

Microfilm in the Age of Digital Imaging

By James Cobban

We exist in a transition period where many archival records are primarily available only in the form of reels of microfilm. Although the holders of many of these microfilm holdings are generally copying them to digital format as fast as their budgets will allow, microfilm will continue to be a key research medium for many years to come.

For researchers there are a number of major problems with microfilm:

1. Microfilm may generally only be viewed in a library or archives reading room. This introduces two sub-problems: the user must travel to and from the microfilm, and the user must bring all relevant research notes and a recording tool (lap-top or paper and pencil) to the microfilm. How often have we all returned home with the fruits of our research, only to find out that we failed to copy a crucial piece of information from the images that were available to us?
2. Microfilm readers are not as ergonomic as our home computers. We have each tuned the physical arrangements of our home or office workstations to our particular comfort. We have positioned the screen to be at the right elevation and tilt to maximize the physical comfort of our necks, and the character of our eyesight. We have positioned the keyboard to the best height and wrist support. We have our favorite chair. If you have not, shame on you; the hours that you spend in front of your computer should be as comfortable as possible. I have yet to encounter a microfilm reader that provided any option for adjusting the position of the display surface. I have yet to encounter a microfilm reader with the zoom lens capability to cover the entire range of documents I study. Meanwhile I am forced to use a lap-top, with its likewise un-adjustable screen and fiddly constrained keyboard, to take notes. Meanwhile through the entire exercise I must keep swiveling my head back and forth between the display of the microfilm reader and the display of the laptop!
3. If you are using a borrowed microfilm at a local library, to minimize your travel time, particularly with respect to archives in other cities, your borrowing period always expires before you have copied down all of the information that you could possibly take advantage of.

These problems largely go away if instead of working directly with the microfilm at the library or archives you scan the images onto a portable medium, and take them home with you.

Instead of lugging a lap-top, which weighs several kilograms, to and from the microfilm, take a Universal Serial Bus (USB) memory stick, also typically known as a thumb drive, which costs only a few dollars, weighs only a few grams, and can be slipped into a pocket, hung on a lanyard around your neck, or slipped into the case for your eye-glasses. Most libraries and archives are gradually replacing their legacy microfilm readers with computer controlled microfilm scanners. This is partly because scanners are required for their own in-house digitization efforts, but mostly because it is no longer possible to get replacement parts for conventional microfilm readers.

Doing Your Own Microfilm Scanning

The scanner will provide a choice of scanning one image at a time, or batching multiple images. Normally you will choose the second option because you want to record as much information as possible during your session, and batching multiple images eliminates the need to assign a separate name to each individual image. There are a couple of specific situations where this is desirable.

To batch images select “Scan to Batch” on the scanning software and select an output format of either Tagged Image File Format (TIFF)¹ or Portable Document Format (PDF)². These two formats are both currently owned by Adobe Systems but are widely supported.

TIFF is the most popular native output format for digital scanners as it is loss-less, that is it preserves all of the original information from the scanned image even when compressing the image.

PDF, as its name implies, is not an image format, but rather provides a convenient and popular container for images, typically stored in the lossy Joint Photographic Experts Group (JPEG)³ format. You may be tempted to scan into this format because it is handled by web browsers in a much more effective way. However if you are going to perform image manipulation before publishing your images then the PDF format has no advantages and some disadvantages over the native TIFF format.

If you are copying a multi-page document, such as a census or set of vital records, you have a lot of pages to record, but you are probably not going to need to examine the individual pages until some aspect of your research directs you to them. I typically scan 5 or 10 pages into a composite document whose name identifies the first page number included.

You may also choose to use the Joint Photographic Experts Group (JPEG) file format. This is a highly efficient format for individual images. Since each JPEG file contains a single image, when you “Scan to Batch” multiple output files must be created the scanner program appends a sequence number to the file name you supply.

