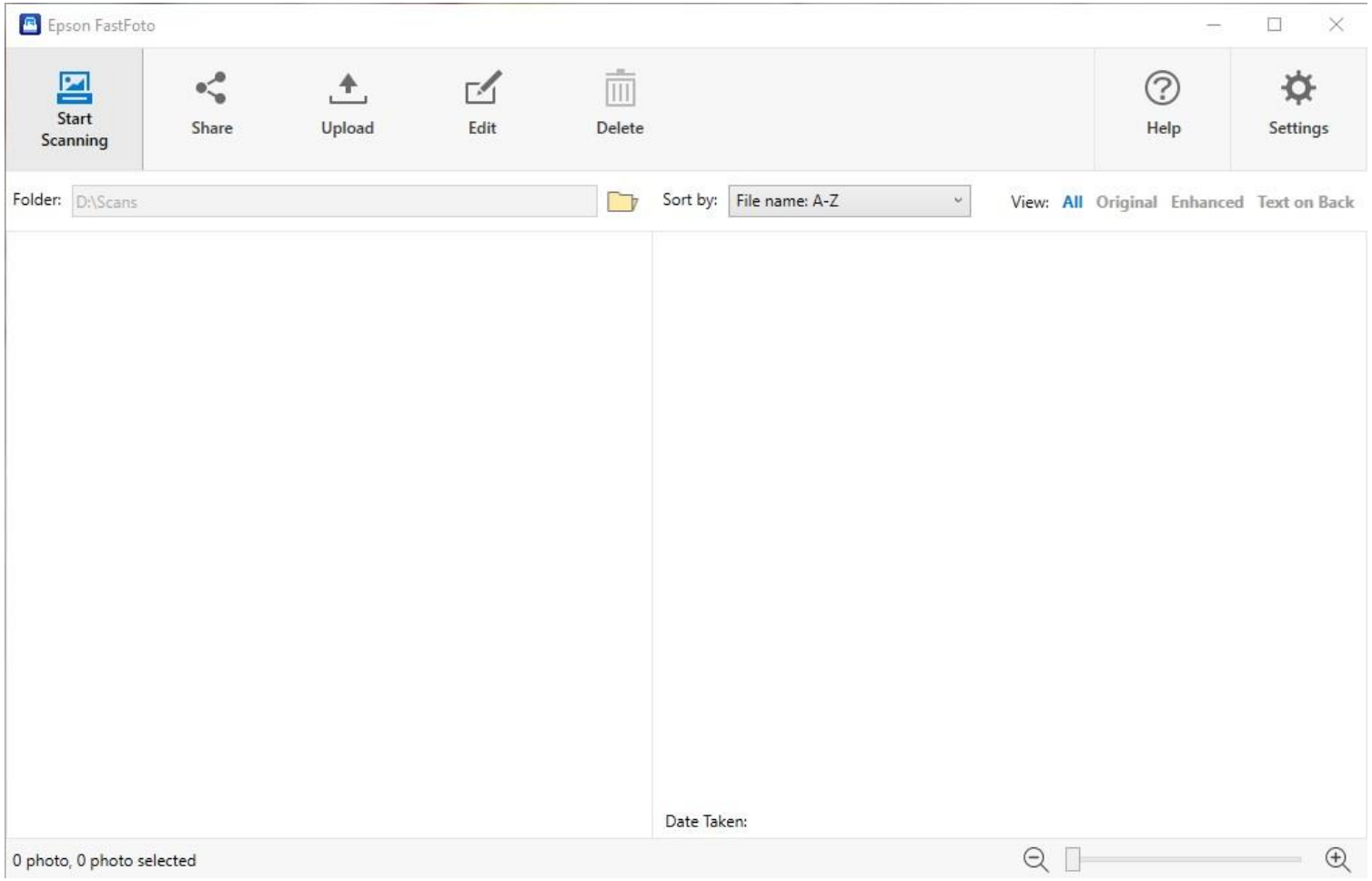
Delicate documents such as most programs, historic photos will be damaged or destroyed by the ADF. Photos mounted on cardboard, that have adhesive such as photos that were previously glued in an album page could damage the scanner and be damaged themselves. Irregularly shaped documents will jam the ADF. Documents like books, magazines and newspapers cannot be moved through the ADF. These photos should be scanned on the HP Flatbed G3110 scanner instead.

Epson FastFoto Scanner

Start the FastFoto Software by clicking the icon on the desktop. The FastFoto window opens.

- Use the software rather than the controls on the scanner
- The software includes photo editing – Use Microsoft Photos instead.

Click the settings (cog icon) top right of the window

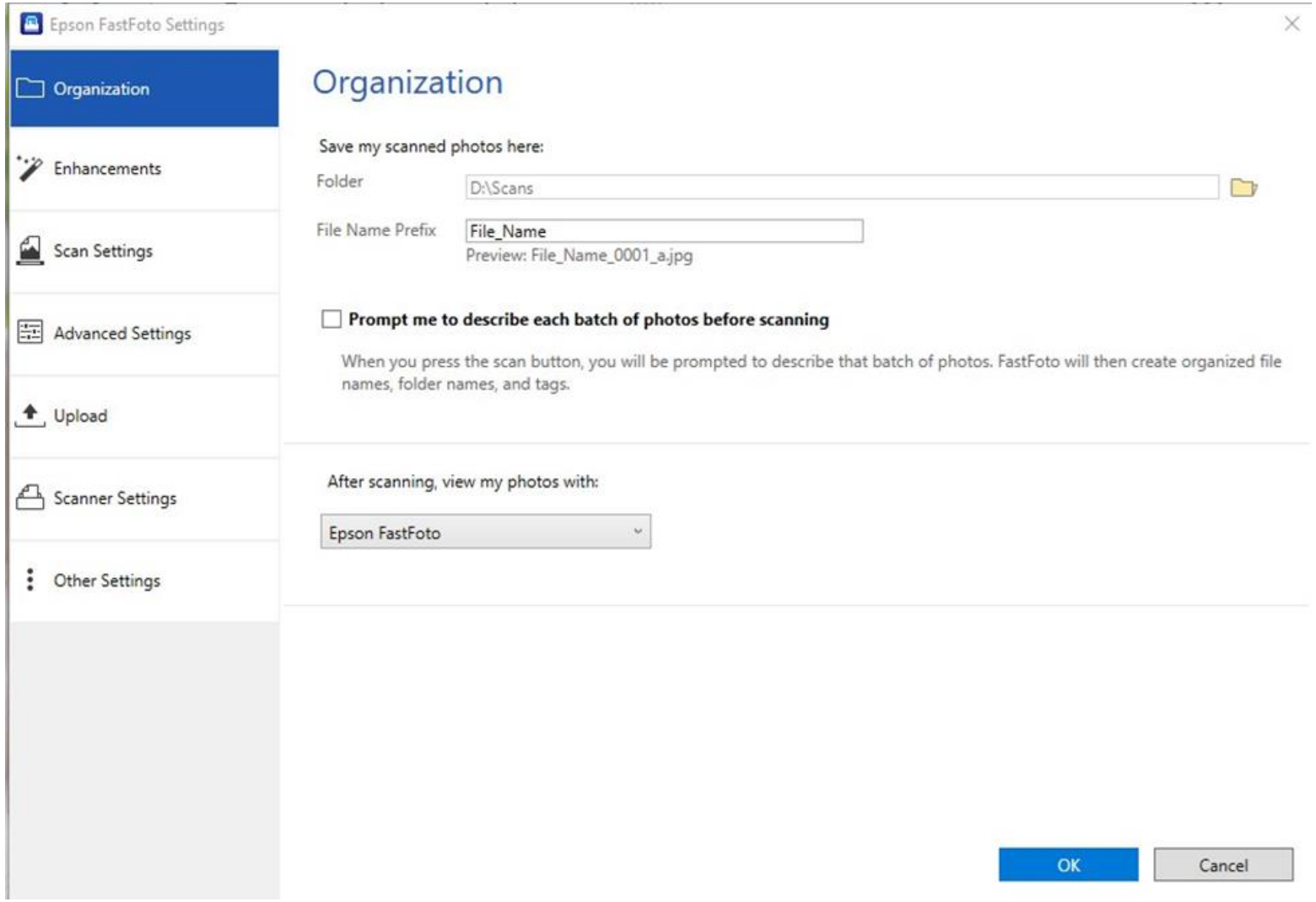


The FastFoto Settings Window opens to the “Organization “ tab:

- It is critical to set the scan location before staring a scan operation. If the path is nonexistent, the software will crash without explanation.

- Enter an appropriate file name prefix and note the naming conventions

Then select the “Enhancements” Tab

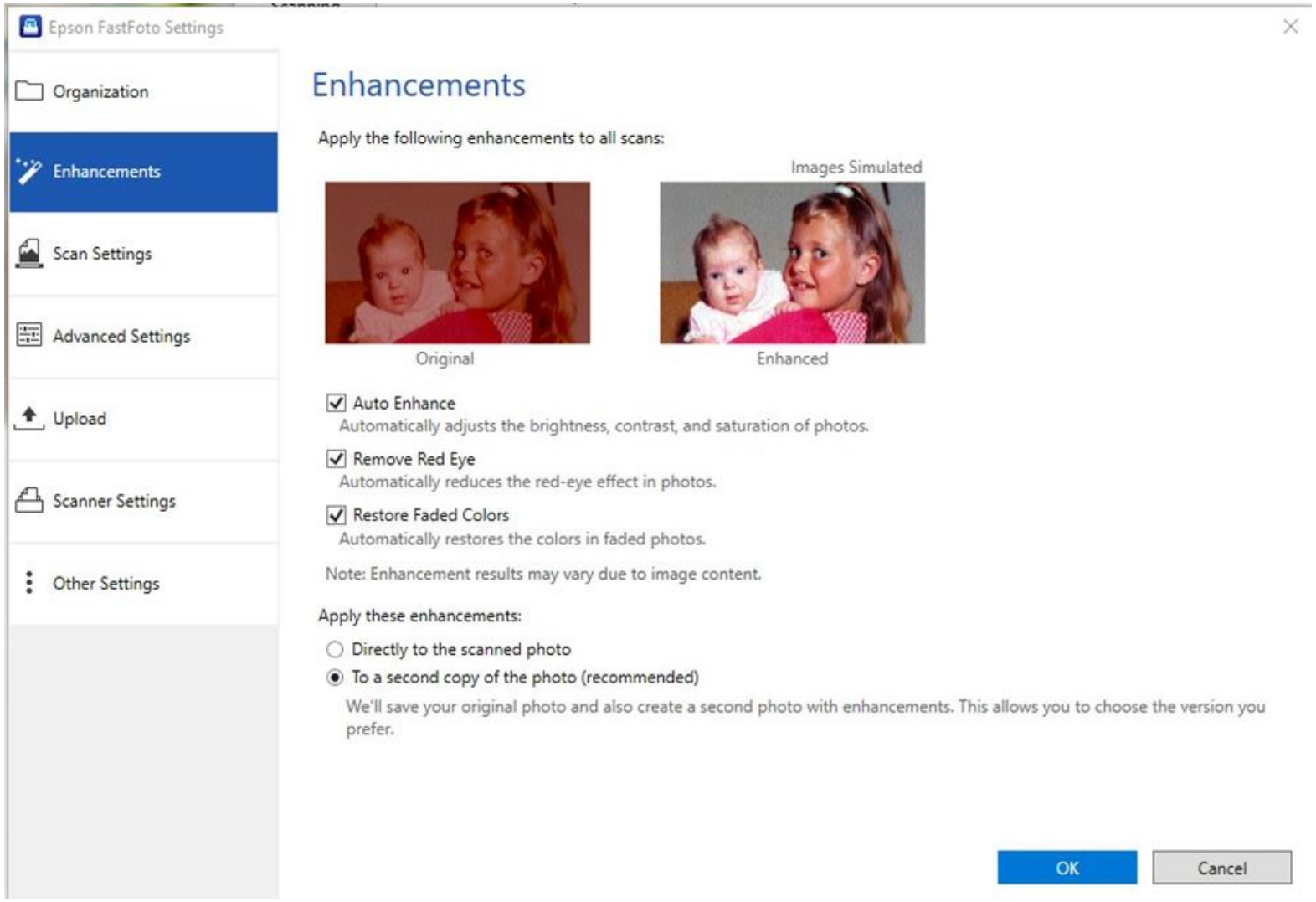


The FastFoto “Enhancements” Window Opens with three options

- Auto Enhance
- Remove Red Eye
- Restore Faded Colors

Select all three

- Apply these enhancements to a second copy.



This picture was taken in July of 1957. The original is on the left and has yellowed and the colors have faded. The version on the right was enhanced by FastFoto software.

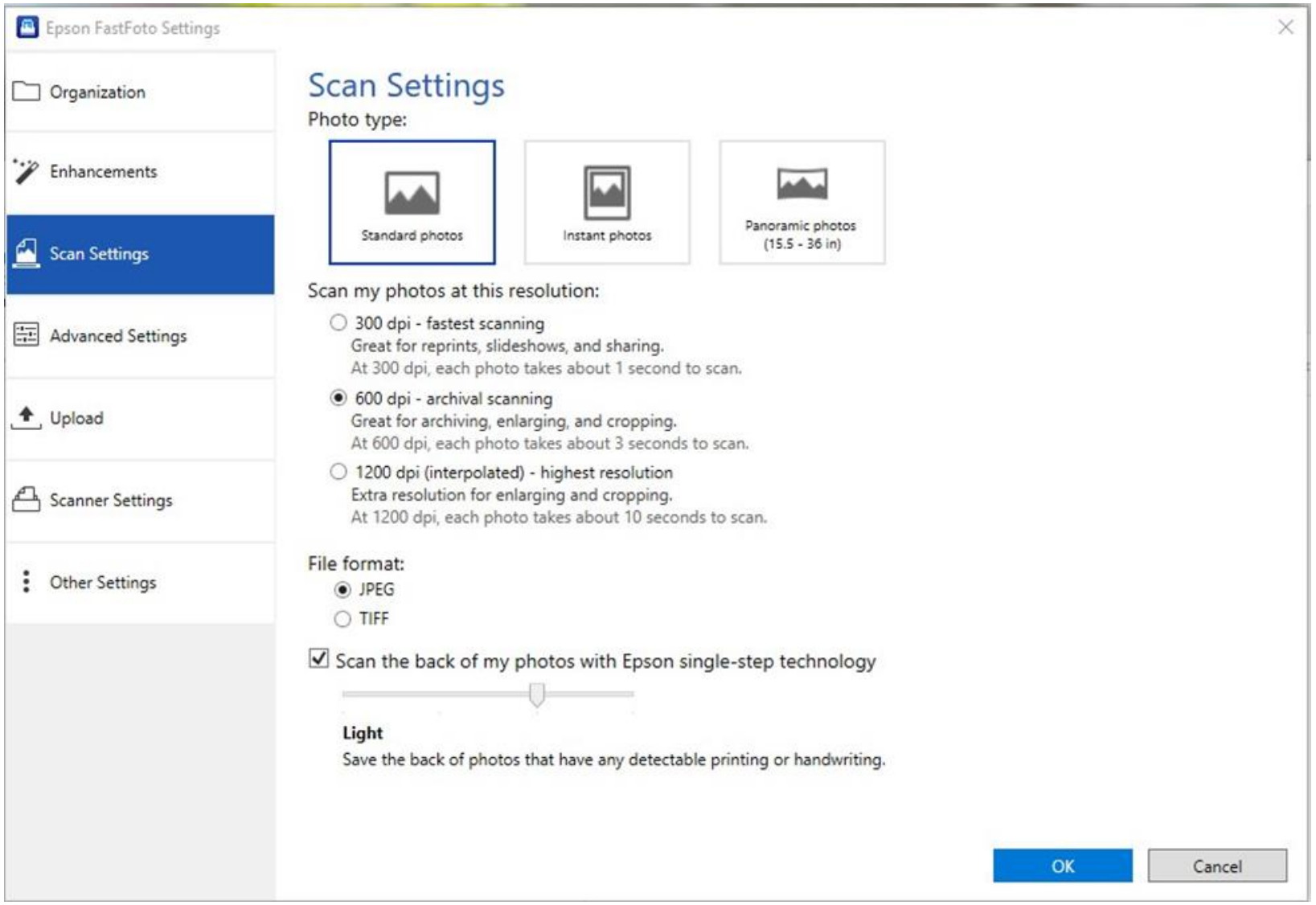
This photo is dated November 1969, original on the left has yellowed, colors have faded, and the subject has red-eye.

Select the scan settings tab and the Scan Settings Window opens.

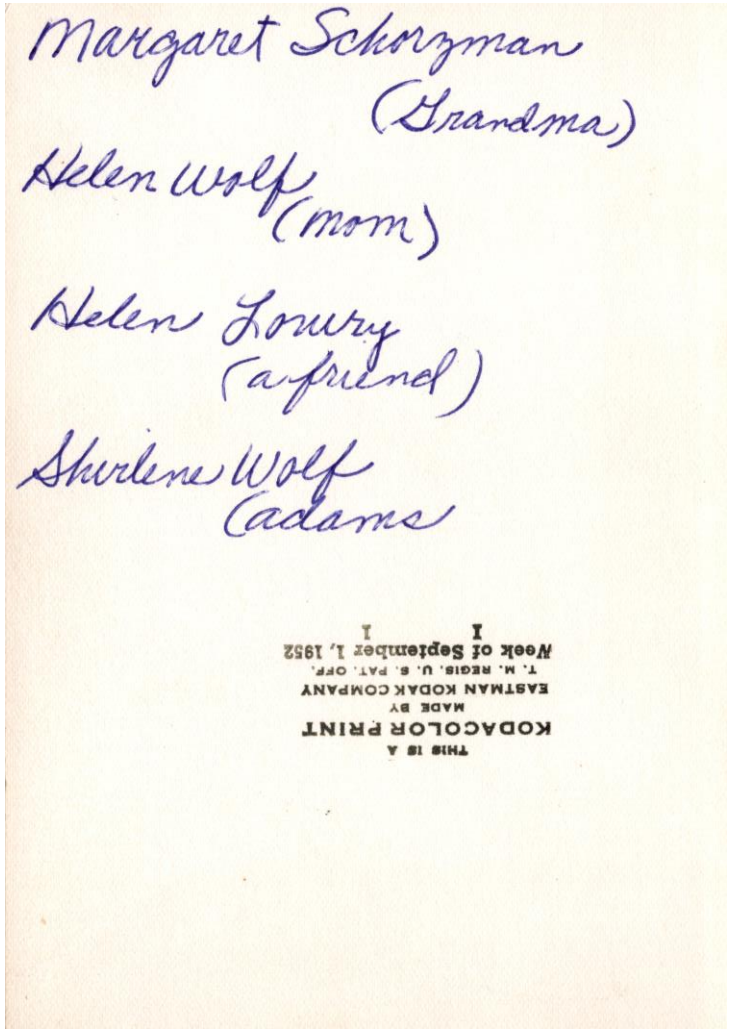
Note that there is a separate setting for “Instant photos” for Polaroid type photos.

600 dpi is recommended for archival scanning and is a good choice.

