Digitisation of the John Thomson collection: technical report

Overview

In the summer of 2007 a strategic decision was taken by The Wellcome Library, as part of an ongoing project based initiative, to re-digitise the John Thomson glass-plate negative collection held in The Wellcome archives. Working alongside the curator, photographic department and the preservation and conservation department, we began to investigate how best to create accurate digital renditions of the items without further degradation to the originals.

We began our research by discussing with colleagues, and other cultural heritage institutions, their experiences both past and present in the digitisation and/or photography of glass plate negatives; and what recommendations they might make. The team then accessed numerous sites on the World Wide Web only to find we had more questions than answers. We then turned our attention to what equipment was required, what equipment we already possessed, and if it was fit for purpose.

What follows is a detailed report on how the project was approached, challenges encountered and the decisions made to avoid a paradigm shift.

Issues and Challenges

In the summer of 2007, the Photographic Department within The Wellcome Library was requested to develop a strategy for the digitisation of the John Thomson glass-plate negative collection held in our archives. Taking a heuristic approach, we were able to overcome various challenges; there were handling issues, conservation issues, file format issues as well as file size issues. What began as a relatively mundane task transformed itself into something more challenging than we first expected.

Handling was the first obstacle to overcome. Because of the nature of the 'wet-collodion' process Thomson used, as well as age, much of the emulsion has become badly scratched or is peeling. Also, the glass Thomson acquired to make his plates varied greatly in thickness and many have become quite fragile; in fact several are now either cracked or broken. Working closely with the curator and the preservation and conservation staff we determined which images were capable of being handled and digitised. Also, Thomson personally made many annotations directly onto the plates either by using tape, writing directly onto areas of the originals or using a scribe to etch into the rebate of the emulsion.

How to capture the images then became our next challenge. Many hours were spent in discussions, research and testing to determine how to achieve the optimum digital reproduction without damage to the original glass-plate. We decided, while imaging the originals as negative transparencies, we should also digitise the entire plate as positive

reflective images to capture (at 25 megabytes) Thomson's annotations. Please refer to Figure's 1, 2, and 3 below for several examples of Thomson's annotations.

Fig.1(Handwritten onto image)

Fig.2(Handwritten onto tape)

Fig. 3 (Scribed into emulsion), with close-up

After much consideration, it was determined that the use of flatbed scanning technology would be the least destructive, while providing the best overall rendition of the original.

File types and sizes are difficult to generalise due to various individual perceptions and the potential costs incurred. Also, as technology progresses, future data migration was a serious issue to consider?

Researching and Developing a Paradigm

How best to proceed, along with what has been performed by contemporaries, was our first objective. While digitisation/ photography of glass-plate negatives is not new, we found there was very little current, explicit or consistent mutual information as to a specific approach regarding a digitisation project, such as ours. We contacted a number of departments at various institutions worldwide but discovered they, as we, could be somewhat vague or biased in this area. Ultimately, after many discussions, we came to the

conclusion that our approach might be construed by some as being, while not perhaps unique, somewhat unorthodox.

Digitisation Equipment

We performed a number of tests and determined the Kodak IQsmart 3 flatbed scanner with a high 10,000 dpi interpolated (5,500 dpi optical) scan resolution using Oxygen Scan version 2.6.1 imaging software was optimum. With a maximum transmission scan area of 45.5 cm x 30.5 cm and the relatively fast scan times, the IQsmart 3 scanner provided us with a safe and effective means for digitisation.

Because the scanner has two glass platens, one top, one bottom (scan bed) we found elevating the scanner hood by approximately 5 mm relieved any undesired pressure upon the original. Also, due to static electricity, glass-upon-glass adhesion, etcetera, this alleviated any chance of the glass-plate being inadvertently lifted and dropped during mounting and dismounting of the original. This was achieved by placing 4 felt (5 mm) pads around the four corners of the scanner hood.

Conservation and Handling

As noted previously, careful handling of the original was paramount. While cotton conservation gloves can be somewhat cumbersome and potentially snag on rough or jagged edges such as glass, some have suggested not wearing gloves while handling glass-plate negatives. We, on the other hand, would strongly advocate the use of cotton conservation gloves as the potential hazards of not using them, such as transfer of oils/acids, cuts, etcetera, far outweigh the drawback.