When using a public microfilm scanner you do not have the option of choosing the lens that is used to magnify the image. Although this will typically be a zoom lens, permitting a range of magnifications, sometimes even the lowest magnification will not permit you to position the entire original page image within the scanning area. For example my local public library has chosen, on the recommendation of the manufacturer of the scanner, a lens which is suitable for viewing 16mm microfilm, and microfilm of newspapers at a useful level of detail. The original size of a single page of a broadsheet newspaper is about 15” (38cm) wide and 24” (60cm) tall. The microfilm image on 35mm film is only 30mm high, so the image must be magnified 20 times to display the text in its original size. If this image were scanned at a low enough magnification that the entire page image could be displayed within the scanning area, the print in the resulting image would be too small to read.

If you cannot fit the entire original image into the scanning area, it is necessary to perform multiple separate scans, and merge them into single images.

1. Use “Scan to Batch” to create a single TIFF or PDF file that contains the multiple images comprising the single original page. Specify to create the file on your USB thumb drive, with a name that will remind you of the contents of the file, for example “NYTimes_1907_05_15_p007.tiff”.
2. The scanning region on many scanners is a rectangular space with either the proportions of an 8½” by 11” or 8½” by 14” space. If you are lucky you have access to a professional model scanner with a 17” by 11” option. You may rotate the image on the display so that the long axis of the image corresponds to the long axis of the scanning area, if this will reduce the number of separate scans you need to do. However you will be able to rotate the scanned image before publishing it, so the actual orientation does not matter much at this point.

¹ See http://en.wikipedia.org/wiki/Tagged_Image_File_Format

² See http://en.wikipedia.org/wiki/Portable_Document_Format

³ See <http://en.wikipedia.org/wiki/JPEG>

Manipulating Images Scanned from Microfilm

Once you bring the raw images home from the library or archives generally some work must be done to permit them to be used comfortably. In the case of commercial images this has already been done for you. Some of the issues that have to be dealt with include:

1. The image files created by the scanner are generally several megabytes in size because the scanner software is designed to preserve as much of the information in the original microfilm image as possible. That makes the images unsuitable for publication on a web-site, and time-consuming to view even on your home computer.
2. The raw images often have structural defects. They are too dark or too light, have too little or too much contrast. Vignetting⁴ makes the top and bottom and corners of the image too dark, while the center is too bright.
3. The raw image is rotated, perhaps because the original was wider than it was high (typical of censuses) and the scanner only supports a vertical scanning area. Note that off-the-shelf web browsers do not provide an option for the user to rotate JPEG images, so you must rotate them before posting them on the web.
4. The raw images are in a compound format, for example TIFF or PDF as discussed earlier, and you wish to stitch the images comprising one original page together into a single JPEG for convenient viewing. The default TIFF viewer in your browser only displays the first image in a TIFF.
5. The built-in image viewer in the browser only shows images in one of two sizes: fit to window or 100%. To ensure that you get as much raw image information as possible you will probably want to scan at a fairly high density, for example 300dpi. For an 8½" x 11" image this is 2550 x 3300 pixels, much larger than the screen size on even the highest end graphics workstations. You should package your images at an effective 100dpi (850 x 1100) or 150dpi (1275 x 1650) resolution.

To address these issues you need to process your scans using photographic editing software. The most popular such software is PhotoShop from Adobe Systems. This product amply justifies its reputation, and is clearly the product of choice for serious applications; however you may find its price tag excessive if all you want to do is perform a little digital darkroom magic on your microfilm images.

An alternative is the GNU Image Manipulation Program (GIMP). This is a completely free product that can be downloaded from GIMP.org. GIMP is available for both Microsoft Windows and Mac OS-X, as well as on all flavors of Linux (which is not true of Adobe PhotoShop except by running it under a Microsoft Windows Emulator). GIMP is an extremely powerful tool because it is designed by image processing professionals to be capable of performing almost any image modification conceivable. It therefore has a steeper learning curve than most other image manipulation products. In the remainder of this section I will discuss how to use GIMP to address the specific issues listed at the beginning of this section.