Select scan the back option



The File Name Prefix was Schorzman and this was the 33rd photo in the set

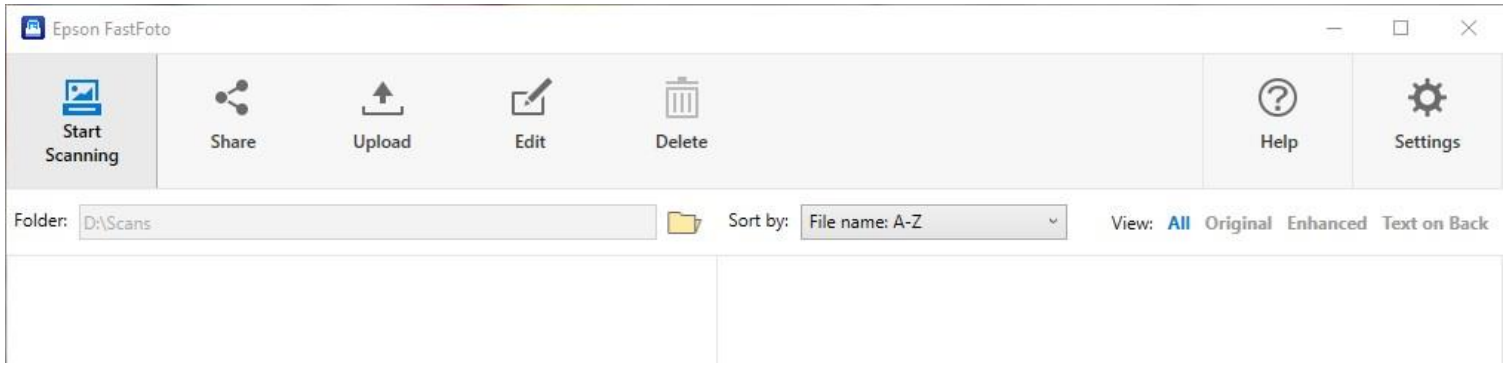


The first photo files are named Schorzman _0033.jpg, Schorzman_0033_a.jpg and Schorzman_0033b.jpg

The remaining tabs on the settings windows

- Advanced Settings
- Upload
- Scanner Settings
- Other Settings

Are either irrelevant or the defaults are acceptable.



Select OK to exit the settings window and return to the Epson FastFoto Window.

Select the “Start Scanning” button at the top right of the page and scanning commences.

Once the scanner has completed the batch of photos you loaded, you can load subsequent batches, press “Start Scanning” again and the process the next batch. The numbering will advance between batches.

HP Scanjet G3110 Flat Bed Scanner for Photos

The function of the Epson FastFoto Scanner and the HP Flat Bed scanner are very similar but the architecture is different.

The Epson FastFoto scanner has an Automatic Document Feeder. The scan bar is fixed in the scan path. The photos move past the scan bar to be scanned

The HP Scanjet scanner has a scanning surface, called the glass, on which objects to scan are laid. The scan bar moves under the glass to scanned the photos.

Consequently The HP Scanjet flatbed scanner is the preferred solutions for small, irregularly shaped, delicate photos. It can also scan books, magazines and newspapers. It is by far the most flexible scanner in the library.

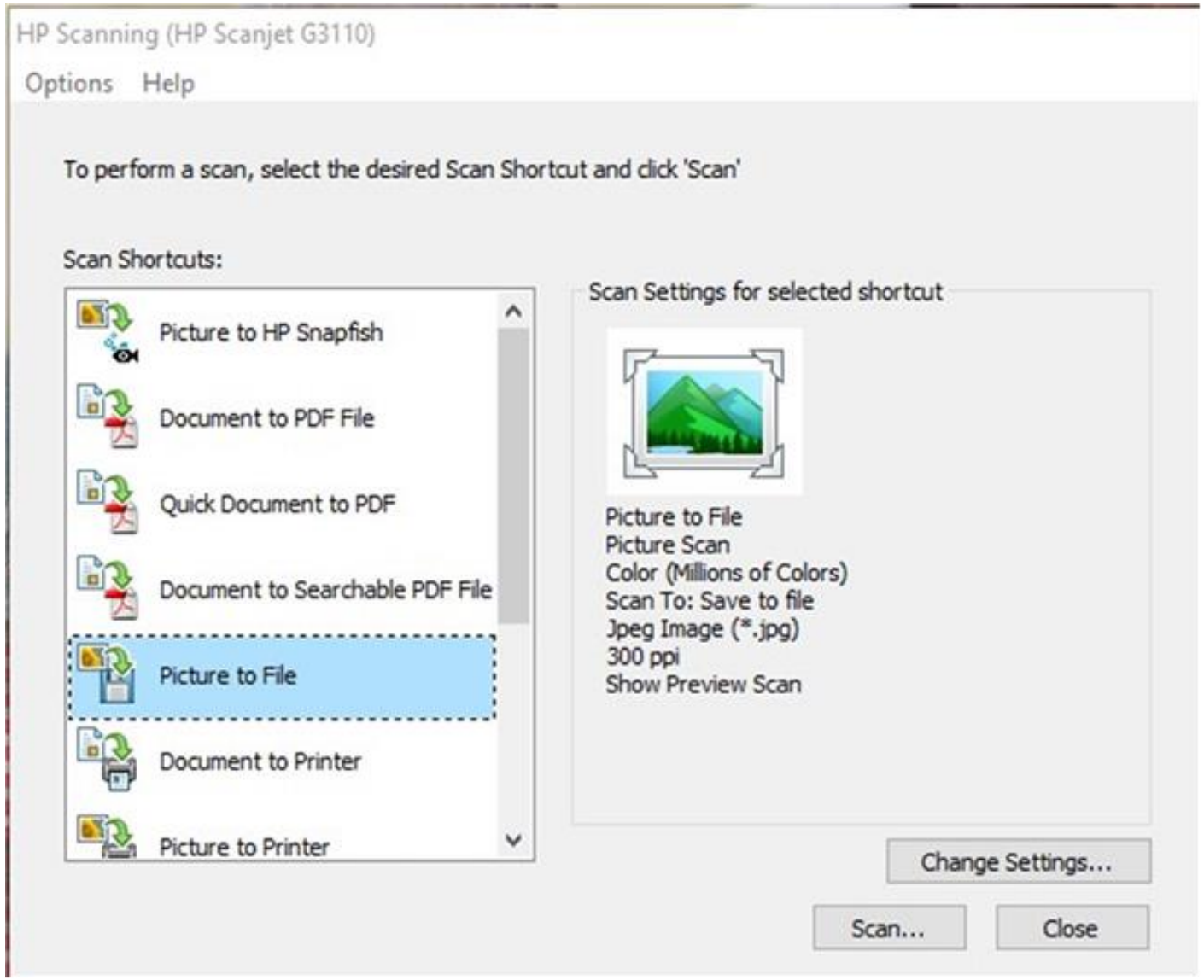
The HP Scanjet has multiple photo scan options as well.

Select the “HPScanning (G3110)” icon on the desktop.

The HP Scanning (HP Scanjet G3110) Window opens.

Select the “Picture to File” option from the “Scan Shortcuts:” window (note we will review the other Scan Shortcuts later in this presentation)

Then select “Scan…”



The window updates to display the Scan Shortcut Settings.

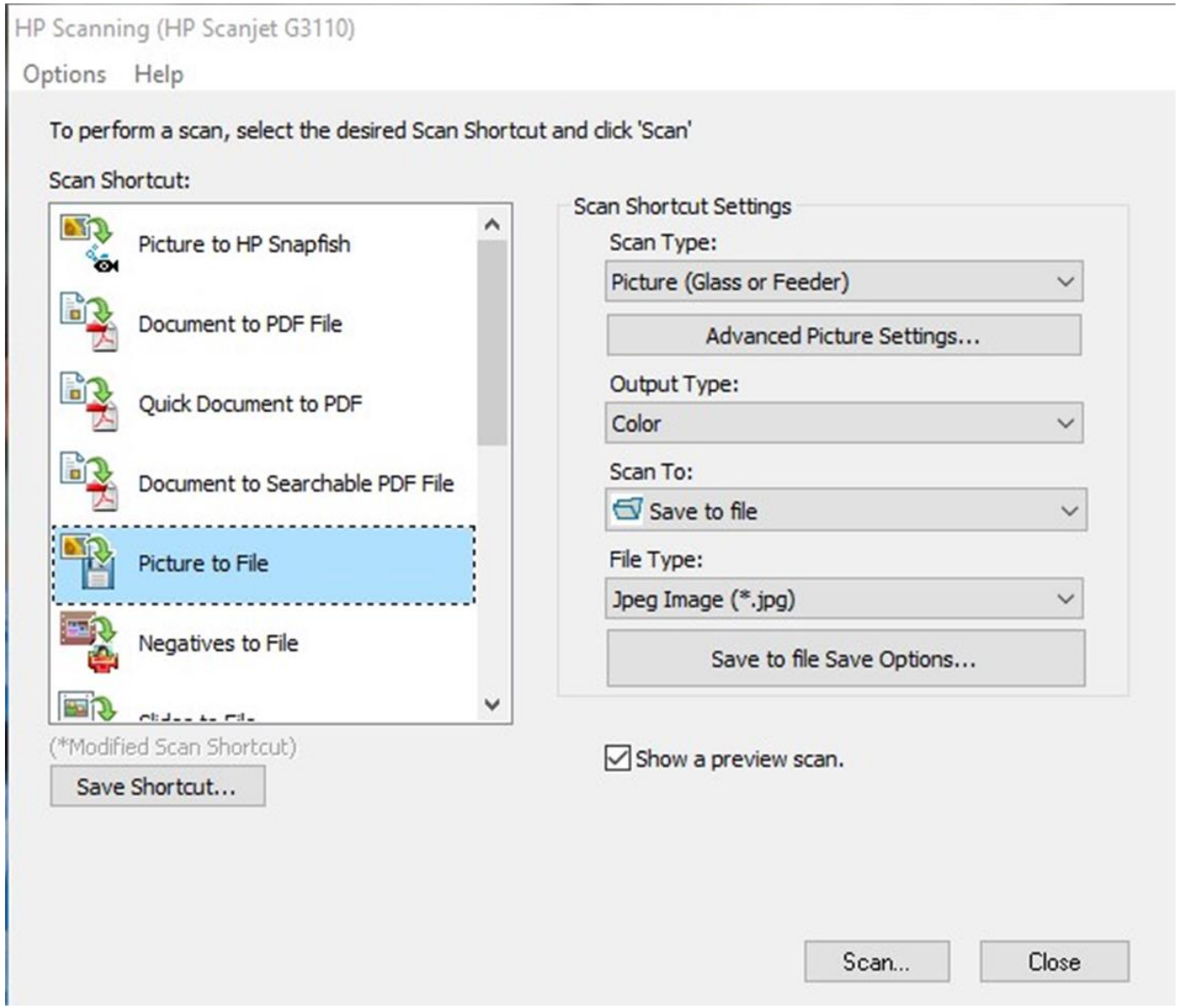
Output Types are Color, Grayscale or Black and White. Often Color is the best option even for a non-color picture.

File Type: There are a number of options. For most cases Jpeg Image is the best choice

Select “Advanced Picture Settings”

Next we will look at “Advanced Picture Settings…”

After that we will review “Save to file Save Options…”



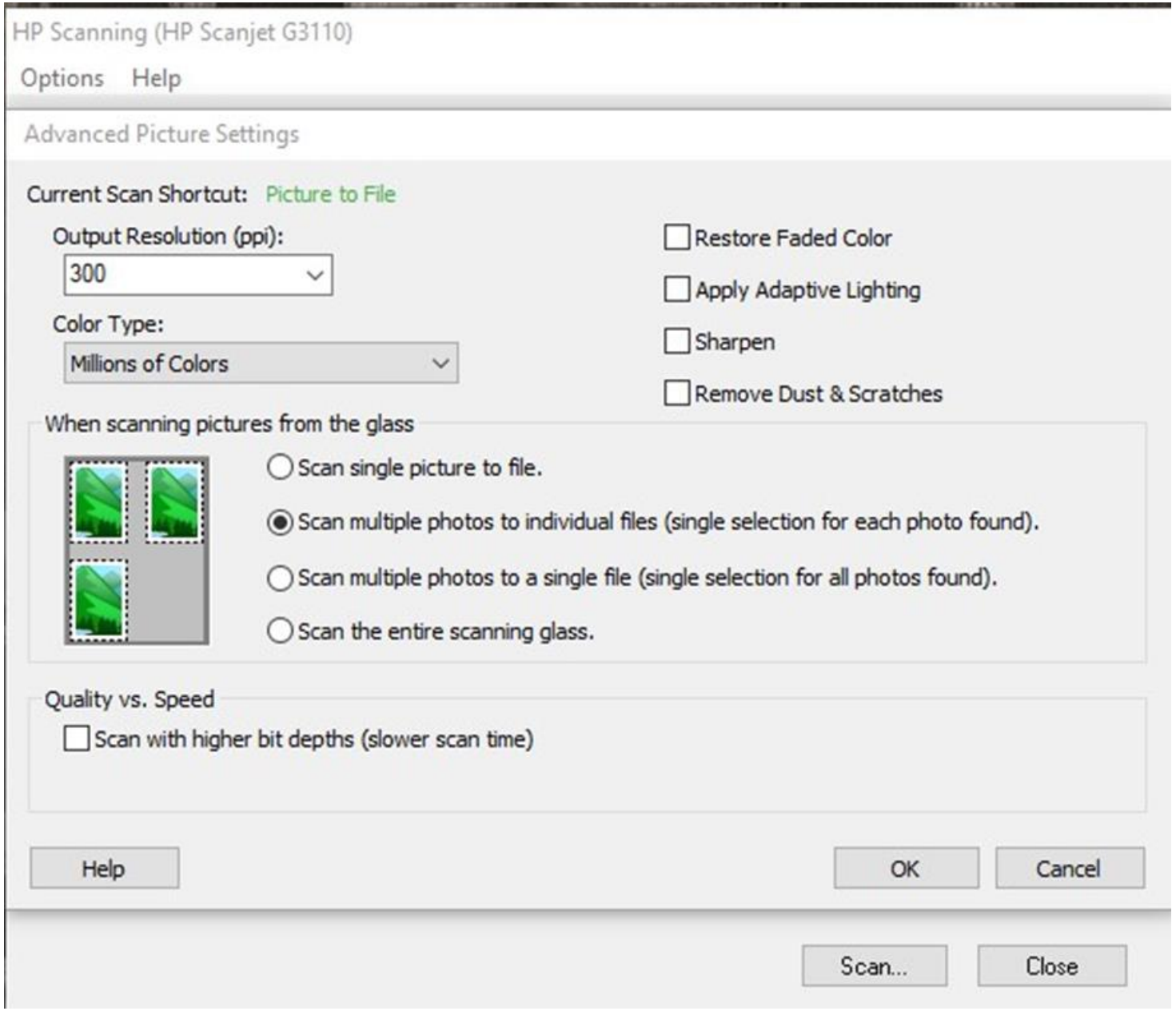
The Advanced Picture Settings Window opens.

Output Resolution (ppi)
PPI is Pixels Per Inch and is equivalent to Epson's DPI (Dots Per Inch) 600 DPI is considered optimum for archival scanning

Select all four photo enhancements
- Restore Faded Color
- Apply Adaptive Lighting
- Sharpen
- Remove Dust & Scratches

(These are equivalent to the Epson enhancements.) They can be selected later.

There are four scanning options: We will demonstrate "Scan multiple photos to individual files. Pass on Quality vs. Speed. Press "OK" to return to the scan shortcut settings window unless the quality needs help.



This picture was taken in July of 1957. The original is on the left and has yellowed and the colors have faded. The version on the right was enhanced by HPScanning software.

The picture on the left was enhanced by FastFoto while the one on the right by HPScanning

From the home page select “Save to the file Save Options…” The Save to file Save Options window opens.

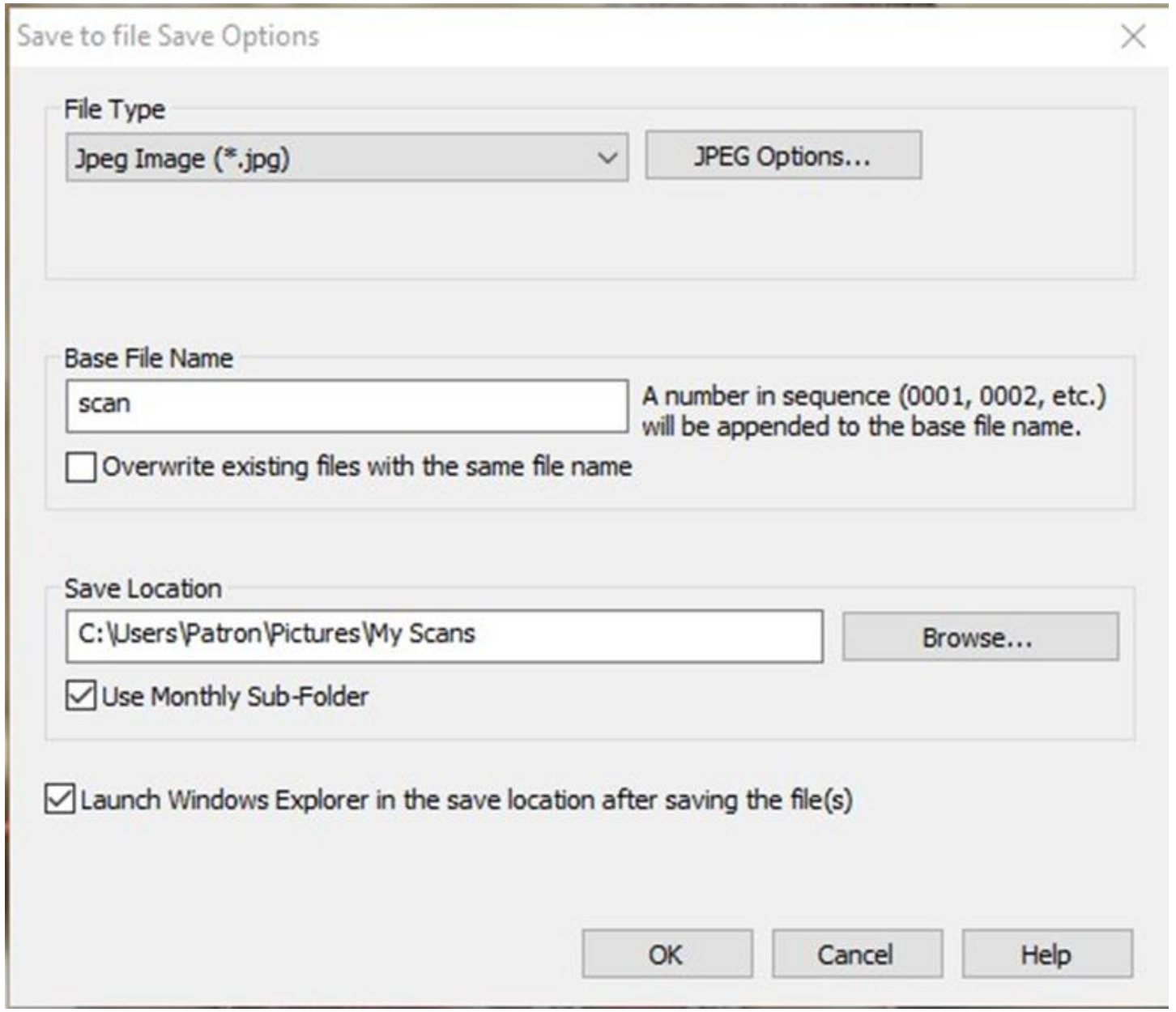
The File Type is repeated from the home page. The JPEG Options defaults are fine.