Because of the fragility of the originals, the decision was taken prior to the commencement of the project that only one box (containing between 6 and 22 glass-plates) would be removed from the iconographic achieve at a time; thus reducing the chance of damage. Each group of originals are housed in separate conservation boxes using rubber conservation supports and acid-free paper sleeves.

Cleaning of the scanner glass should only be undertaken using a clean, dry, non-abrasive cloth. Compressed air can be used for the removal of dust from some types of photographic media, but should never be used on or near original glass-plates as this may cause flaking of the emulsion. No solvents of any type should be used to clean either the scanner glass nor the original glass-plates as they can, and in all likelihood will, cause degradation of the emulsion, particularly if the emulsion has already become unstable.

File Sizes, File Formats and Resolutions

File sizes, file formats and resolutions are interesting areas to research as perceptions vary widely. Some advocate smaller file sizes and resolutions, something between 50 Mb and 80 Mb at 300 dpi, while others suggest 100 Mb to 120 Mb at 600 dpi. Tagged Image File Format (Tiff) appears to be the universally accepted file format although later formats such as PNG (1995) and JPEG 2000 (2000) have also been suggested.

After a number of digital imaging tests, at various file sizes and resolutions, we determined 550 Mb at 1000 dpi was optimum for digitisation and storage both now and for any future data/format migration. Because of the unique nature of the Thomson negatives, the ‘wet-collodion’ process he used, detail contained in each image and their varying degrees of

degradation it was our desire to ensure the collection might never need handling or digitisation in the future. For more in-depth information regarding the history of John Thomson and the collection, please visit the collection webpage.

While we are cognisant of the fact these files sizes and resolutions appear, at first glance, to be excessive, a brief overview of our server system might be beneficial. The Wellcome Trust incorporates the Enterprise Pillar Data Systems server array, presently using between 35 and 40 terabytes. The server is ghosted or backed-up to 2 additional offsite servers of the same capacity, or more, but is constantly being upgraded. The present configuration considers that for every gigabyte used, the figure should be multiplied by four. That is to say 1 Gb x 4 = 4 Gb used.

It would be imprudent to suggest file sizes and resolutions such as these be a benchmark for all digitisation projects however, in this instance and taking into account the nature of the subject matter and the detail contained, these were the figures for which we arrived.

The file format was our next consideration. While there was rarely any genuine doubt that the Tagged Image File Format (Tiff) would ultimately be our preferred means of providing images for publication, (JPEG has also been incorporated for public access) the initial digital capture format was different. As a photographic department and with the diverse subject matter both historic and contemporary we are traditionally accustomed to, the RAW file format is our preferred means of digital capture of originals. Because RAW varies from tens, if not hundreds of file types calling themselves RAW, the term is actually quite generic. But, because the RAW format effectively contains all of the information the sensor or CCD (charged coupled device) array detects, we could then manipulate the digital information more accurately.

In the instance of the Thomson collection and, considering the decision to use the Kodak IQsmart 3 flatbed scanner, we opted for the DT (digital transparency) file format developed by Kodak/Creo. Within the Oxygen scanning software there resides an option to capture files in the DT file format. The DT file format is a 16-bit raw data TIFF and provided us with the flexibility we desired. Further detail will be provided in the digitisation process section.

Digitisation Process

The digitisation process was actually relatively straightforward due to the heuristic approach we undertook. We agreed that no more than four glass-plate negatives would be on the scan bed at any given time; the idea being that if, for any reason, there occurred a catastrophic event (such as the scanner hood dropping) there would be fewer plates potentially damaged; the second failsafe being the 5 mm felt pads mentioned earlier.

We then incorporated rubber conservation mats next to the scanner to which the negatives, when not in their box or sleeve, could be safely transferred. After ensuring the scanner glass was thoroughly clean and dry, we laid each negative upon the scan bed using a woodwind (saxophone) reed, see Figure 4 below. We found the use of these woodwind reeds was ideal due to their flexibility and finely tapered ends. This provided a safe and effective way of mounting and dismounting the originals and were easily manipulated wearing cotton conservation gloves.

After mounting, we then previewed the image using the Kodak Oxygen scan software and set the agreed image capture parameters, as indicated in Figure 5, below.