Start up GIMP in the usual way for your operating system. By default the installer will have created an icon on your desk top. Also if you right-click with your mouse on an image file you will normally be

⁴ Vignetting is a character of all images created using a lens: the edges, and particularly the corners of the image are darker than the center. This is completely unavoidable in any camera system. It arises primarily because the edges and corners of the image are farther away from the lens than the center and are at an angle. The amount of light that falls on each area at the edge and corners is therefore reduced by the inverse square law of distance, and by the cosine of the angle between the image and the ray of light from the lens (which is equal to the sine of the angle from the center of the image to the point on the edge, as viewed from the lens).

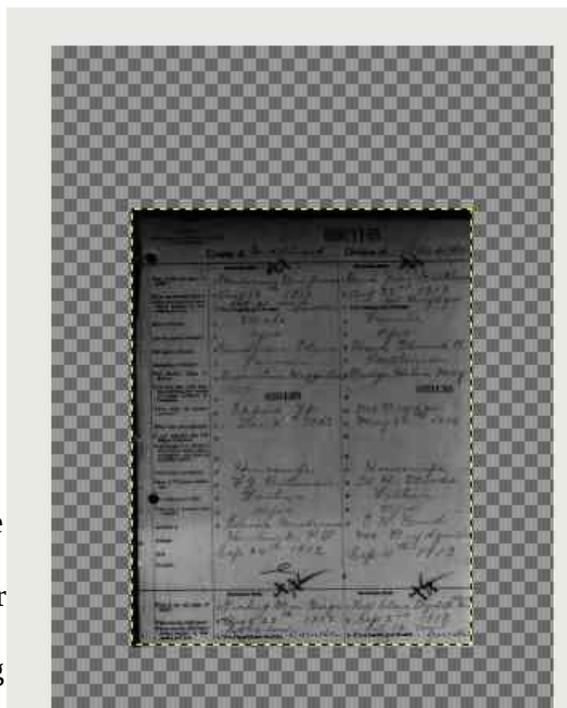
given the option to open it in GIMP.

Open the raw scanned file. If you scanned to a TIFF or PDF you are presented with a dialog that shows thumb-nail images of the component scans, so that you can specify which images you wish to manipulate. For example if you used “Scan to Batch” simply to save having to name each individual page image, then you can now select an individual scan for processing. If you are going to be combining multiple raw scans into a single image to publish then you open the images as “layers”. The alternative is to open each separate image in a separate window. If you used “Scan to Batch” to digitize an image that was too large to fit within the scanning area with the lowest magnification of the lens on the scanner, then click on “Select All”.

Stitching Together Multiple Scans into a Single Image

The multiple scanned images of the TIFF or PDF that you opened are stacked one on top of another in the main window. Note that there is a second “Toolbox” window open. You will need this window later, but you can grab a corner and shrink it so it does not cover so much of your display.

The final resulting image will naturally be larger than the component images, so the first thing you need to do is increase the size of the drawing “Canvas”. With your mouse click on “Image” and “Canvas Size”. The keyboard shortcut for this is Alt-I then V. It is not critical that the dimensions you set here are perfect, you will correct them trivially when you are finalizing the image. Note that the width and height options are initially linked together, as shown by the unbroken chain link between the two values. Just set the width to some significantly larger value, based upon the number of scans it took you to cover the width of the original image. When you hit enter to close this dialog, the stack of images will now be sitting on a larger mottled background.



Now you want to spread out the images, much as you would spread out a stack of photographs of a panorama on a desk.

Reduce the magnification level, either using “View” and “Zoom” (Alt-V Z) or by using the displayed value at the bottom of the window, until the entire canvas is visible in the window.