Enter a meaningful Base File Name.

If you select “Launch Windows Explorer …” the default “Save Location” is fine alternately you can browse to your USB memory device. But it resets after every scan operation.

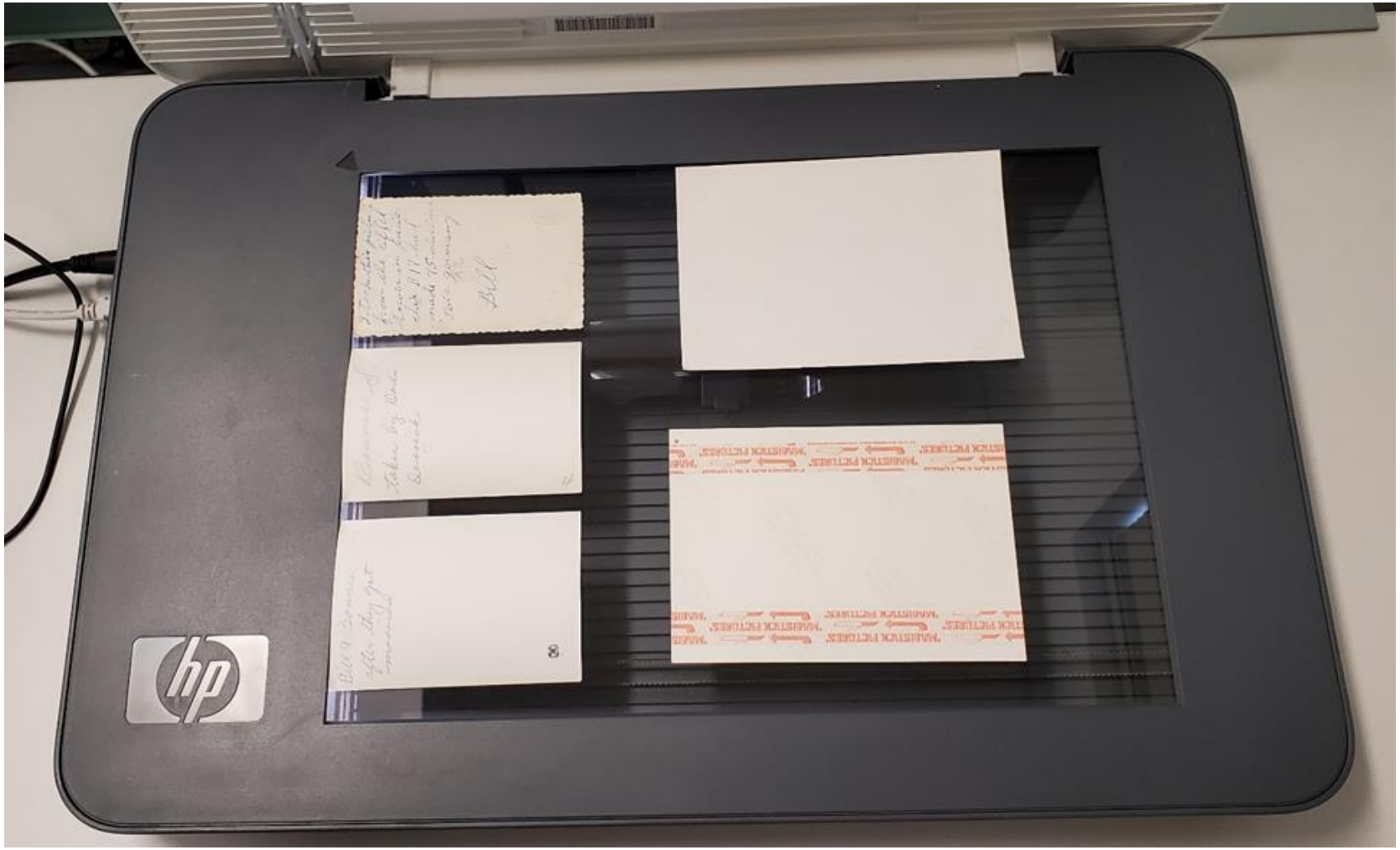
I prefer to not use the “Monthly Sub-Folder”

Press “OK” to return to the scan shortcuts setting window.



Now that the software is configured place your photographs on the glass and close the lid.

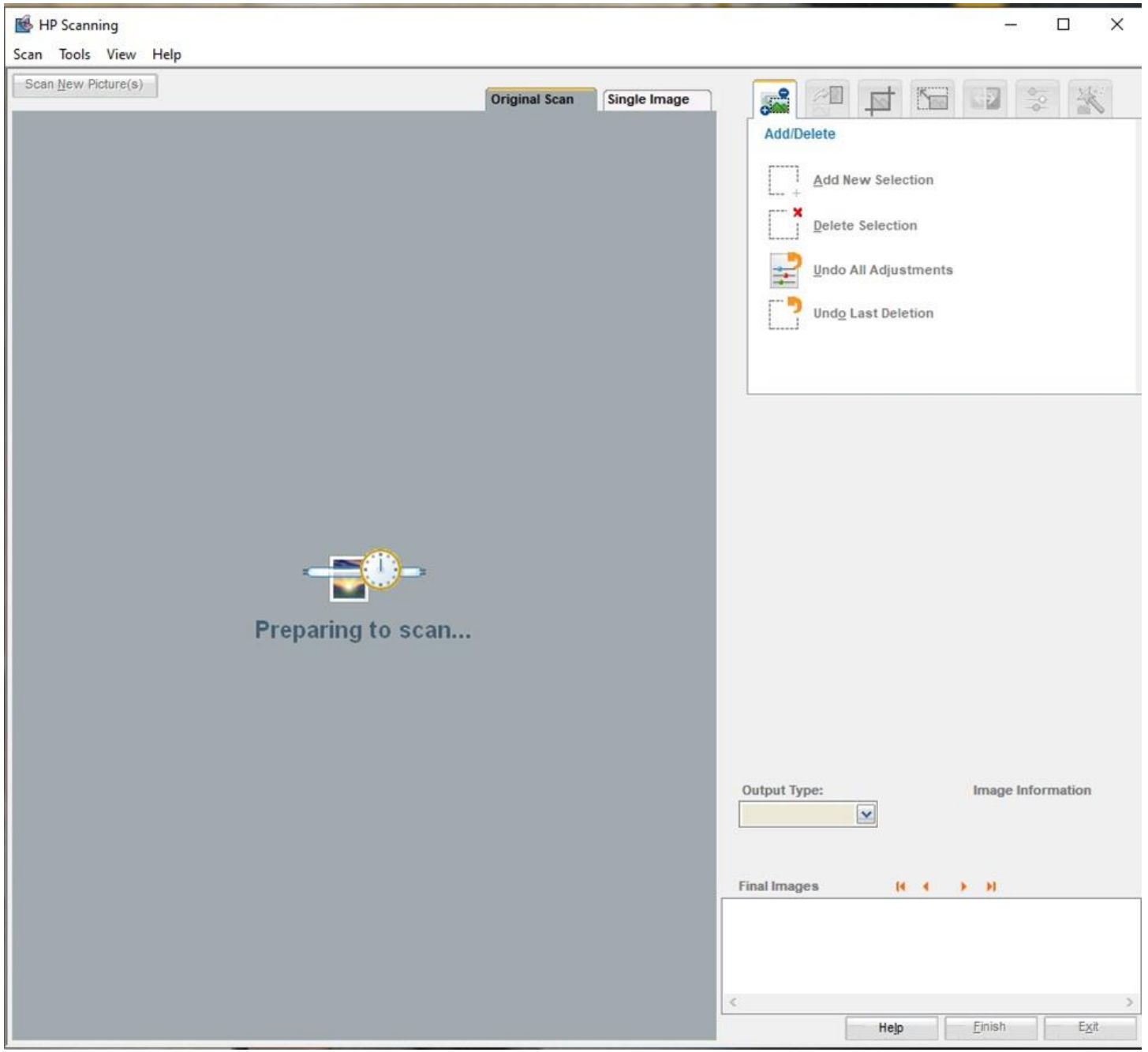
From the “Scan Shortcut Settings” window select “Scan…” to start the scanning process.



The Scan Shortcut Settings window closes and after a pause, the HP Scanning window opens with a banner “Preparing to scan…”

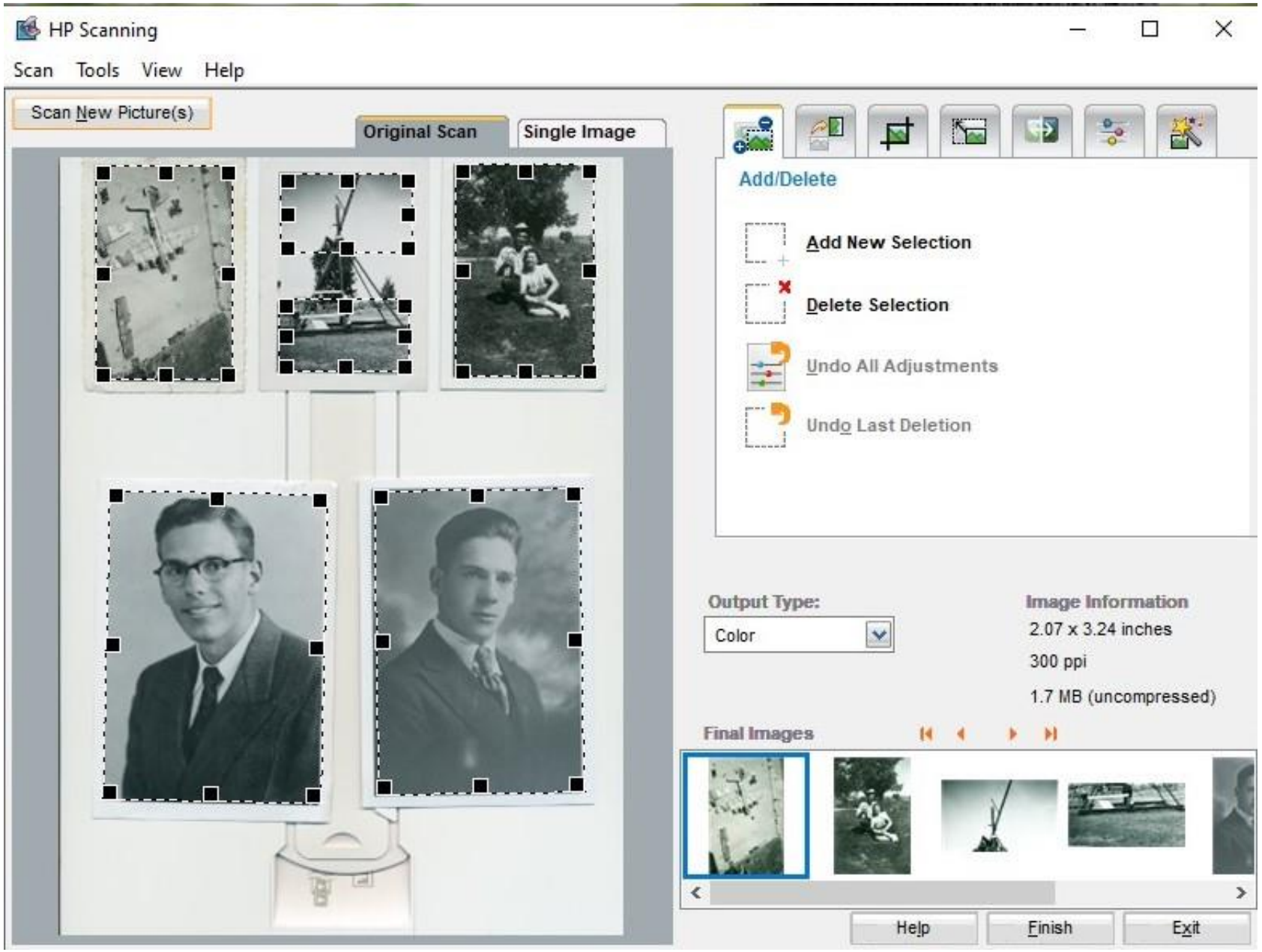
The scan lamp turns on, the motor engages and moves the length of the scan bed. It’s noisy – you can hear the motors.

The window updates…



The scanner displays the scan results on the left. Since we selected “Scan multiple photos to individual files (single selection for each photo found)” the scanner software will attempt to identify the pictures. The individual pictures are displayed at the bottom left under “Final Images”

The selected image is highlighted in blue in Final Images. It can be deleted. Add new selection will provide a new crop selection that you can manually place.

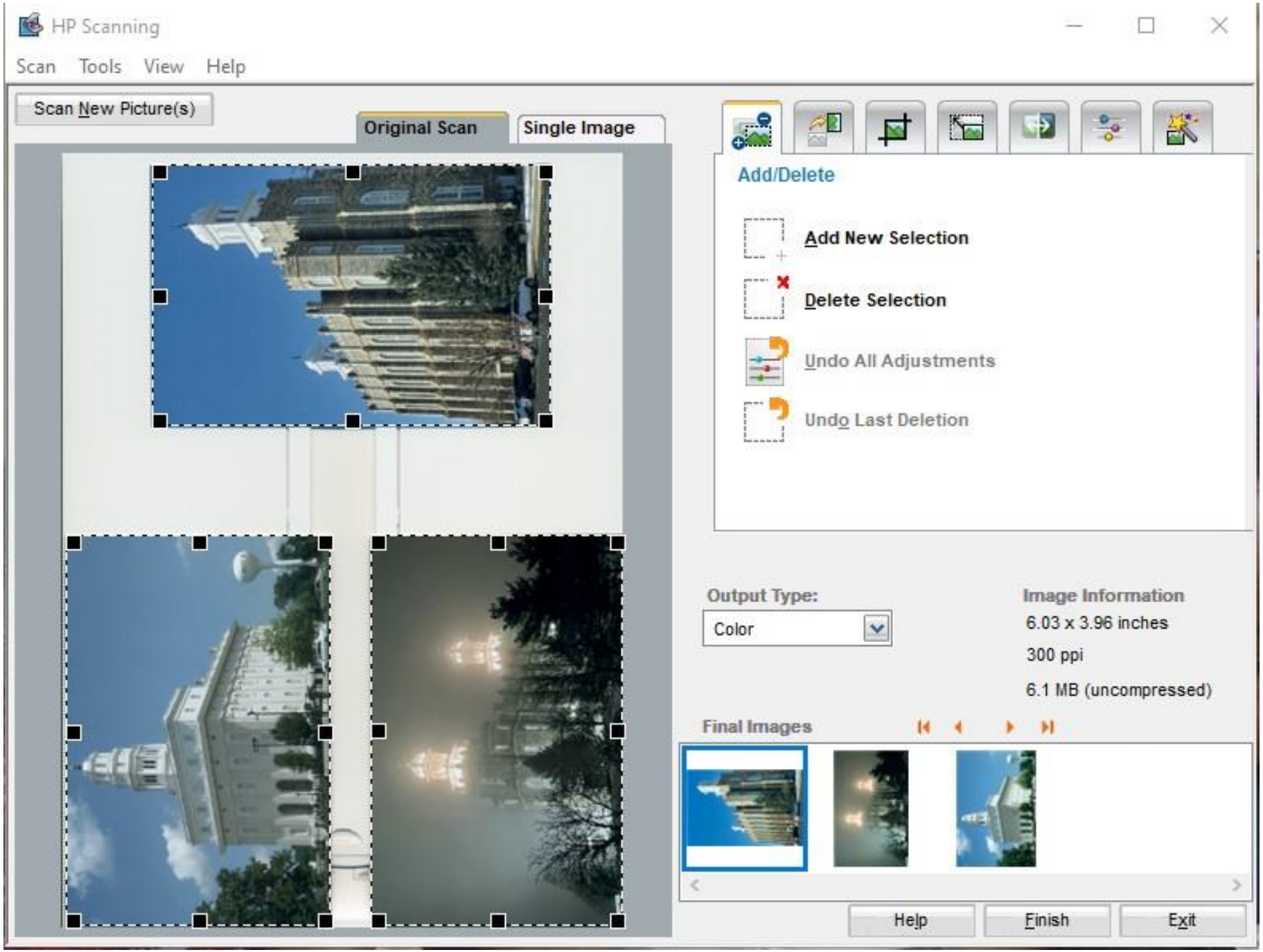


Another Example

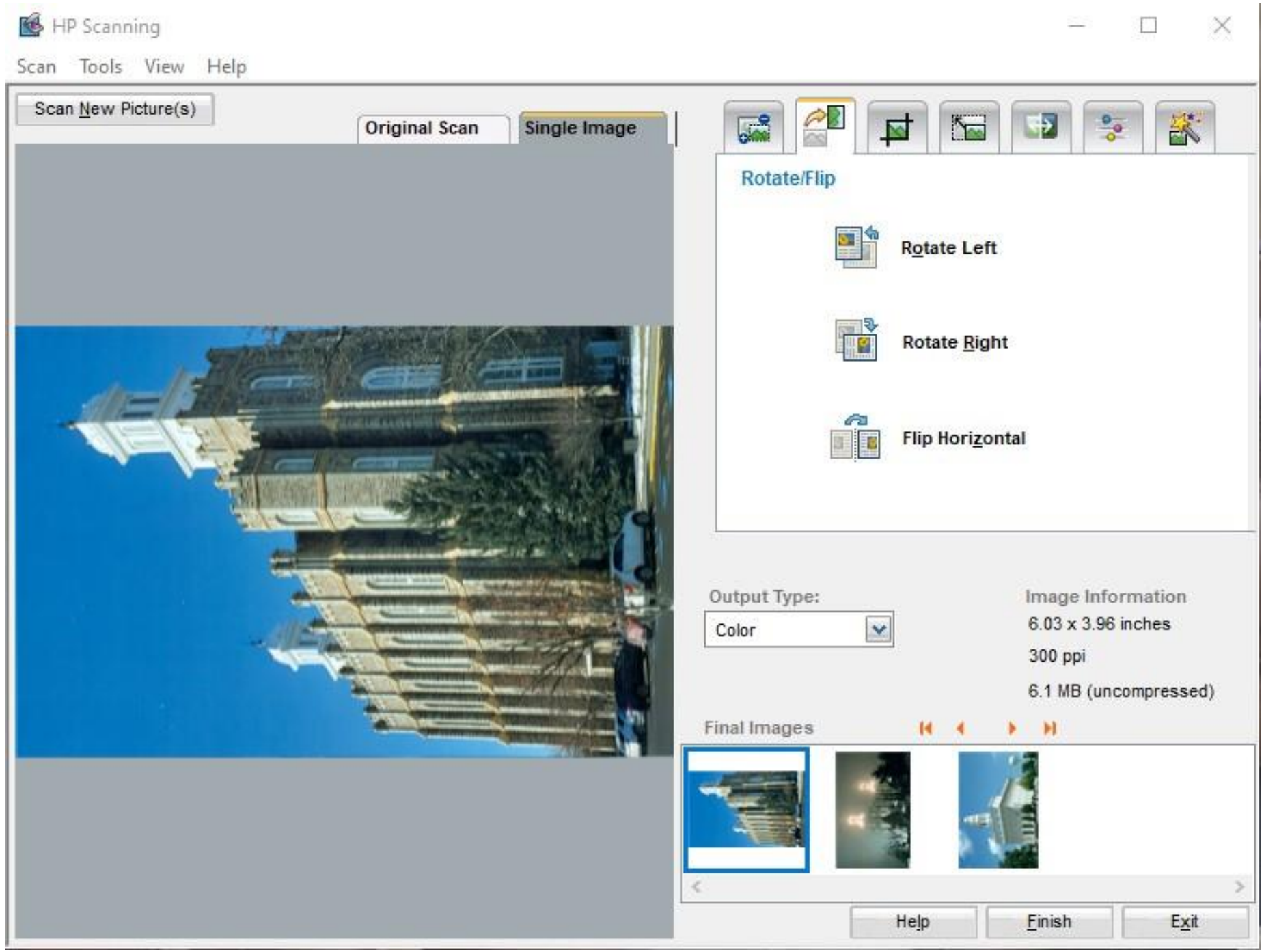
The Original Scan Tab is selected rather than Single Image.