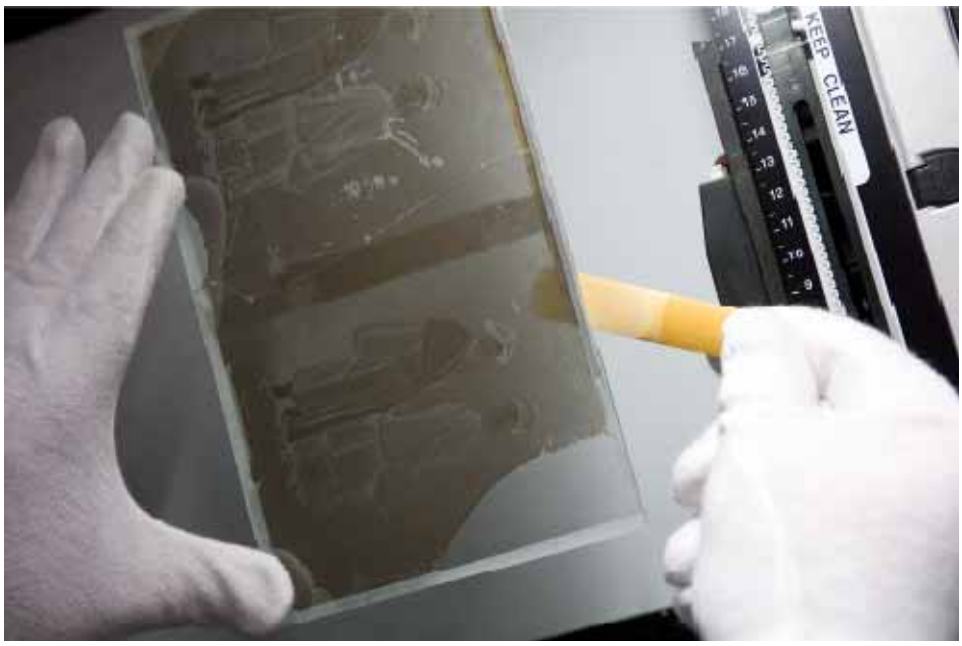
KEEP CLEAN

Fig. 4

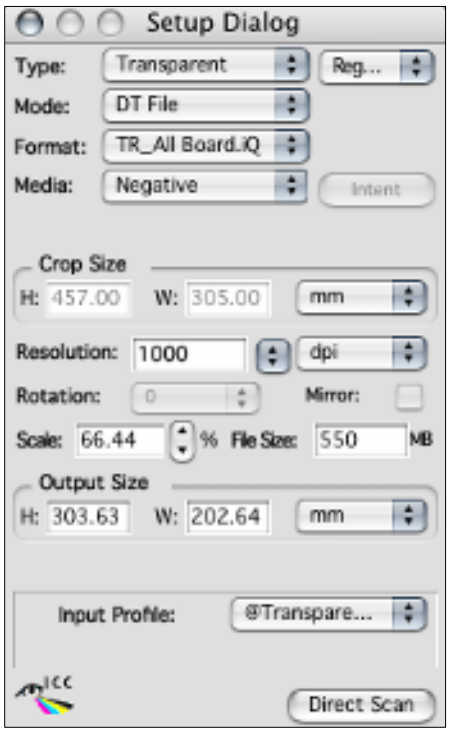
Setup Dialog
Type:
Transparent
Reg...
Mode:
DT File
Format:
TR_All Board.iQ
Media:
Negative
Intent
Crop Size
H: 457.00
W: 305.00
mm
Resolution:
1000
dpi
Rotation:
0
Mirror:
Scale:
66.44
%
File Size:
550
MB
Output Size
H: 303.63
W: 202.64
mm
Input Profile:
@Transpare...
Direct Scan

Fig. 5

It would be appropriate at this time to elaborate further on the file format, resolution and file sizes used during this project. As discussed previously, we agreed upon the Kodak/Creo DT format for initial capture of the original. The DT or digital transparency format is a type of 16-bit raw-data TIFF. The concept behind imaging of the original in the DT file format was to capture as much (RAW) information as possible during the initial stage. While using a 16-bit pixel depth will obviously make file sizes much greater (twice the size of 8 bit) the advantages are increased information, colour range and avoidance of banding in tonal gradations (in this case 16 bit RGB/Black &White when converted in Photoshop CS3) in short, greater image diversity.