You spread the images out using the “Move” tool. Among other ways you can activate this tool by clicking on the “Move” icon



in the “Toolbox” window, or just by pressing the “M” key. The “Move” tool is very intuitive. You simply position the graphics cursor (arrow) over any visible portion of an image, with your mouse. Press and hold the left mouse button, and then drag

Illustrat



Illustration 2: Rough Arrangement

the image to where you want it in the canvas. Note that “Move” operates on whichever image is **visible** under the graphics cursor, not on the layer that is currently selected (identified by the dashed outline). Therefore the first image you move must be the top of the stack. If this image is already in the spot that you want it to be in when you have finished laying out the images, you will first have to drag it out of the way, so you can see the next image on the stack, and and at the end drag this top image back into its final position before proceeding to the next step.

Once you have all of the images roughly positioned with respect to each other you may find that your canvas is either not wide enough, or tall enough to be able to see everything. This is easily addressed. Just click on “Image” on the main menu, and then “Fit Canvas to Layers” (Alt-I I). This will ensure that the canvas is just big enough that all of the images are completely visible, with no wasted space around the edges.

Now zoom in to 100% in order to make the fine adjustments in the relative positions of the layers.

Since it does not matter where we start, lets assume that the top-most layer is already in the right position, and adjust the position of all of the other layers relative to it, before stitching them together.

To facilitate manipulating the layers, we open the “Layers” dialog. This dialog displays thumb-nails of the images in each layer, permitting us to select a specific layer using the mouse, to add some special effects, and to be able to rearrange the stacking position of the currently selected layer by clicking on up and down arrows. The long formal way to get to this dialog is to select “Windows” on the main menu, then “Dockable Dialogs”, and then “Layers”. However when we follow this path we note that there is a keyboard shortcut: Ctl-L. Leave this dialog up for the rest of your session, just moving it around with the mouse if it obscures a key part of the image.

For each layer below the top layer (the one we are adjusting with respect to) we perform the following steps:

1. Select the second layer in the Layers dialog with the mouse.
2. Click on the up arrow at the bottom of the Layers dialog to raise this layer to the top of the stack. In the main window the image in the second layer now covers up part of the original top layer.
3. There will probably be a distinct difference in the darkness of the edge of the new layer against the background layer. This is caused by vignetting in the scanner. This edge would be annoying if left in the final image. Secondly the best index points to use to align the new layer against the background may not be just at the edge of the new layer. Moreover it may be desirable to position the line at which the underlying image is overlaid by the new layer at a less critical position in the final image. In particular most government records, such as censuses and vital statistics, as well as newspapers, are laid out on a rectangular grid, often separated by horizontal and vertical lines. These make excellent reference points for lining up portions of the image while at the same time containing little relevant information. We want to make portions of the new layer more or less transparent so that the layer beneath it shows through. The first step in this process is to add a mask onto the top layer. This can be most easily done by positioning the graphics cursor over the thumb-nail image of the top layer, clicking on the right button of the mouse, and then selecting “Add Layer Mask”. This pops up a dialog. Accept the default of a full opacity mask by pressing Enter.
4. Now we want to modify this mask to make the top layer selectively transparent. In the Layers dialog, you will see that a

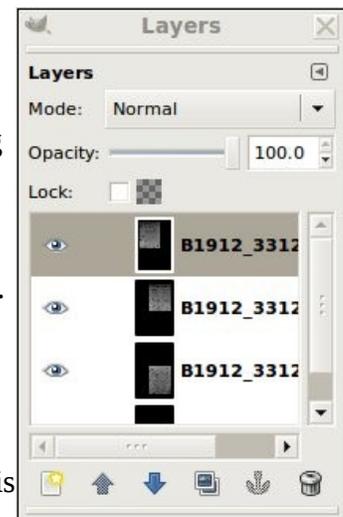


Illustration 3: Layers Dialog



Illustration 4: Layer with Mask

thumb-nail of the layer mask has appeared next to the thumb-nail of the top layer that we are manipulating. Click on this thumb-nail so that subsequent operations apply to the mask, and not to the image portion of the layer.