There are 7 tabs on the right for editing the picture.

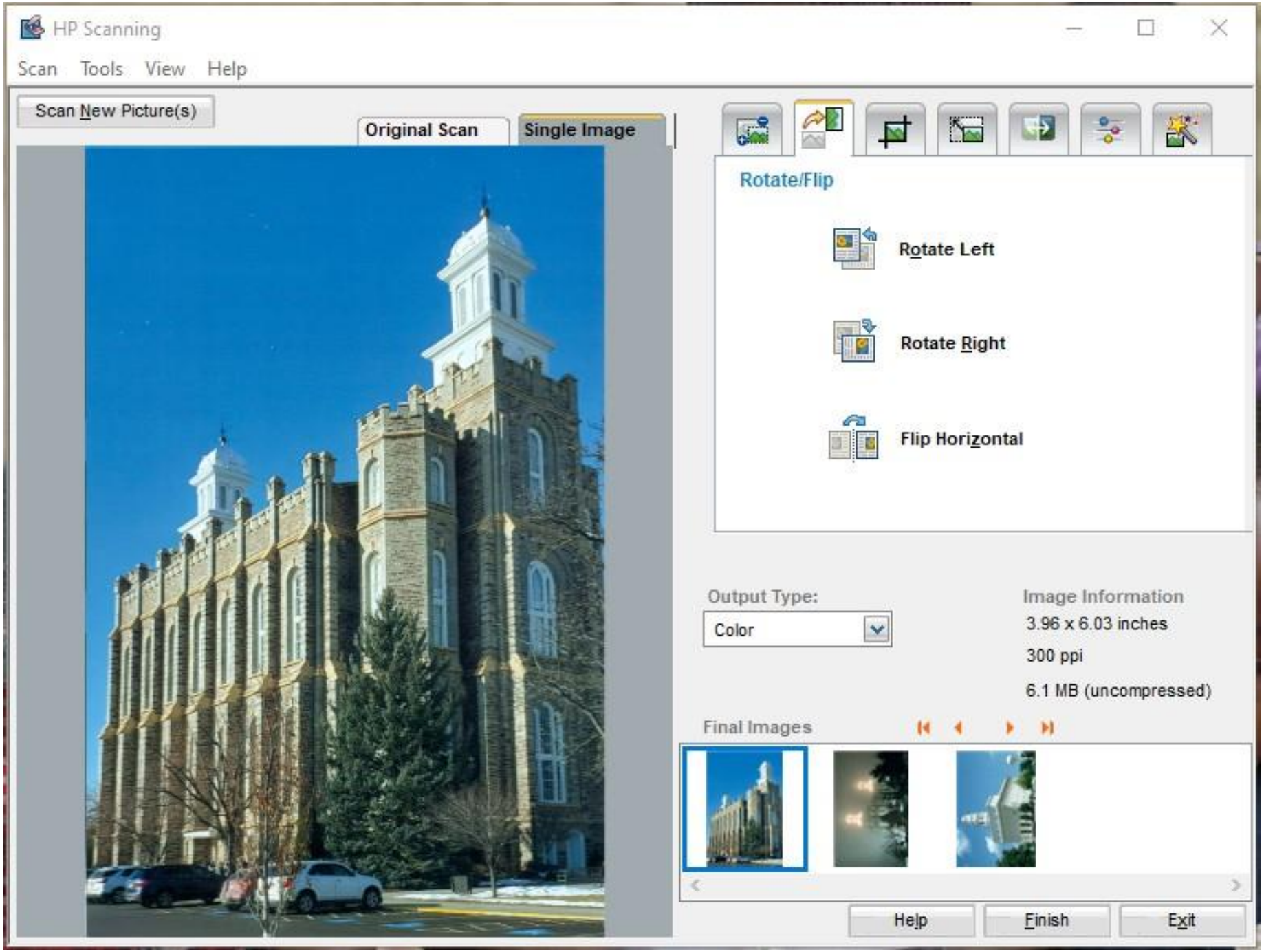
You can select anyone of the three images. The selected image has a blue outline around it in the Final Images



Selecting Single Image displays the selected image And moves from the Add/Delete Tab to the Rotate/Flip Tab

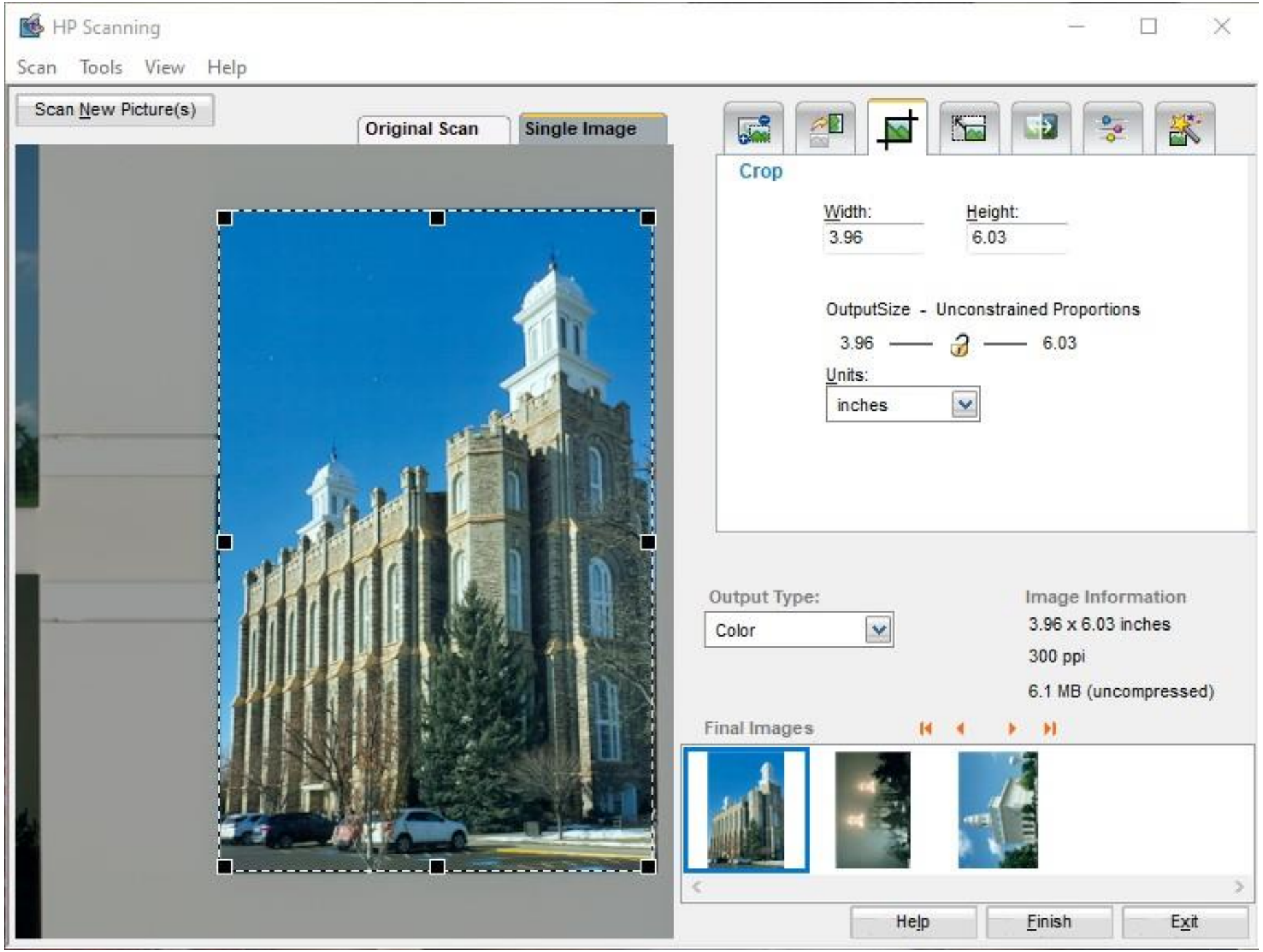


The image was rotated to be upright

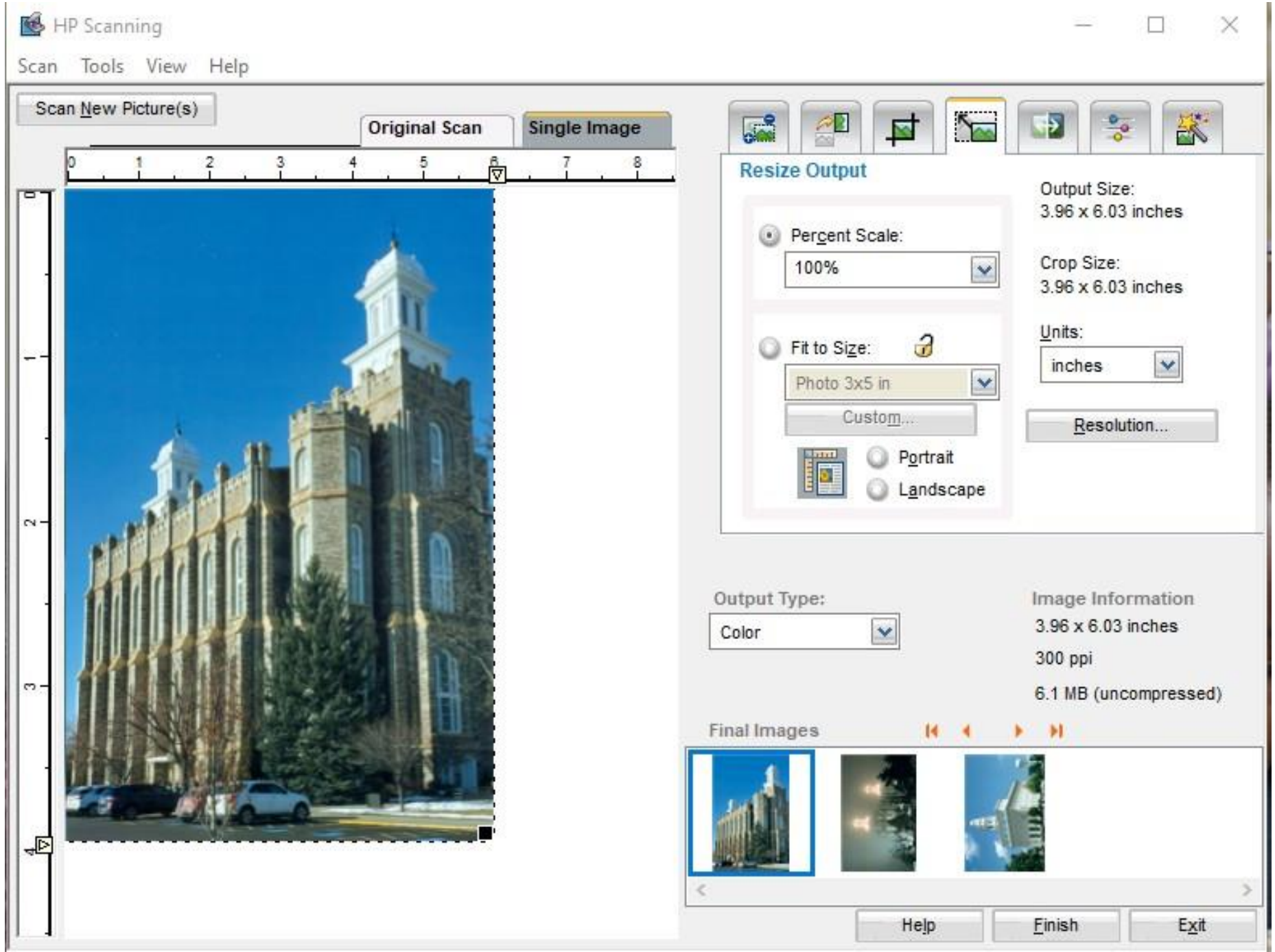


Crop

With finer Controls

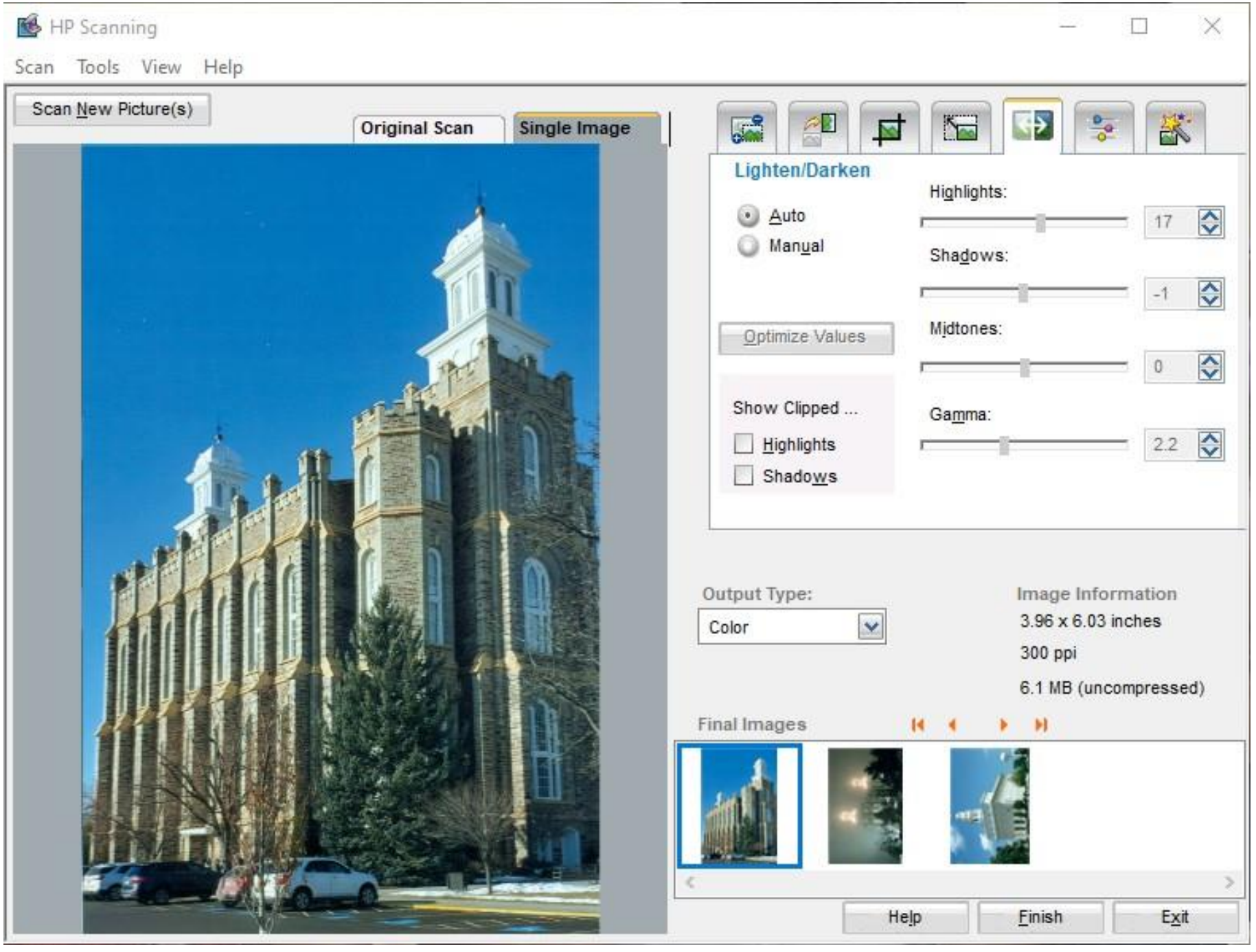


Resize Output



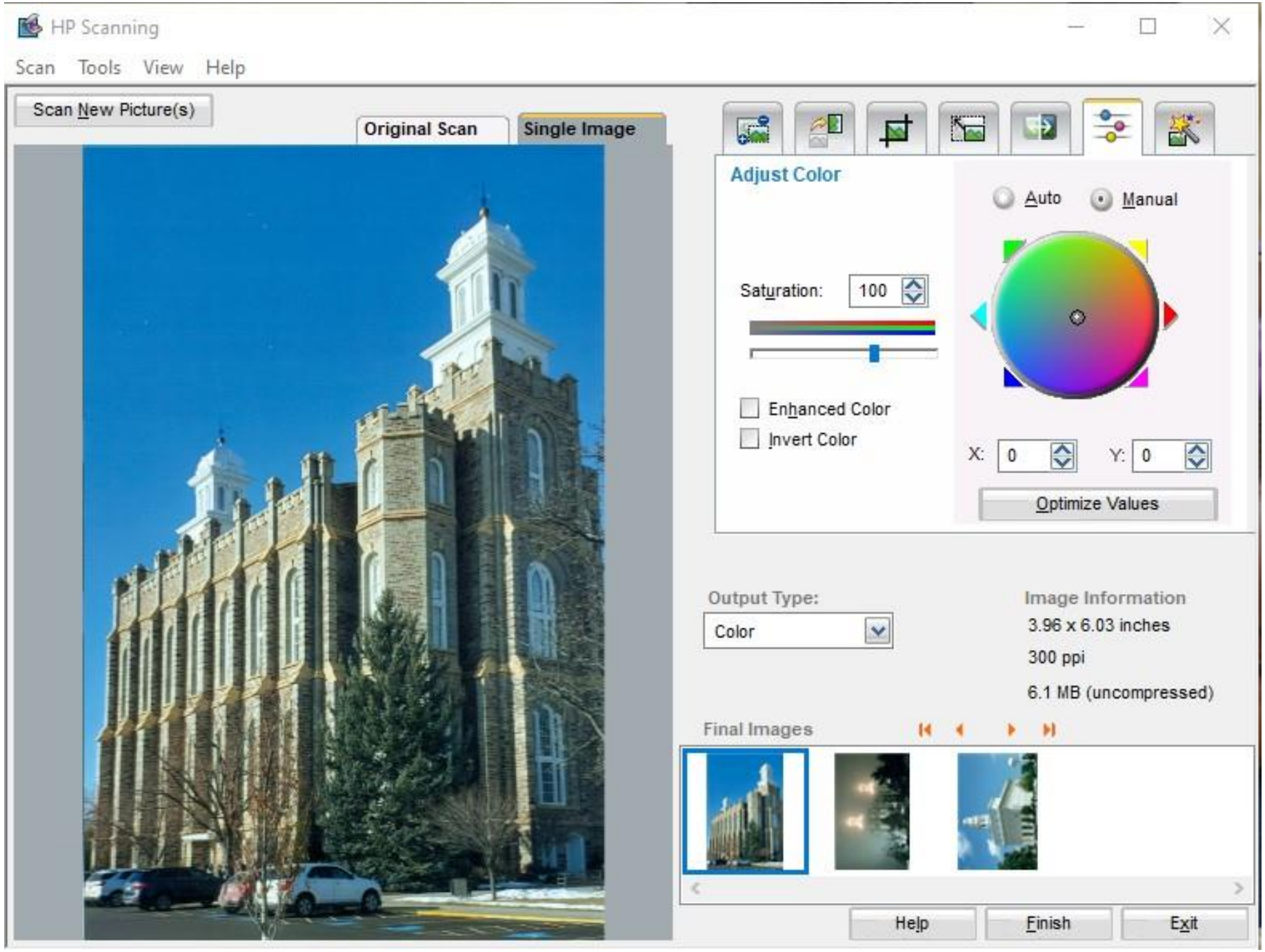
Midtones are the areas of an image that are neither bright nor dark. Highlights refer to the bright parts of an image, and the shadows are the dark parts.