In Figure 6 we selected a resolution of 1000 dots per inch. To be assured of a very high quality rendition of the originals, the decision was taken to use 1000 dpi as our optimum resolution. While this increases the file size exponentially, it also provided us with greater reproduction quality. Also, increasing the file size, in this instance 550 Mb, flexibility was enhanced significantly.

Interestingly, the Thomson collection is comprised of several various sized glass-plate negatives but for this narrative we will concentrate on the types of cameras Thomson employed for our collection, full-plate and stereo.

Figure 6 represents a full-plate (approximately 10 in x 8 in) and Figure 7 represents a stereo image (approximately 8 in x 4 in). It might be expected, based upon the physical dimensions of the original, the full-plate image would create a larger file size than the stereo image. On the contrary, the stereo image was roughly twice the file size of the full-plate. Figure 8 and 9 should help to explain this anomaly.

Fig. 6

Fig. 7

Fig. 8 Fig. 9

Capture (DT Format)

During the capture phase, we cropped the full-plate image to just inside the dimensions of the original (see Figure 8) set the file size to 550 Mb and increased the crop area to just slightly beyond the outside of the glass. For a full-plate image at 550 Mb the results are fairly obvious, approximately 570-580 megabytes. However, during capture of the stereo images we cropped the image in half (see Figure 9) set the file size to 550 Mb and then opened the crop box up to capture the entire plate. By increasing the crop of the image, the file size increases by approximately; 2 x 550 Mb = 1.1 Gigabytes.

Conversion to RGB/tiff format

Once the initial capture in the DT format was accomplished, the next step in the process was to convert the raw images to RGB/tiff's. Below are two examples of the DT to tiff conversion dialogue boxes.

Figure 10 shows the setup dialogue alongside the DT image. To proceed with the conversion process, the DT image was first opened in Oxygen Open and a crop area was determined. In this case, a stereo image was opened and cropped to one half of the total image area. This decision was taken because there may potentially be requests for only half of a stereo image as, all too often, the other half is damaged or obscured (note Figure 10). At this stage the desired parameters were set within the setup dialogue box. As discussed, for our project, we determined a resolution of 1000 dpi and a file size of 550 megabytes. The Mode was set to Color RGB 16-bit, Media set to Negative and the Input and Output profiles set to iQsmart3.icc and AdobeRGB 1998 respectively. We found the iQsmart icc profile gave us the optimum input colour profile while AdobeRGB 1998 was our desired standard (generic) output for accurate colour management when moving images between various colour systems.

Also, you may notice we selected no sharpening; this was due to the over-sharpness (by the scanner) of the captured image when sharpness was applied.
Figure 11 shows the same setup dialogue alongside the DT image but note the expanded crop box and file size.

Fig. 10

Fig. 11

Once the parameters had been set, it was a simple matter of selecting the scan option in Oxygen Open for the conversion from DT to 16-bit RGB/tiff as the software runs in a virtual scan (no scanner attached) mode. The same process was used regardless of the type of image (full-plate or stereo) captured. It is worth mentioning here that there was no batch conversion mode available during this process so each image was opened separately and the process repeated.

RGB/Black & White Conversion in Photoshop CS3

The final step in the process was to convert the RGB/tiff to its constituent 16-bit Black and White/RGB image. Our reasoning behind the black and white conversion was that this would have been the most prolific means of printing to a broader audience at the time. The figures below illustrate how we approached this specific project and may not, necessarily, be appropriate for various other glass-plate negative collections held within our achieves.

By opening the RGB/tiff image in Photoshop CS3 and converting the image to black and white using the function Layers, we were able to achieve the flexibility desired for best overall image quality. Conversion here performs a black and white conversion while retaining both RGB colour integrity while not affecting the file size; unlike performing the Image/Mode/Grayscale conversion option, which strips the image to a grayscale component only. Figures 12 and 13 below illustrate this procedure.