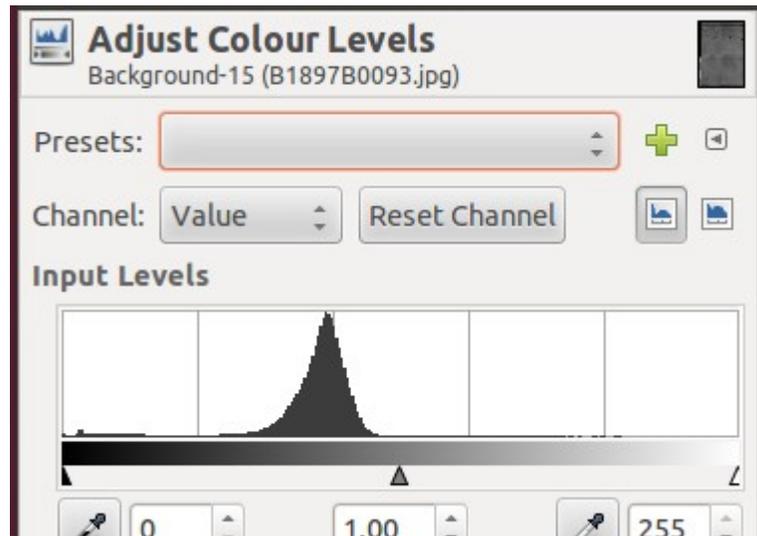
5. Over on the “Toolbox” dialog click on the  “Gradient” icon. For the layer mask black corresponds to fully transparent, and white to fully opaque. This tool can therefore be used to blend gradually from seeing the background image to the foreground image. For example if you want to blend along a vertical boundary between the two images, you will use the mouse to drag horizontally from a point where you want the background to show through, until you get to a point where you want the foreground to completely hide the background. Before releasing the mouse hold down the Ctrl key. This guarantees that your selection is exactly vertical or horizontal. I find that a 30 or 40 pixel blend is quite adequate to hide the boundary while giving enough contrast to facilitate aligning the foreground on top of the background. If the source images have to be arranged in more than one row to tile the result, then the images in the second and later rows (with the exception of the first image in each row) need to be blended both with the previous row and with the previous portion of the current row, which requires blending on two edges of the foreground image. To blend on more than one edge, after selecting the “Gradient” tool from the “Toolbox” dialog make sure that the Blend Mode, which appears in the bottom half of the “Toolbox” dialog when you select the “Gradient” tool, is set to “Multiply” instead of “Normal”. You can leave this option set from now on, because it only means something when you draw more than one gradient on the same mask.
6. You will observe that the background now shows through the foreground layer where you have set the mask. Activate the Move tool by pressing the letter “M” or clicking on the  tool in the Toolbox, click on the thumb-nail of the top image in the Layers dialog (failing to do this will cause the Move tool to act on the mask instead of the image, grab an opaque portion of the top layer with the mouse and drag it until the two images are aligned to your satisfaction. Note that the up, down, left, and right arrows on the keyboard can be used to fine tune the alignment.
7. Select “Layer” on the main menu, and then “Merge Down”. The foreground image is merged with the background.
8. Repeat from step 1 for the remaining layers of the image.

You now have created a single image from the components. The Layers dialog shows a single thumb-nail of the composite image.

Brightness and Contrast Adjustment

Depending upon the quality of the original microfilm the image may be too dark or too light, and the contrast between the background and the text may be inadequate.