Gamma can be described as how smoothly black transitions to white on a digital display.



Adjust Color

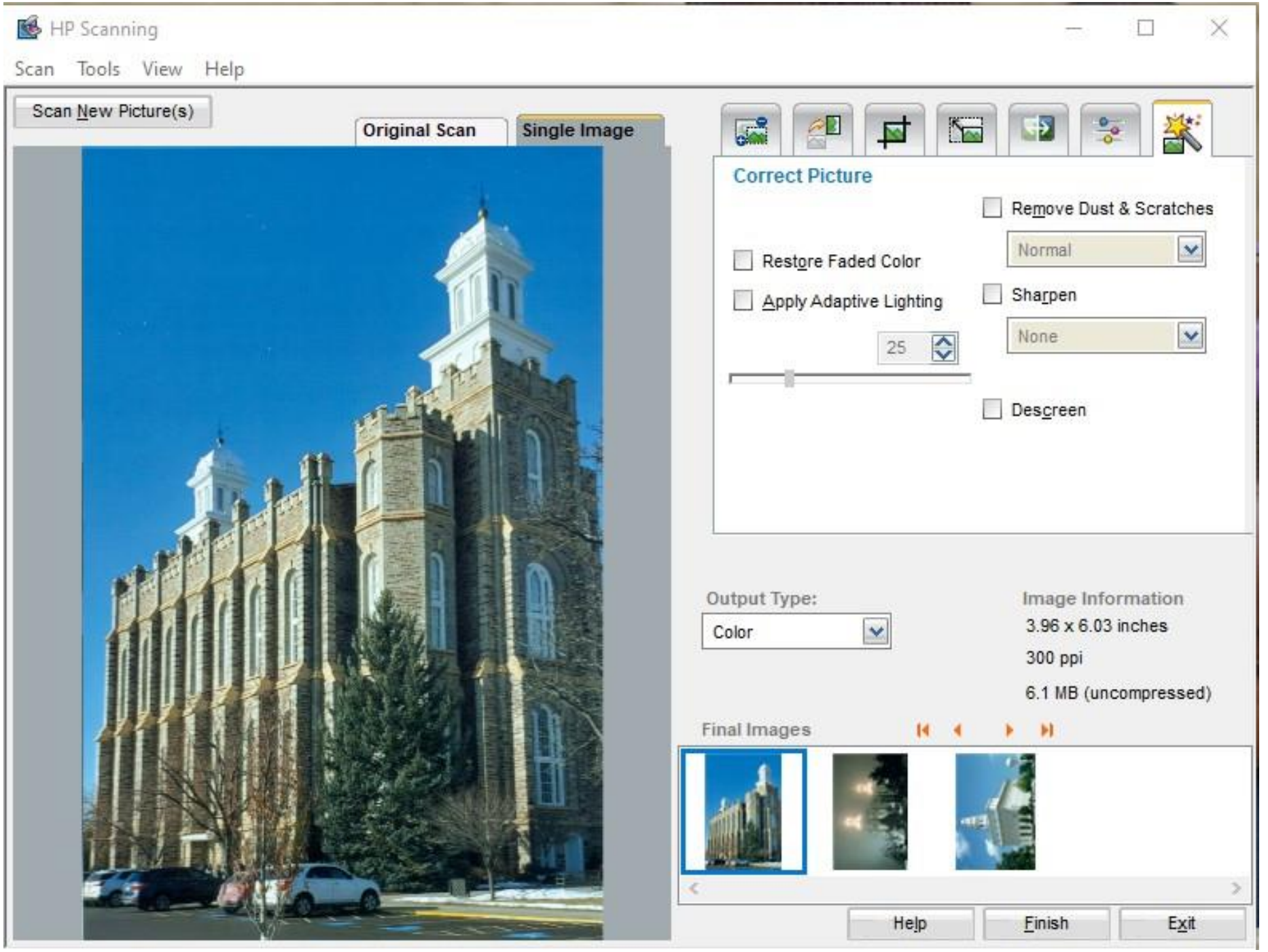
Enhanced Color
Invert Color
Saturation Adjustments
Color Controls
Auto or Manual
Optimize Values



Correct Picture

- Restore Faded Color
- Apply Adaptive Lighting
- Remove Dust & Scratches
- Sharpen
- Descreen

Adaptive lighting may lighten dark areas in the scene, or darken light areas



Sharpen: Focus edges to increase clarity in an image. Sharpen is the opposite of blurred. (Image sharpening refers to any enhancement technique that highlights edges and fine details in an image.) Sharp is the opposite of blur.

Descreen: Reduce undesirable patterns in scans of printed items. Specifically half toned newspaper and magazine printed items. It removes Moiré-pattern artifacts that are artifacts of scanning half toned print items.

Moiré patterns occurs in a image when a scene or an object being scanned contains repetitive details (dots, lines, checks, stripes) that exceed the sensor resolution. The scan produces a strange-looking wavy pattern


George Fox

Scanning a text file

FamilySearch only accepts PDF documents for uploading text documents like histories to memories.

The HP Scanning software supports options for generating pdf documents from a printed original.

There are two options for scanning documents

1. You can scan a text document and create an image file (a compressed photograph) of the text.

2. You can scan the text document and with some effort, create an editable text file that enables you to edit the text, reformat the document or include it in another document.

Filename.pdf – Portable Document Format

Created by Adobe Inc, a technology company that makes publishing software including printer drivers (Adobe Postscript) and photo editors (Adobe Photoshop). Adobe created the portable document format in 1993. Adobe provides free readers (Adobe Acrobat Reader) and charges for the software to create and edit pdf document. PDF was based on printer driver technology and has since become in international standard.

PDF files are the only option for uploading text documents to FamilySeach.

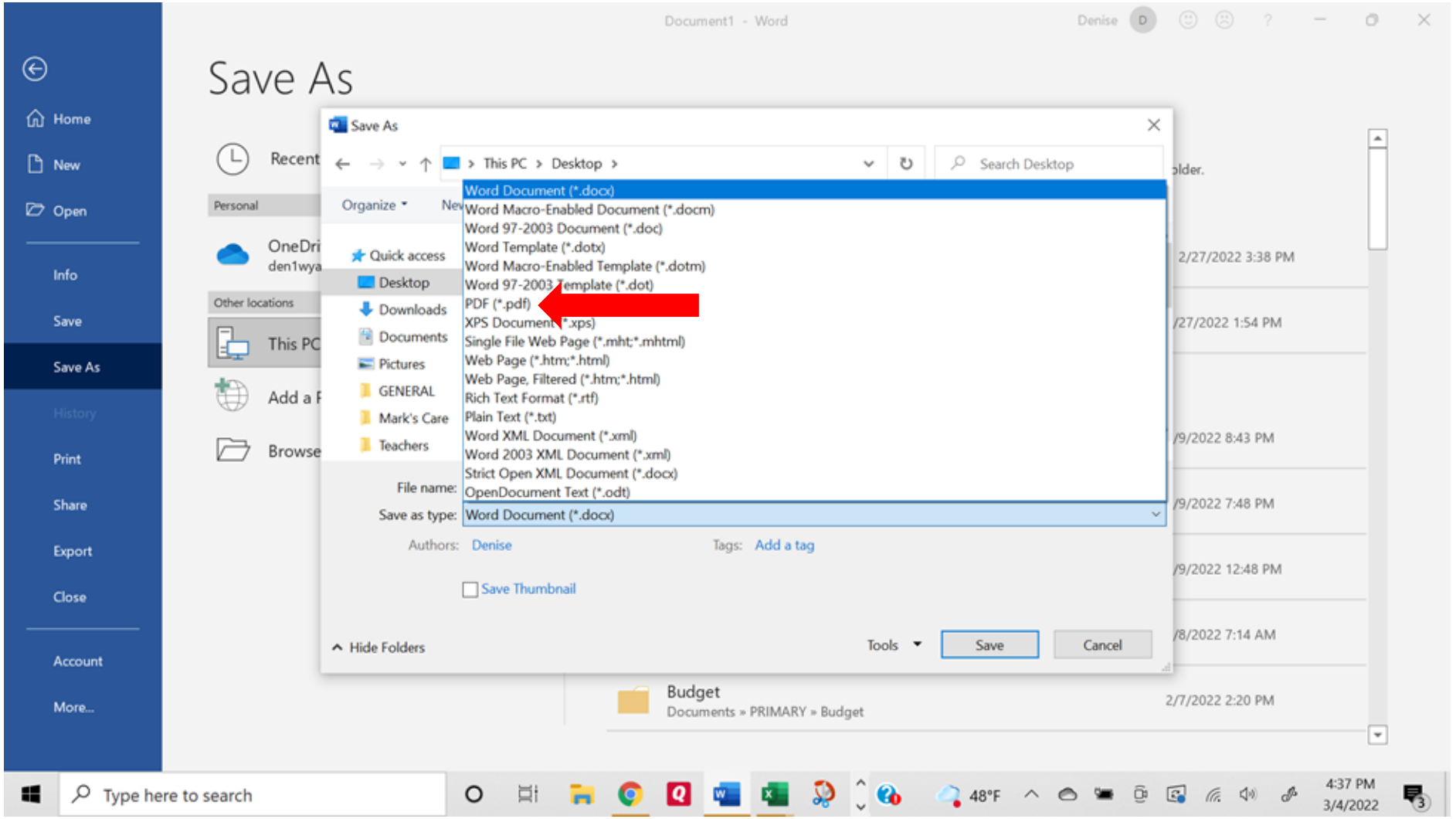
PDF files can be created by a save operation of a document file in Microsoft Word, or an export operation in LibreOffice Writer.

They can also be created directly by scanning a document on the HP Scanjet G3110.

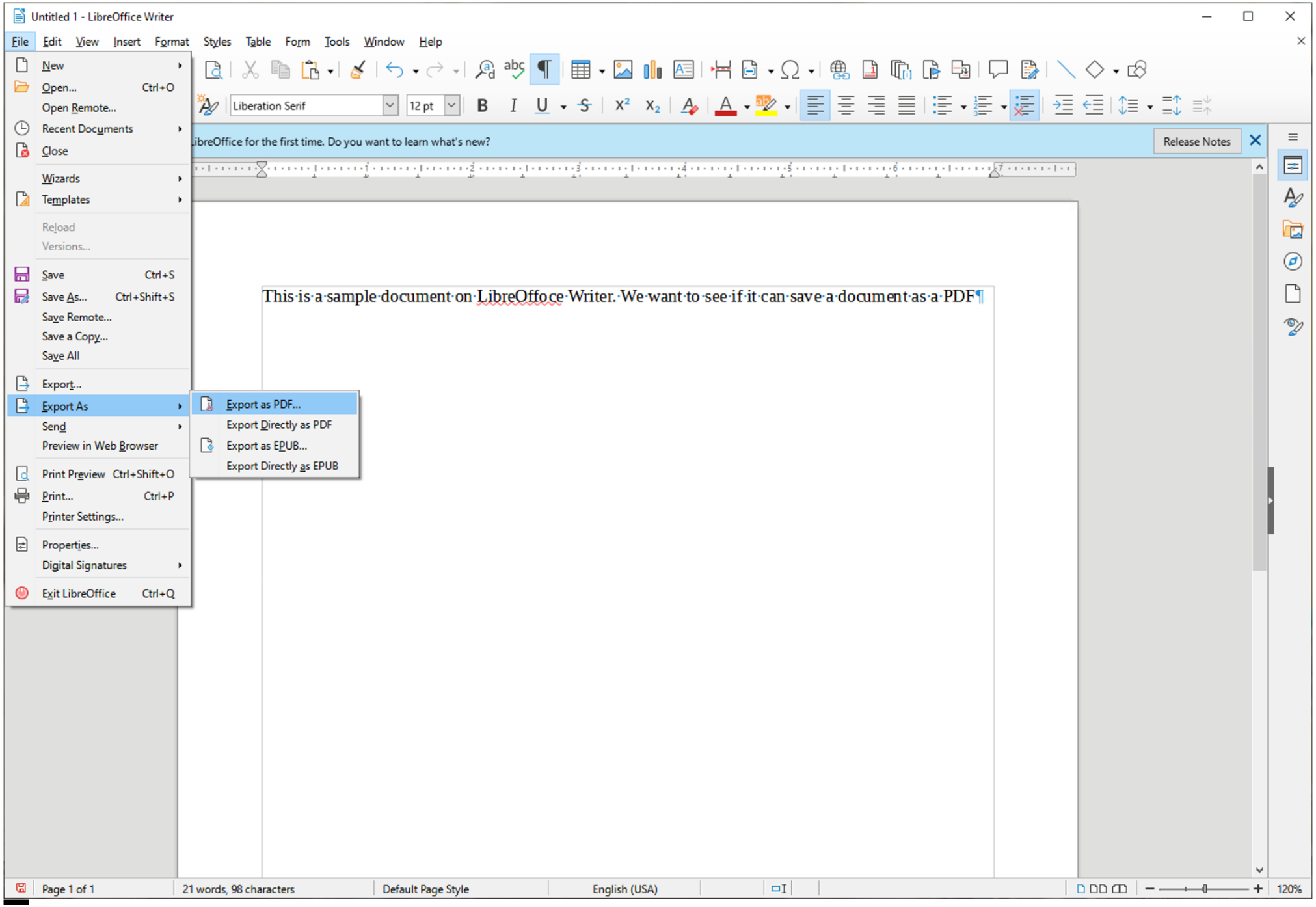
Word processors have the option of saving a file as a PDF.

Here is an example from Microsoft Word.

The options are found in a “Save As type:” drop down menu



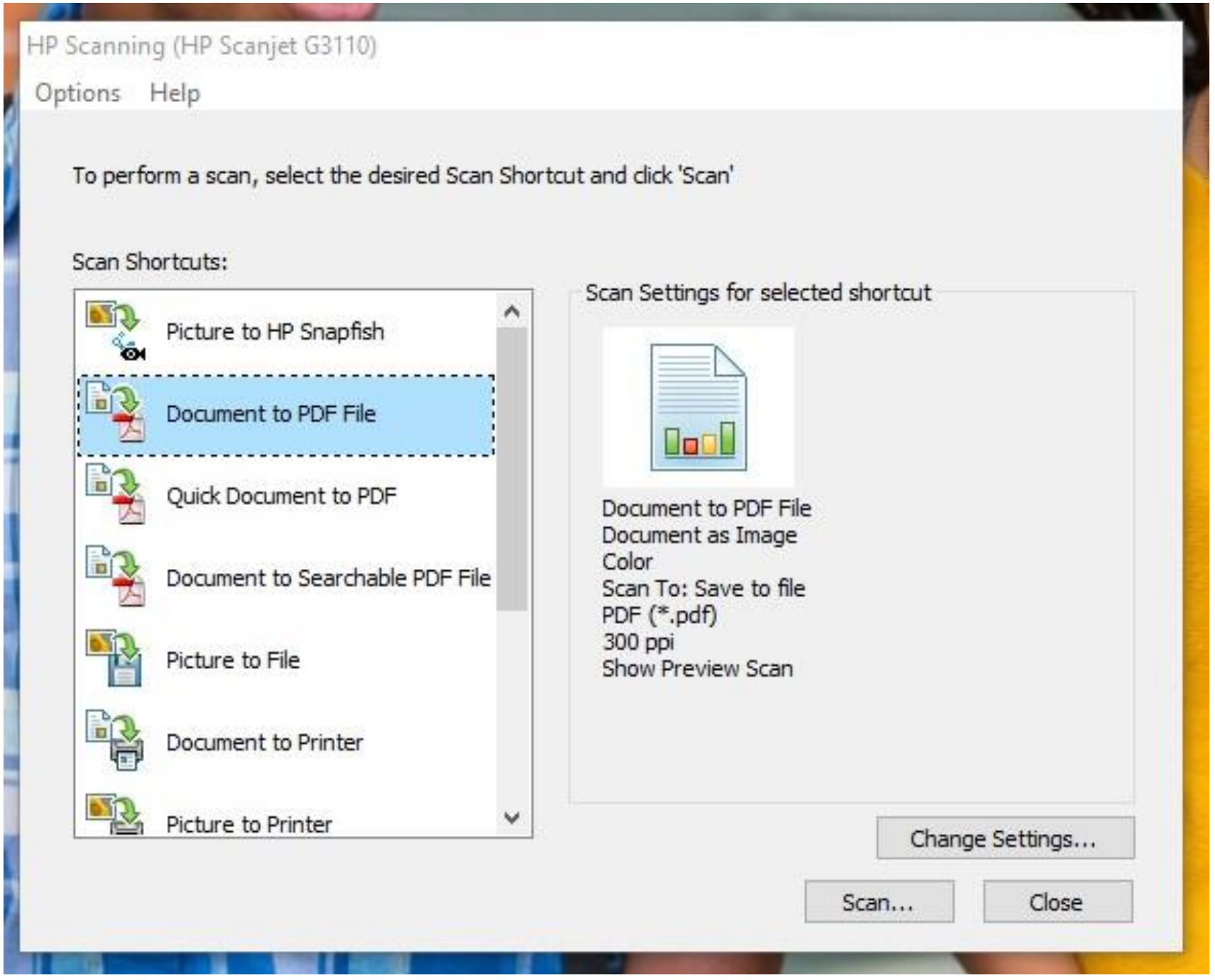
In LibreOffice Writer you can export a document as a PDF