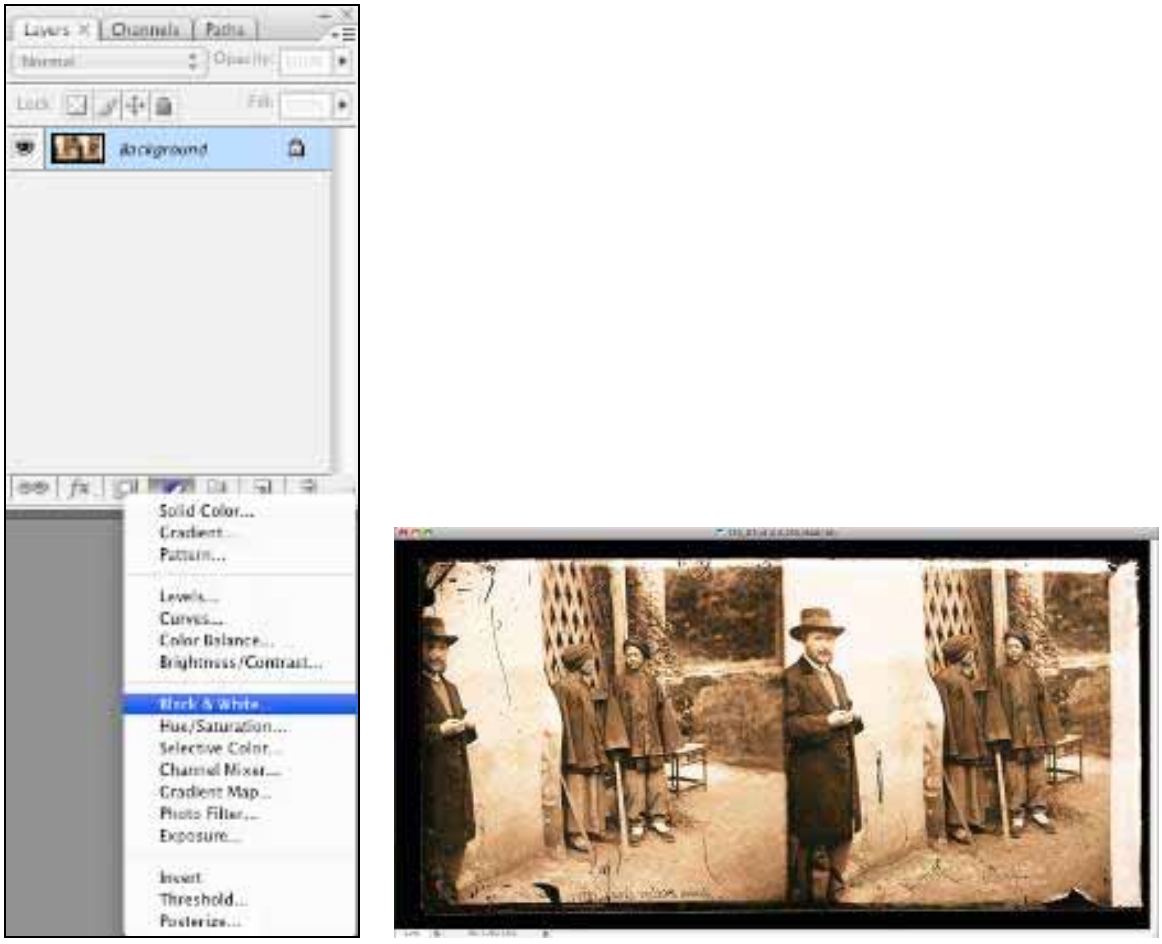

Fig. 12

Fig. 13

The following steps to complete the conversion process could be, and were, quite subjective. Due to the subject matter, process (wet-collodian), exposures and image degradation, pinpointing a specific area to define the overall image was challenging. Referencing the histogram to adjust highlights, mid-tones and shadows as well as avoiding undesired clipping was extremely important. Figure 14 shows the histogram in Levels before and after adjustment to the image above.