“Colors” on the main menu provides a number of tools. For archival documents, which conceptually consist of black ink on white paper, the most effective tool in this menu is “Levels”. This tool displays a histogram of the number of pixels in the image which have a particular brightness. For a typical document this dialog looks like:



For an archival document most of the pixels are going to correspond to the background of the document and relatively few to the “black ink” of the handwriting. So in the histogram above the big peak is the background, and the much lower cluster near the left-hand, black, end of the graph is the “ink”. Grab the white triangle below the histogram and drag it toward the left. All pixels with values brighter than the the position of the white triangle are displayed as white. Similarly all pixels left of the black triangle are displayed as black. The contrast of intervening values is stretched accordingly. So you want to drag the white triangle until it is left of most of the big peak in the histogram, and the black triangle until it is right of most of the low peak. This will enhance the contrast so that the handwriting is easy to read. The appearance of the image in the main window alters as you drag the triangles.

When you are satisfied click on “OK”.

Trimming to Size

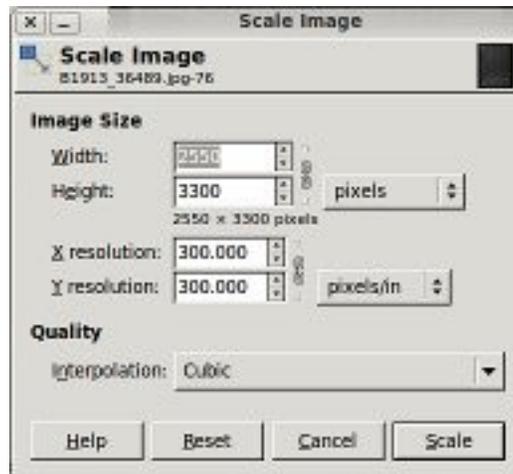
Having finished the fine tuning of the image you may find that the canvas leaves space, or omits portions of the image. Click on “Image” on the main menu, and then “Fit Canvas to Layers”. This will ensure that the canvas is just big enough that the whole image is completely visible, with no wasted space around the edges.

In scanning the microfilm you may have scanned portions that did not contain image. You therefore may wish to use the rectangular Crop tool. This is available on the “Toolbox” dialog under the icon of an X-Acto knife . It is also available under the “Tools” menu by selecting “Transform Tools” and then “Crop”. However you will soon get used to the keyboard shortcut Shift-C. You then use the mouse to draw a rectangle around the portions of the image that you wish to retain. Pressing Enter then discards the outside.

Scaling

To ensure that you get as much raw image information as possible you will probably want to scan at a fairly high density, for example 300dpi. For an 8½” x 11” image this is 2550 x 3300 pixels, much larger than the screen size on even the highest end graphics workstations. You should package your images at an effective 100dpi (850 x 1100) or 150dpi (1275 x 1650) resolution. The actual pixel dimensions of your image are displayed in the title of the image editing window.

To adjust the pixel dimensions of your image select Image → Scale Image (Alt-I, S). This displays the Scale Image dialog:



Since you do not want to distort the image make sure that the chains linking the width and the height are unbroken. Change the pixel width of the image to either half (1275 pixels in this case) or 1/3rd (850 pixels). You can set it to anything, but choosing an even fraction will minimize image artifacts.

Exporting to JPEG

Images that are put up on the Web are almost always in JPEG format. This format is handled natively by all web browsers, and generally provides the smallest file size for a given image. It achieves this by using a variety of techniques that permit discarding raw image information while preserving the appearance. Note that even if you scanned originally to the JPEG format, you may want to process the image through the image editor in order to reduce the file size.

To export an image to JPEG, just use “File” “SaveAs” (shift-Ctl-S), and specify a file suffix of “.jpeg” or “.jpg”. A dialog pops up warning you that the JPEG format cannot represent all of the information in the image you have created, because of limitations of the JPEG format itself.

Finally a dialog is presented which gives you control of the “Quality” of the image. This instructs the image editor as to how aggressive it can be at compressing the image to reduce the file size. You will need to play around with this to find out how large the files are at a particular quality setting, and how much image quality is sacrificed.

Start Over with the Next Image

Do not close the current image, as this will discard a lot of the options you have set along the way, in particular the canvas size and the JPEG quality. Just delete the current layer (Alt-L, D) before opening the next input file as layers.