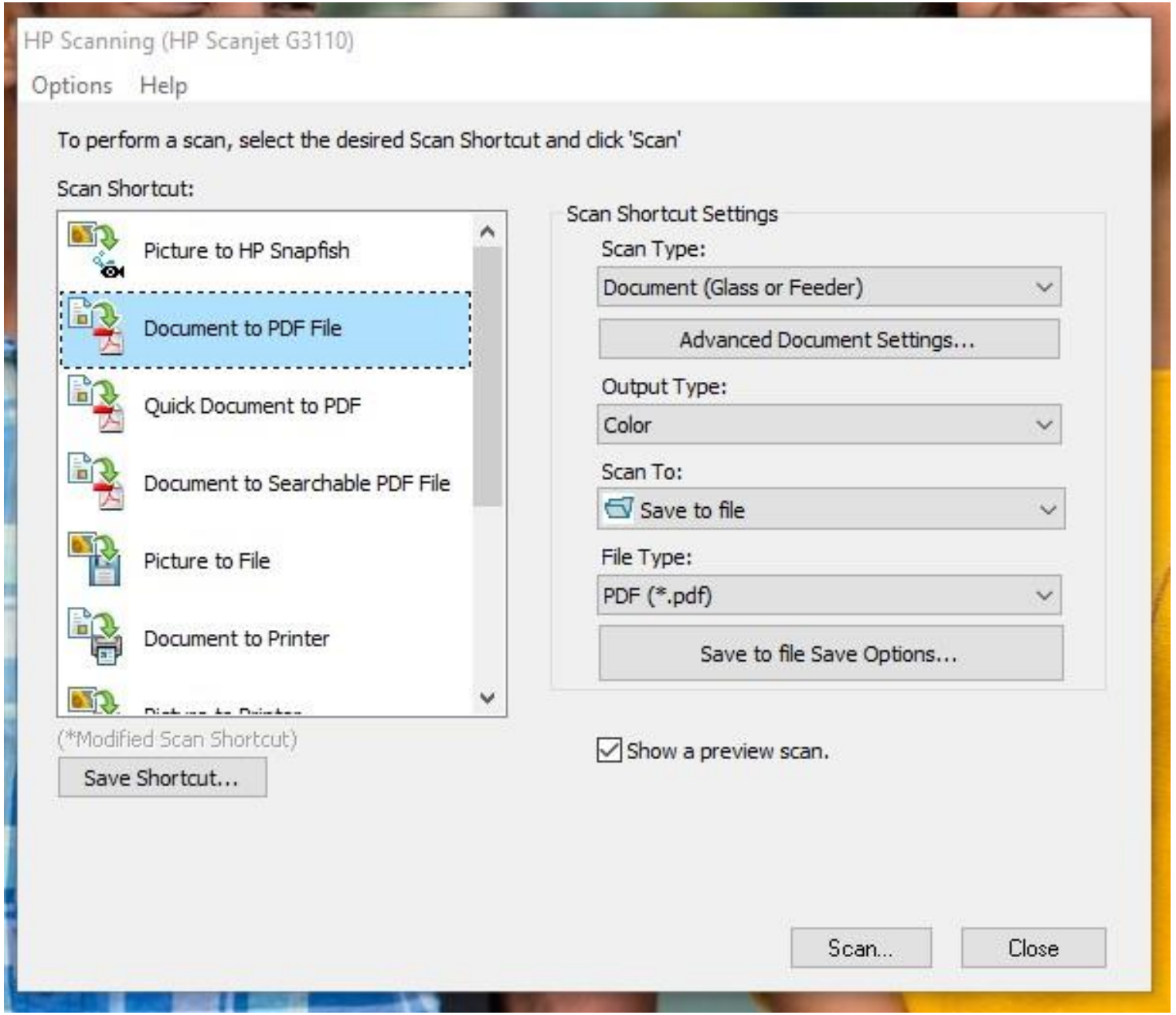
Scanning a text document to a PDF file using the HP Scanjet G3110 .

Click the Icon to open the HP Scanning page.

Select “Document to PDF File”
Then “Scan…”

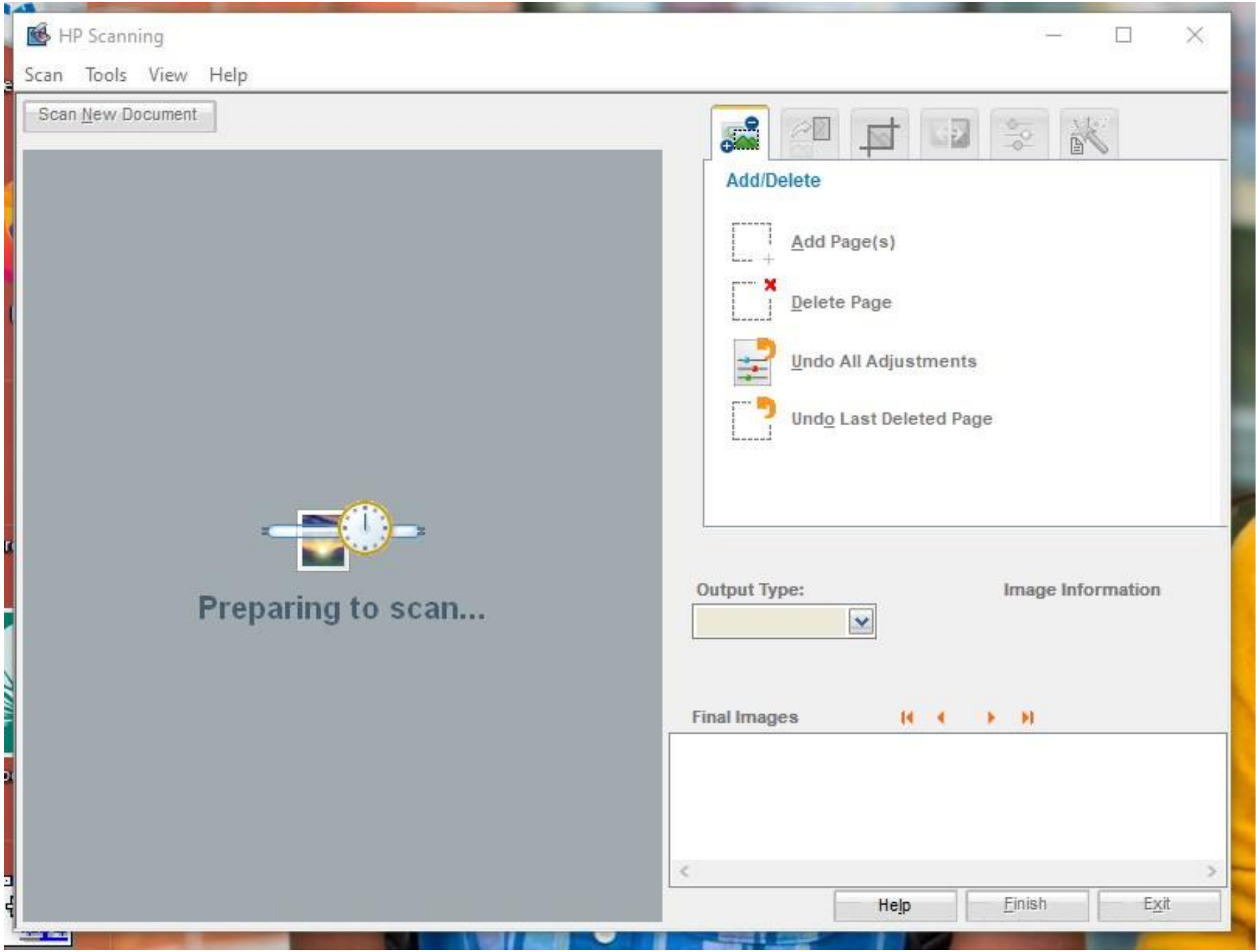


2. The HP Scanning Window updates.
Note that the “Output Type:” Color works well for many applications. The defaults are fine.
Use the “Scan to:” to select the scan file name and where to save it.
Place the document on the glass, close the lid and press “Scan…”
The scanner takes its time before responding.

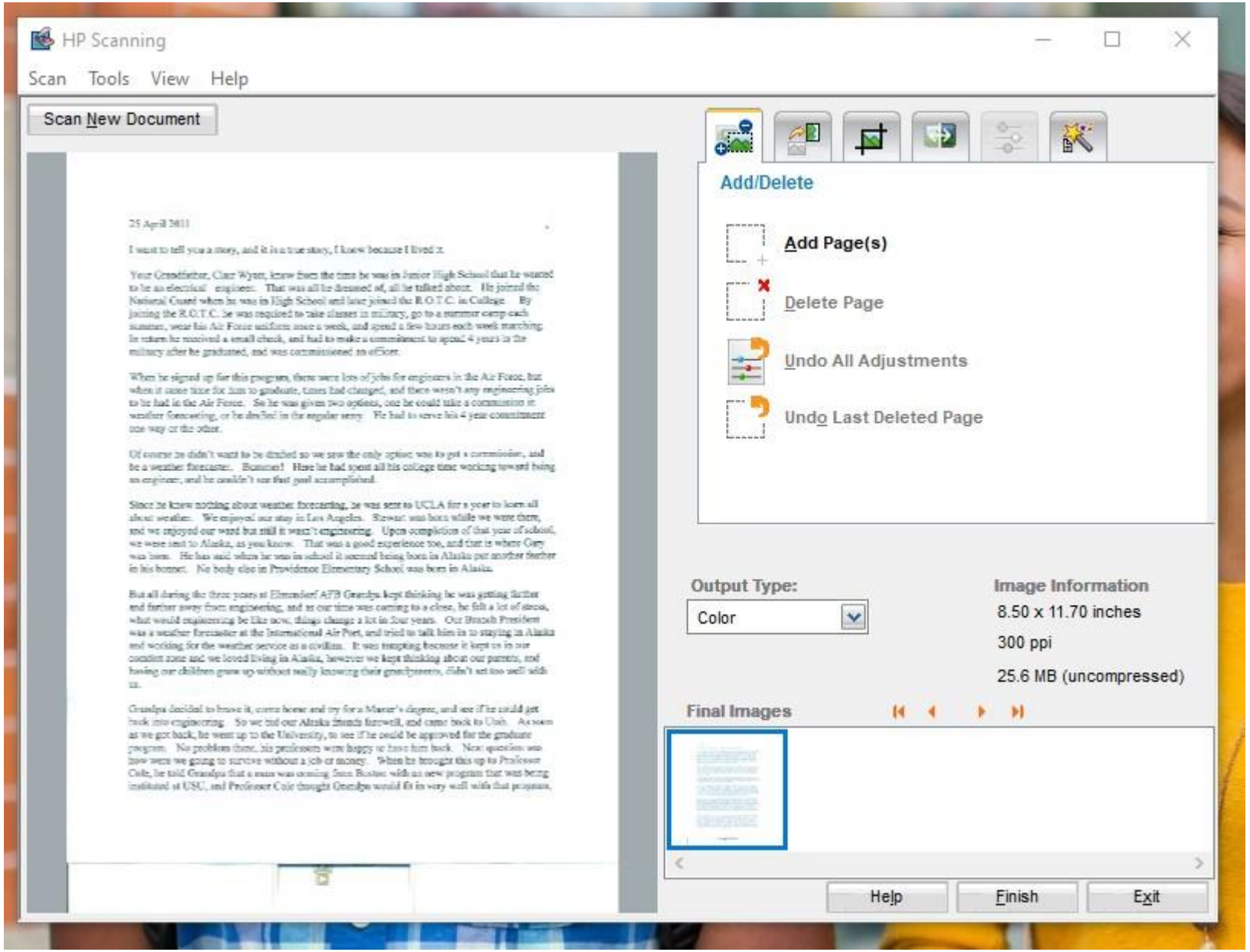


3. The Scanning window opens with a “Preparing to scan…” message.

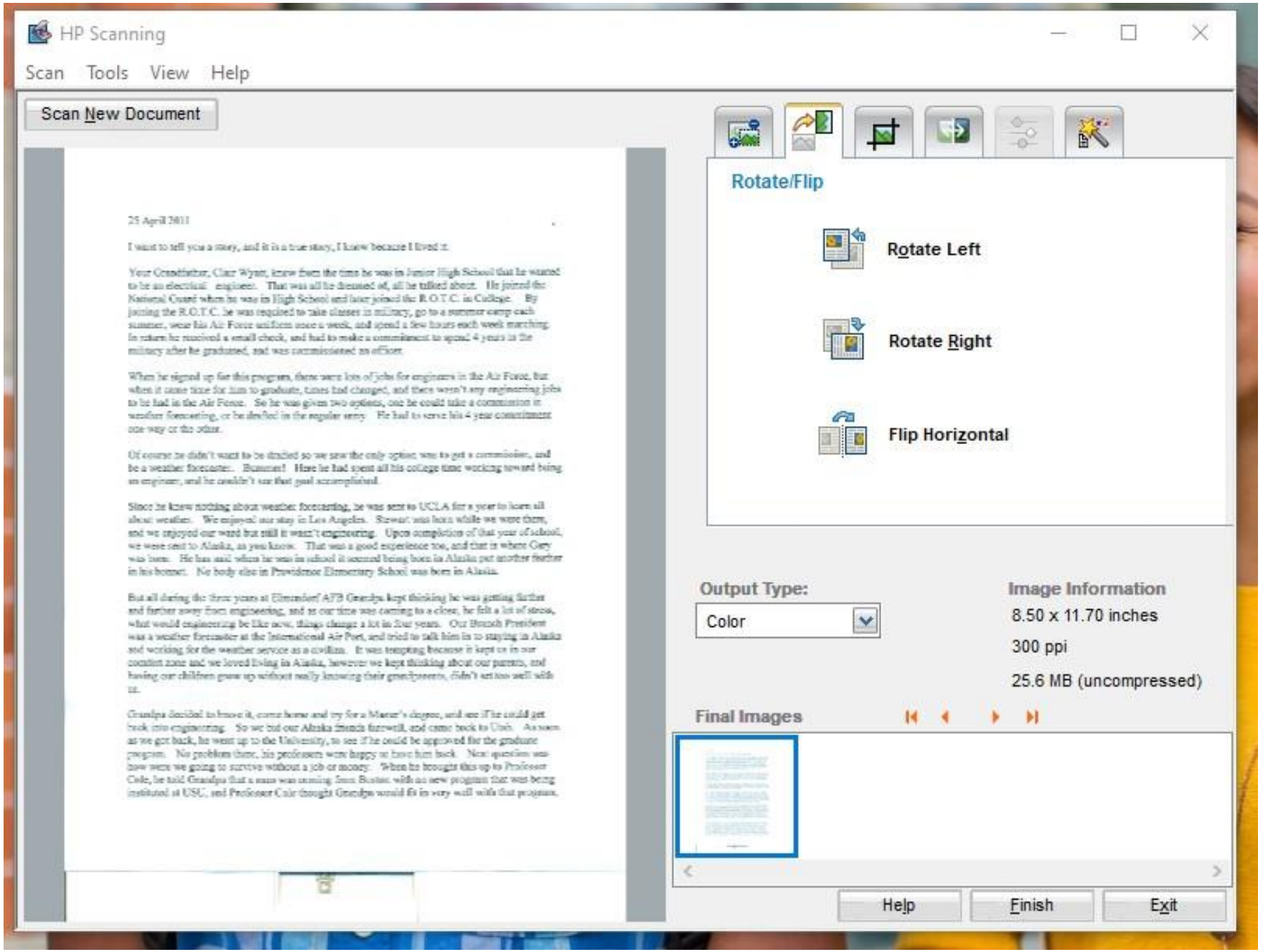
The scanner lamp will turn on and you will hear the scan motor running. The lamp will move with the scan bar.



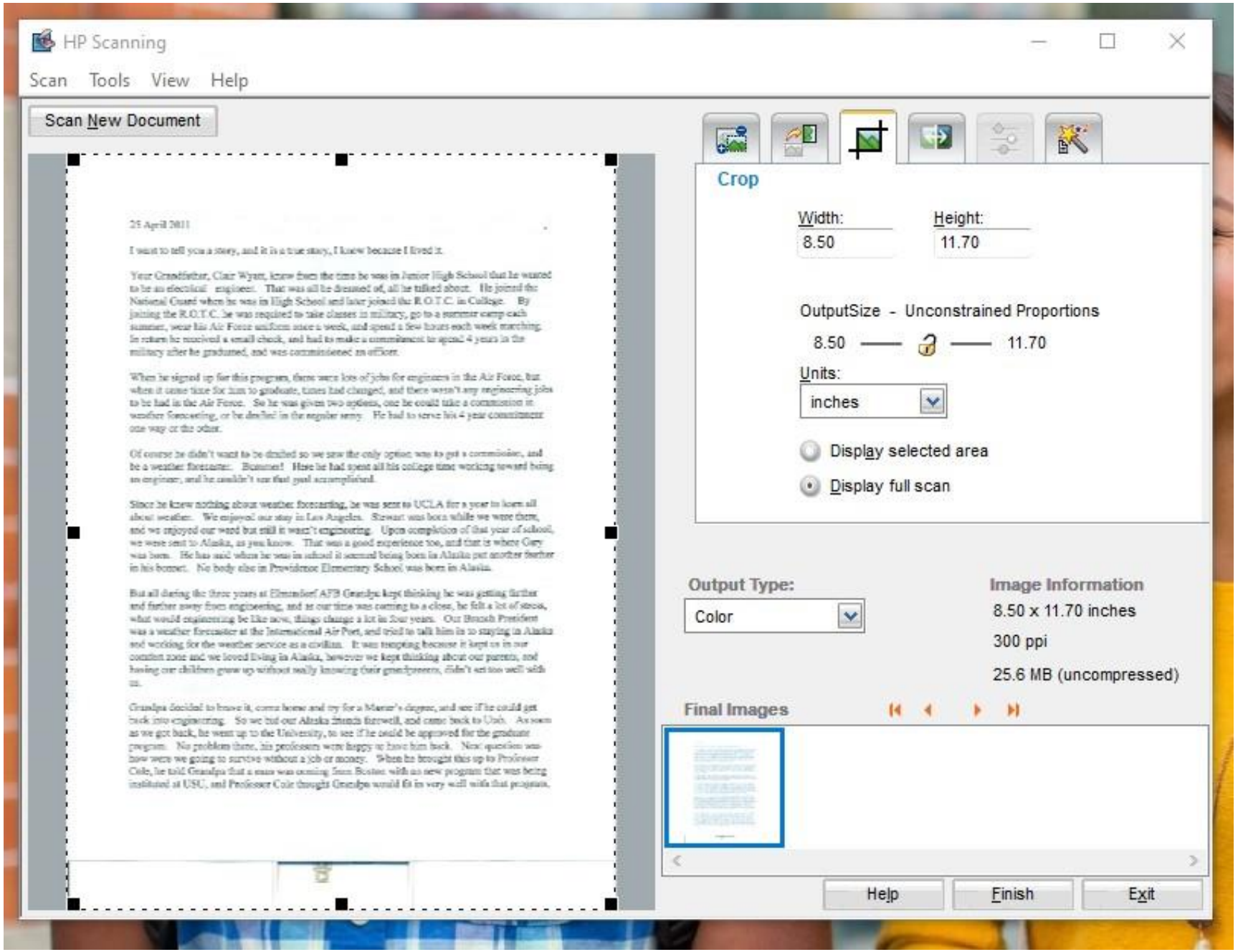
4. The Scan completes and the window displays the most recent scanned image on the left. The editing options opens to the “Add/Delete” tab.