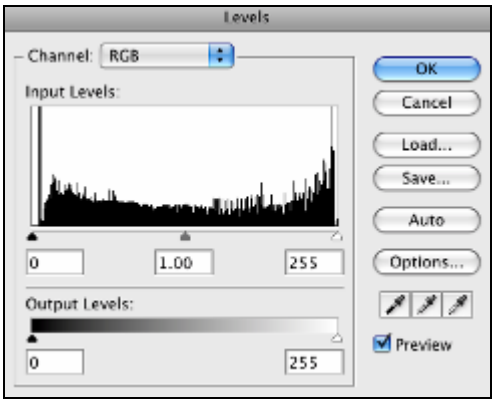

Fig. 14

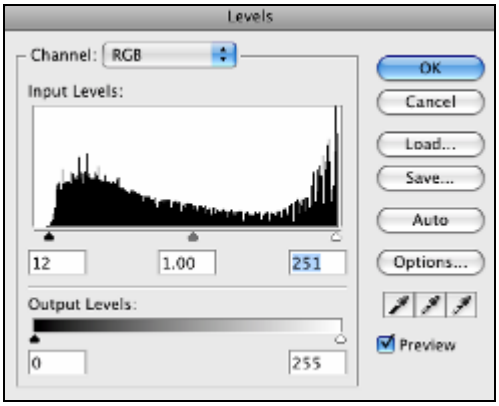

Fig. 15

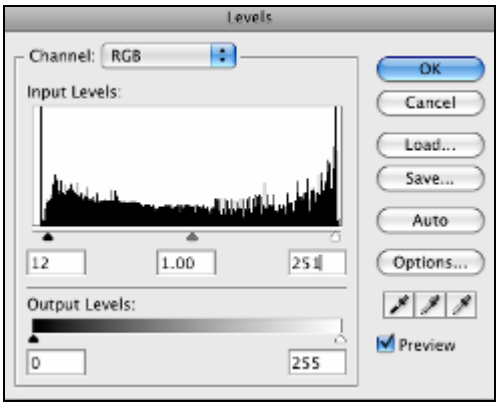

Fig. 16

The histogram in Figure 14 shows a spike in the shadow area indicating the shadow detail is being referenced from the black area (scan bed) around the outside of the image as opposed to the shadow detail within the actual image. This was overcome in Photoshop using the Rectangular Marquee tool, selecting an area within the image, choosing Levels and making a temporary image adjustment (Figure 15). Taking note of the values we believed optimised the overall image we then cancelled Levels, deleted the Marquee box, selected Levels once more and applied the noted values as can be noted in the right histogram in Figure sixteen.

To complete the image adjustment tasks, we found some slight adjustments in Curves was sometimes necessary but was again purely subjective.

Conclusion

Pre-project planning and the heuristic approach undertaken were integral to the overall success of the Thomson Project. The hours spent investigating, imaging and reviewing procedures paid huge dividends once the project was underway. While there are always potentially unanticipated circumstances that may arise, when properly planned and executed, they tend to present few obstacles which cannot be overcome.

We have previously utilised camera configurations such as Kodak DCS Pro Back on a Mamiya camera body using high powered, daylight balanced light boxes (in a copy stand configuration), to capture glass-plate negatives with a great deal of success. The process was fast and accurate but with lower resolutions and file sizes. We have also utilised older flatbed scanner technology such as the Heidelberg Topaz 1 and Topaz 2 operating on Linocolor imaging software with relatively high resolutions and file sizes but with slower imaging times. As an innovative and progressive department, we always welcome any constructive and relevant input.

There are several additional processes such as assimilation and insertion of accurate metadata, ingestion and public access which fall outside of the photographic department's remit and have therefore not been addressed within this report.

Digitisation Project Principals:

Curator: Mr. William Schupbach
Project Manager: Ms. Christy Henshaw
Photographic Manager: Mr. Richard Everett
Conservation Manager: Ms. Stefania Signorello
Imaging Supervisor: Mr. Laurie Auchterlonie
Head of Access and Stewardship: (the late) Ms. Bridget Kinally who's leadership and passion for this, as with all public access programs, cannot be overstated and whom shall be greatly missed.

Further information

The Wellcome Trust is sponsoring the John Thomson collection to be exhibited throughout China beginning in the spring of 2009, requiring VLF (very large format) prints for which the file type, file sizes and resolutions described are ideally suited. For further information about the John Thomson Beijing Exhibit visit Diarmuid Mitchell on the BBC China website:

For more information on the John Thomson collection, please contact the Wellcome Library
183 Euston Road, London, NW1 2BE, library@wellcome.ac.uk.

Thomas Cox, Imaging Technician
December, 2008