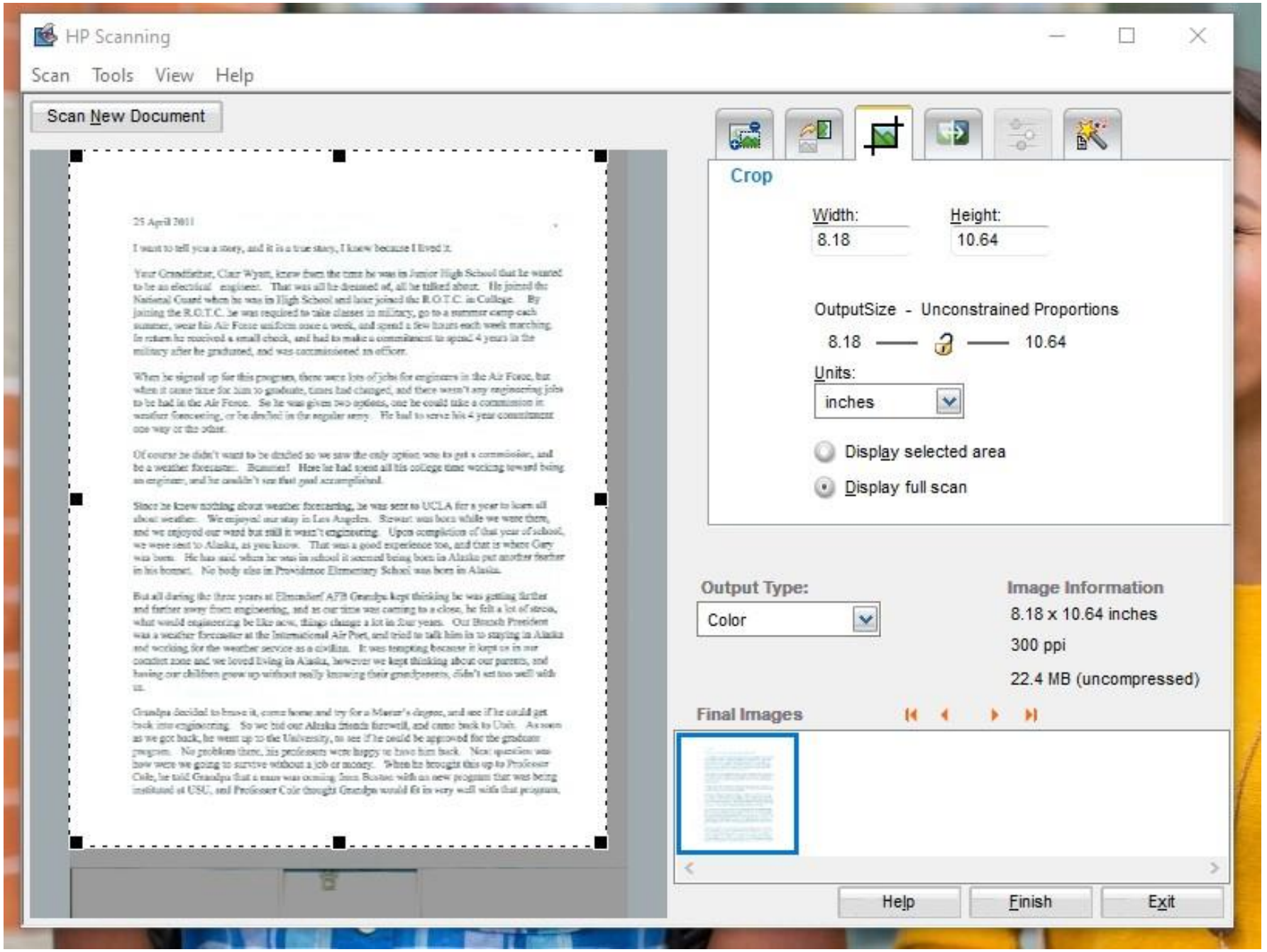
5. Select the 2nd tab, “Rotate/Flip” which allow you to rotate or flip the document if needed.



6. Select the 3rd tab, “Crop”. The scan always covers the entire scan bed, 12x9 inches. When scanning an 8.5 x 11 page, the scan image will include the bottom inch of the lid. Use the crop option to return the image to the size of the scanned page by dragging the black squares with your mouse. The dash lines mark borders of the saved scan. The actual crop occurs when you change tabs.



Here the crop lines have been adjusted to the size of an 8.5 by 11 page.

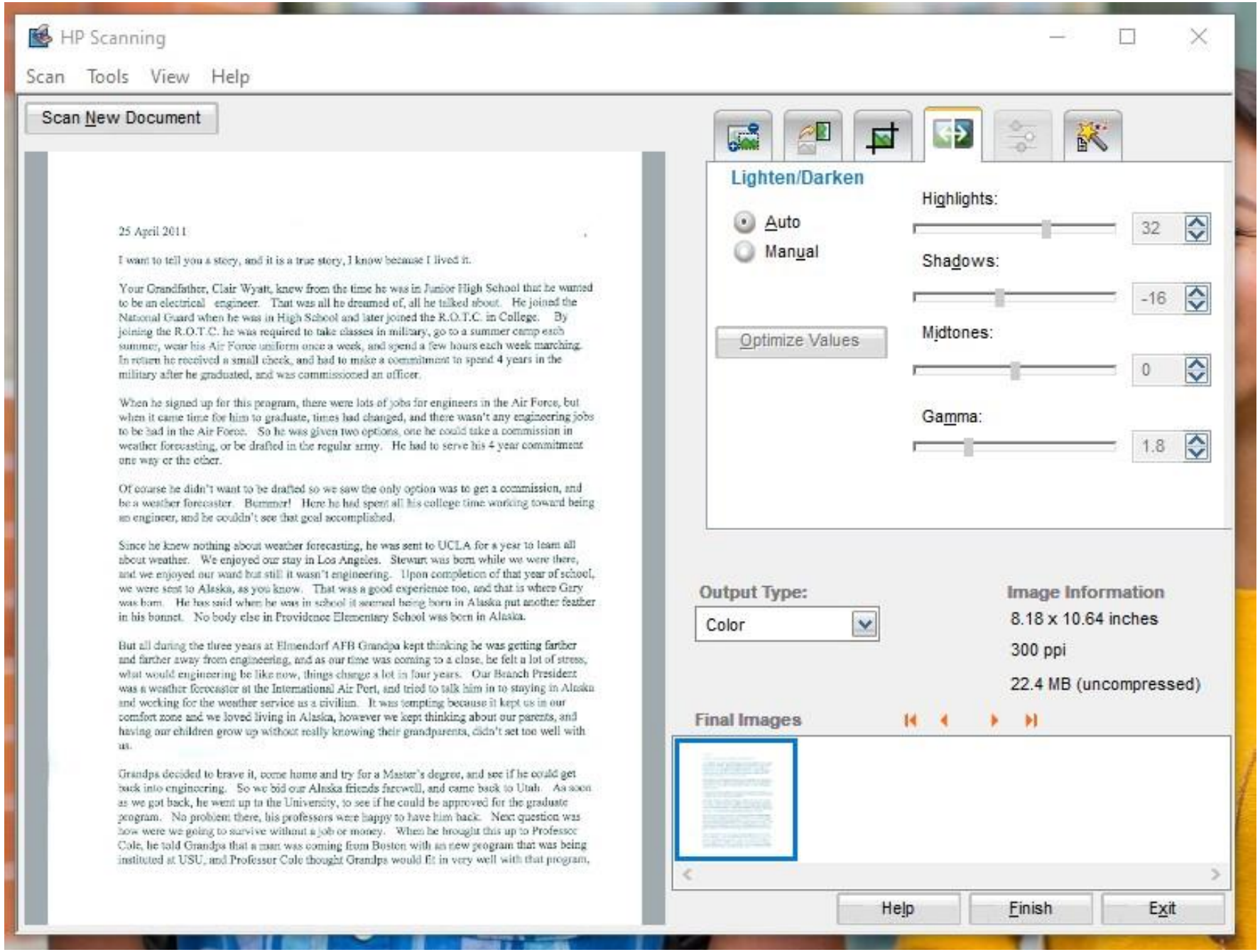


7. Select 4th tab, “Lighten/Darken” to adjust highlights, shadows , midtones and Gamma.

Midtones are regions that are neither dark or light.

Gamma can be described as how smoothly black transitions to white on a digital display.

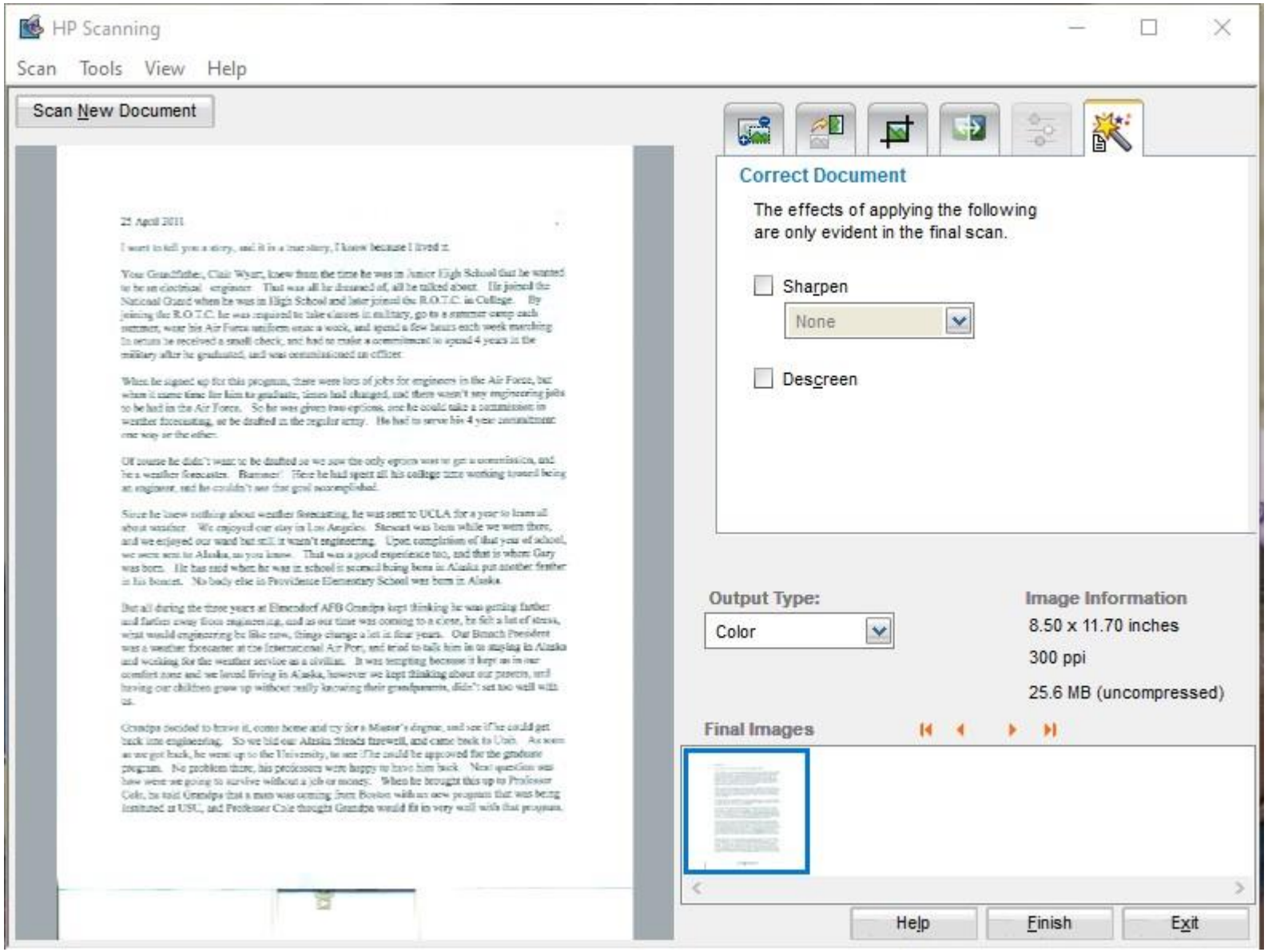
Note the 5th tab is unused in this application.



8. The 6th and last tab is titled “Correct Document” and states that the effects of applying the following are only evident in the final scan.

Sharpen: Focus edges to increase clarity in an image. Sharpen is the opposite of blurred. (Image sharpening refers to any enhancement technique that highlights edges and fine details in an image.)

Descreen: Reduce undesirable patterns in scans of printed items. Specifically half toned newspaper and magazine printed items. It removes Moiré-pattern artifacts that are artifacts of scanning half toned print items.

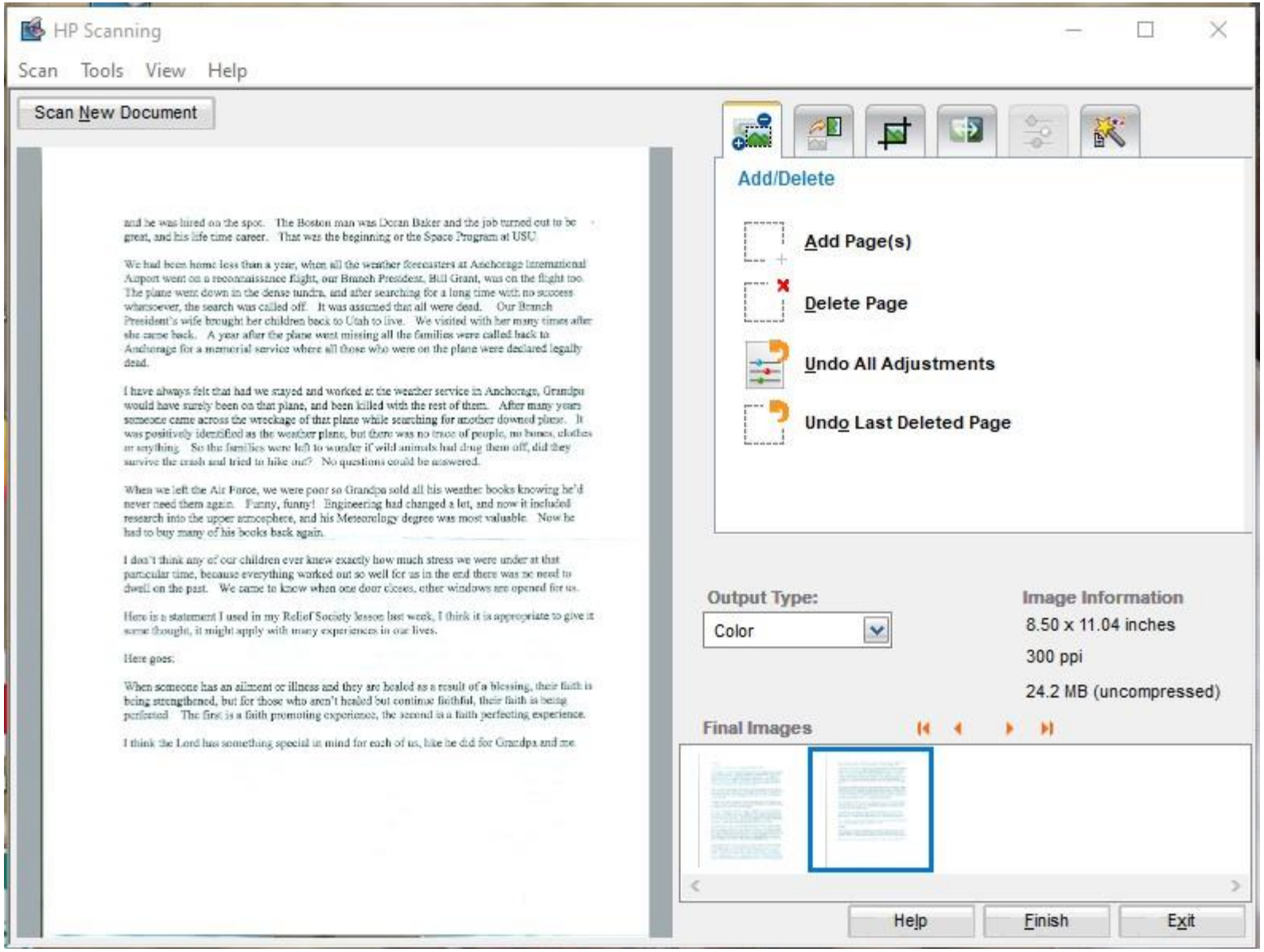


Having completed editing the first page , we return to the 1st tab, “Add/Delete” where we have options including “Add Pages”.

All previous scanned pages are visible at the bottom right in the “Final Images” pane.

Load the next page on the scan bed before selecting “Add Pages”

When you have scanned all the pages, select “Finish”.



Continue the process until all pages are scanned.

When you have scanned the entire document and everything looks good, select “Finish”. And the document will be created with all of the pages linked together in a single document and substantially compressed.

In this example I didn’t set the “Scan to:” in the “Scan Short Cut Page” so the scanning software will use the defaults. The document will be stored at C:\Users\Patron\Document\My Scans and will be called scan000x. HP Scanning will open File Explorer at that location which you can rename, cut and paste into your flash drive.

HPScanning (G3110) Options

~~1. Picture to HP Snapfish~~

2. Document to PDF File (Document as Image)

~~3. Quick Document to PDF (Document as Image)~~

4. Document to Searchable PDF File [Searchable text (OCR). Save to file PDF-Searchable (*.pdf)]

5. Picture to file [Save to File Jpeg Image (*.jpg)]

~~6. Document to Printer~~ (Use the photocopier instead)

~~7. Picture to Printer~~ (Use the photocopier instead)

8. Text (OCR) to RTF File [Editable Text (OCR) Save to file Rich Text File (*.rtf)]

9. Text (OCR) to WordPad [Editable Text (OCR) Scan to WordPad]

~~10. Document to TIF File [Save to Tiff Image (*.tif)]~~ (A FamilySearch Option – but Tiff files are typically larger than pdf files.)

Definitions and explanations

Filename.**txt - Text Files**: The TXT/text file is a plain text file that does not contain any formatting information like font selection, bold or italic print, or either end of line or paragraph information. A file containing formatting information and saved as a text file will lose all formatting information. Microsoft has a text editing program called Notepad that can be found at Start -> Window Accessories -> Notepad

Filename.**rtf: Rich Text Format files** contain limited formatting information including line breaks. Microsoft offers a rtf editor called Wordpad which can be found at Start -> Window Accessories -> Wordpad.

OCR: Optical Character Recognition Software

A technology that recognizes text within a digital image. It is commonly used to recognize text in scanned documents and images.

OCR software can be used to convert a physical paper document, or an image into an accessible electronic version with text.

The 1950 Census effort will use OCR and artificial intelligence to initially read the census data. OCR can be used for processing all sorts of documents, from hand written to printed.

The HP Scanning software that came with the HP Scanjet G3110 can effectively recover text from a clean, typed document.

Using the HP Scanning software to scan an image containing text to create a text file.

The previous example scanned a text file to create a pdf image file. The Scanjet can use OCR software to create various types of text files (RTF) from a image containing text.

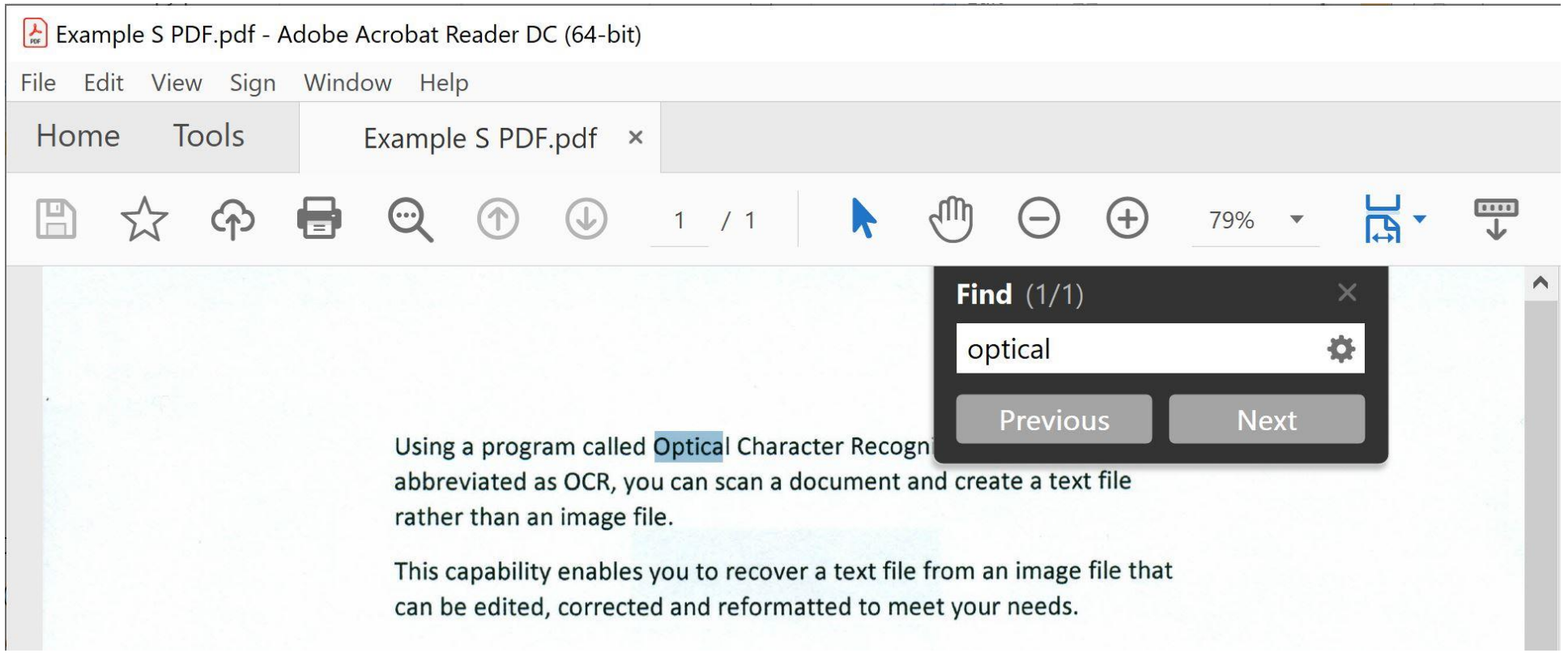
1. Document to Searchable PDF File [Searchable text (OCR). Save to file PDF-Searchable (*.pdf)]

2. Text (OCR) to RTF File [Editable Text (OCR) Save to file Rich Text File (*.rtf)]

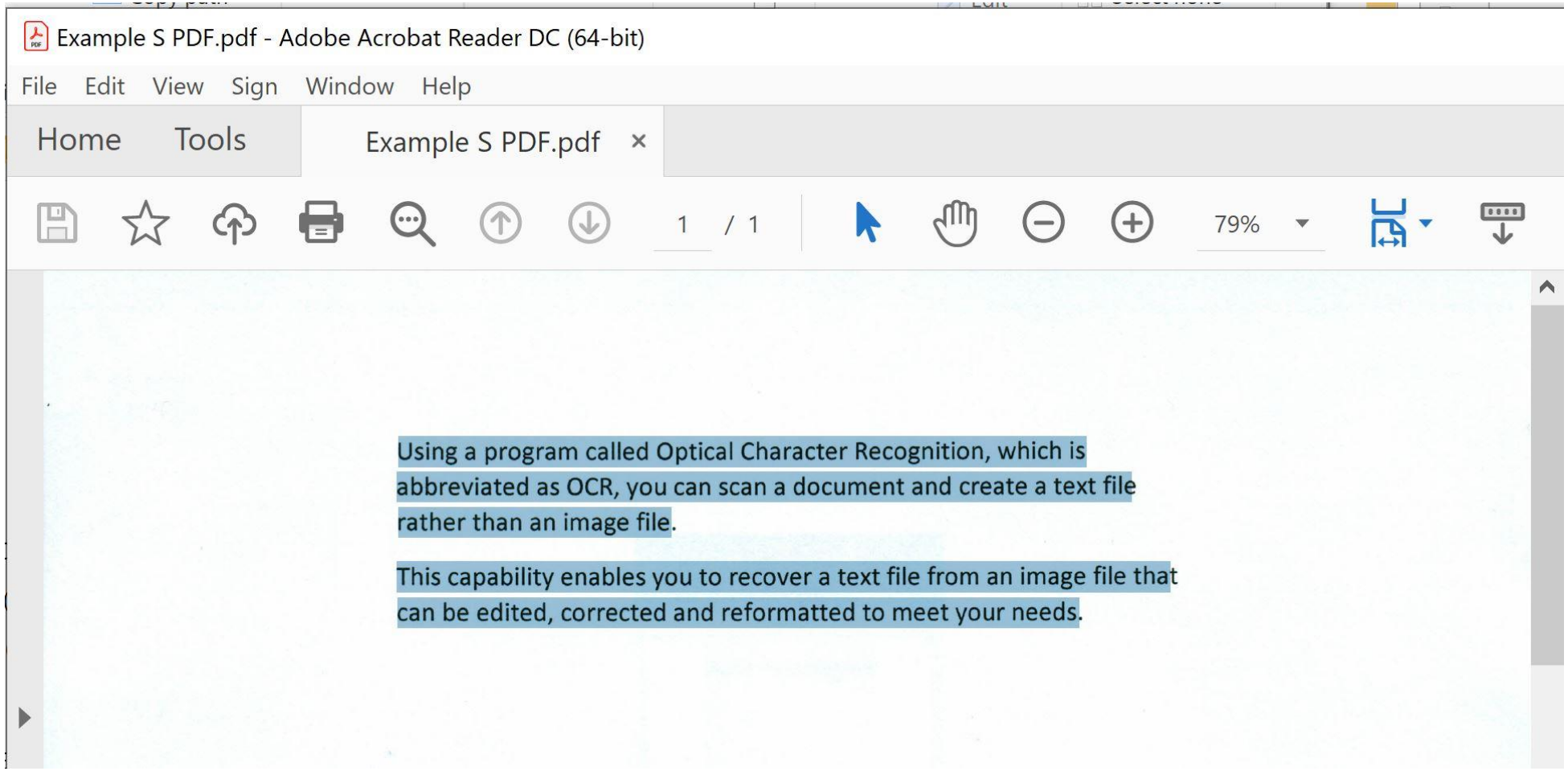
3. Text (OCR) to WordPad [Editable Text (OCR) Scan to WordPad]

Document to Searchable PDF File

With this option the document is scanned to a PDF file, but the file contains text rather than an image. This example illustrates searching in a scannable PDF document.



The text in a Searchable PDF can be edited, copied and pasted into another document.



Two other options: Text (OCR) to RTF File or Text (OCR) to WordPad. In the first case the scanner saves the scan to an RTF file. In the second case the scanner opens Microsoft WordPad (an RTF editor) loaded with the scan text results.

We will use the text (OCR) to RTF file to describe the process and challenges. There are two issues with the OCR process.

The first issue is with line breaks. The RTF format uses line breaks to mark the end of a line of text. This isn’t a problem with the RTF format itself, but becomes a problem if you edit the file or read it into a word processor like Word or LibreOffice Writer. Word processors can accommodate line breaks but typically only use paragraph breaks.

Here is the original document I scanned to RTF. The scanned version is a perfect copy with no errors.

Using a program called Optical Character Recognition, which is abbreviated as OCR, you can scan a document and create a text file rather than an image file.

This capability enables you to recover a text file from an image file that can be edited, corrected and reformatted to meet your needs.

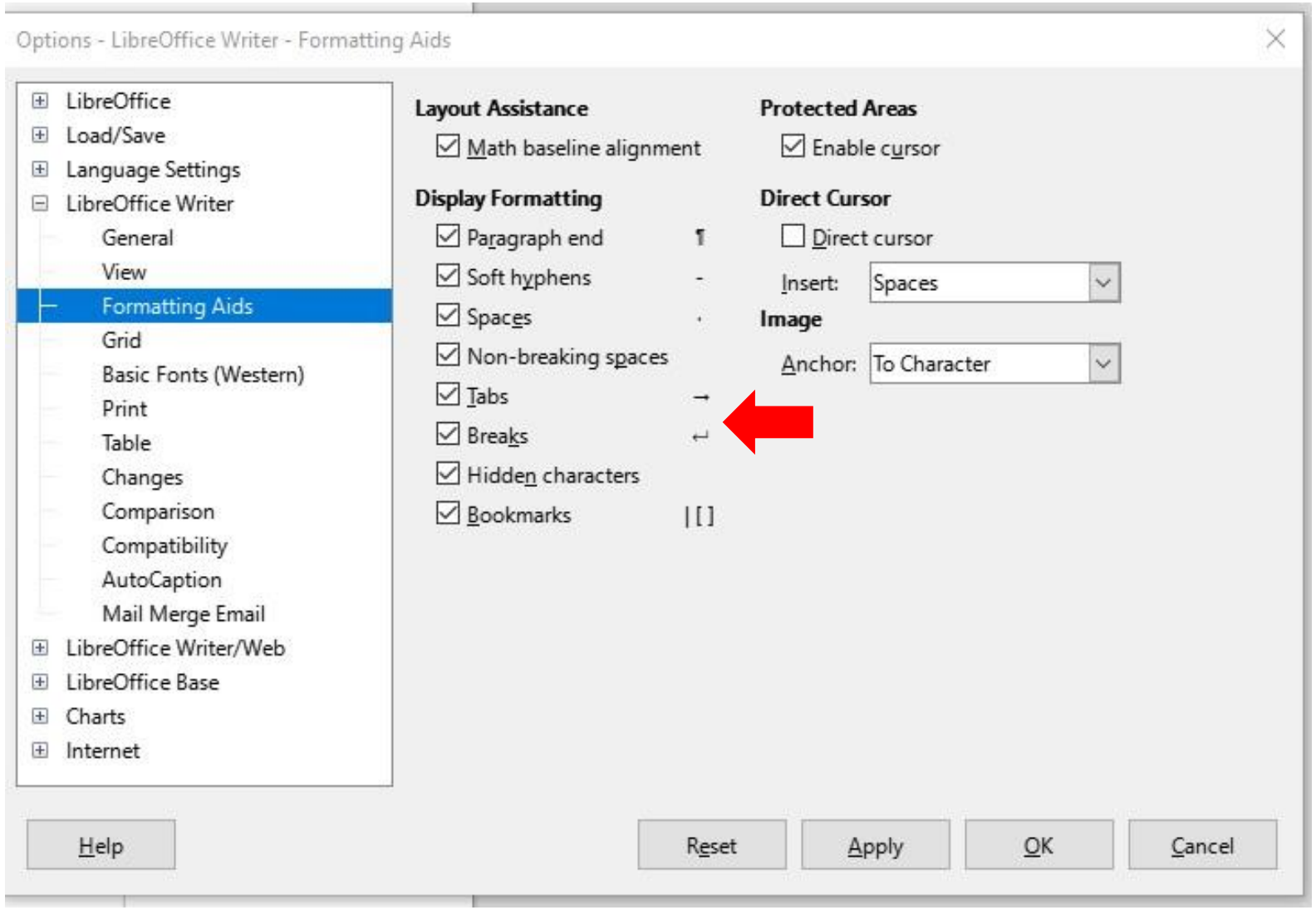
To edit the RTF file, open it in a word processing program such as LibreOffice Writer or Word.

This makes detecting and fixing problems with the scan output easier.

It also prepares the document for subsequent editing and formatting.

To illustrate and fix the problem, it is helpful to see the formatting marks used inserted in the RDF text. To view the marks in LibreOffice Writer:
Tools->Options->LibreOffice Writer->Formatting Aids – the formatting Aids window opens

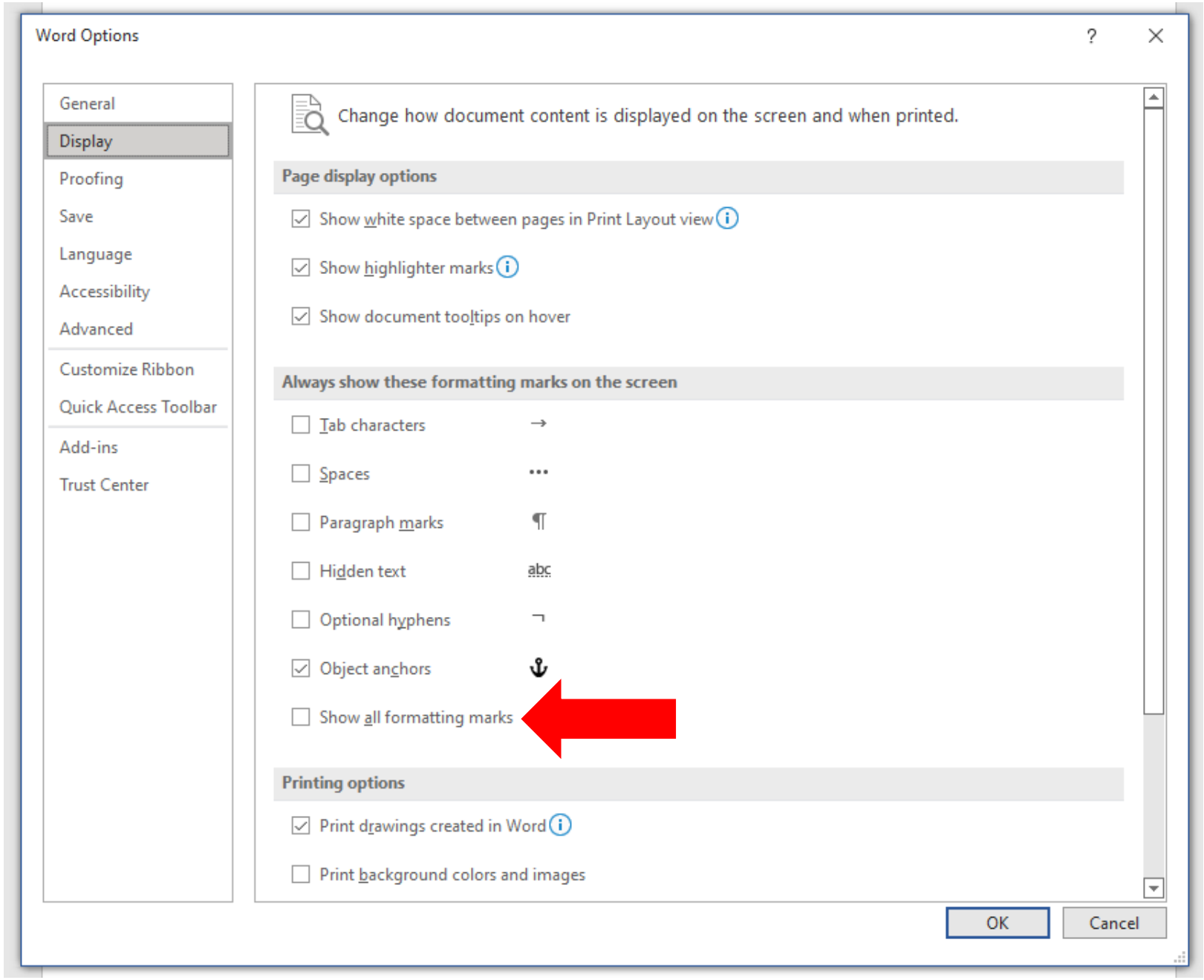
Select Paragraph end, spaces and breaks.



Displaying formatting symbols in Microsoft Word

File->Options The “Word Options” page opens. On the Word Options page select “Display” at the top left

Select “Show all formatting marks”



Here is a short RTF Scan displaying formatting marks with a change in margins. The embedded line breaks creating formatting havoc.

Using·a·program·called·Optical·Character·Recognition,·
which·is·↵
abbreviated·as·OCR,·you·can·scan·a·document·and·
create·a·text·file·↵
rather·than·an·image·file.·¶

This·capability·enables·you·to·recover·a·text·file·from·an·
image·file·that·↵
can·be·edited,·corrected·and·reformatted·to·meet·your·
needs.·¶
¶

I can fix this by deleting the line breaks (left pointing arrows) which can be deleted like any other character. You may need to add a second line break to define the end of a paragraph.

The second problem with the OCR is scanning accuracy.

The program may have problems distinguishing between :

- An upper and lower case “O” and zero.
- Upper case “I” lower case “L” and the number 1.
- A number “5” and an upper case “S”

If there are marks and damage on the original copy, they may be interpreted by the OCR software as additional letters and numbers.

Human readers can use context to resolve character uncertainty. This OCR program does a best guess without any other consideration.

Here is a example of a scan of a poor quality typed document resulting in OCR Errors

Original Document

Biography of
JACOB CLOWARD
Utah Pioneer of 1851

by

Madoline Cloward Dixon
Great-Granddaughter

Jacob Cloward was one of the pioneers who sacrificed his property and his health to join the Mormons in their trek to Utah. However, he did not live to realize the joys of Zion , for he died only a few weeks after reaching, Provo,Utah, the fall of 1851.

Scan Results with massive errors

Manual Column Break

Biography ·of ·↵
JACOB ·CLOW1\ ·IUJ·↵
Iltah ·Pioneer ·of ·ln51 ·¶

oy·¶

Hadoline ·Clovard ·Dixon ·¶
Great-Granddaughter ·¶

Jacob ·Clovard ·vas ·one ·of ·th~ ·~loneers ·vho ·sacrificed ·his ·property ·↵
~nd ·}lis ·h"alth ·to ·join ·the ·Hormons ·in ·their ·trek ·to ·utah. ·Hovever, ·↵
he ·did ·not ·live ·to·r·e ·a ·Li·z·e ·the ·joys ·of ·~don ·, ·for ·he ·died ·only ·a ·fev ·↵
weeks ·after ·~ei\ching, ·Provo,Utah, ·the ·fall ·of ·lR51. ·¶

If you read an RTF file with scan errors into a word processor, the scanning errors will usually be highlighted as spelling errors which makes them easy to find and fix.

If there aren't too many OCR scanning errors, the file may be recoverable. As the frequency of errors increases, you have to make a decision as to whether it is easier to retype the document from scratch or to fix the RTF errors.