

SCANNED IMAGES: How Well Do They Depict the Subtle Features in Handwriting?

Janet Fenner Masson

Forensic Document Examiner, 908 Town & Country Boulevard, Suite 120, Houston, TX 77024

It is increasingly common for companies and government agencies to retain scanned images of documents in lieu of originals. In many instances, the original document is destroyed after imaging. As a result, the best image of a document may be a digital image. This project is not intended to be an inquiry into the parameters document examiners should use in scanning documents for their own records. Rather, it is a study of digital images captured using guidelines like those utilized by companies and agencies in maintaining records. A study was made to determine which features are reliably and accurately depicted in these scanned images and which are not. A comparison of the image quality obtained using various scanning parameters and transmission methods was made. In addition, examinations from originals, from first-generation photocopies, and from scanned images were considered.

Introduction

Documents are increasingly being preserved digitally rather than being stored in original, paper-and-ink form. This is seen in bank records, company records, and court records. While there are known limitations to the examination of photocopies, most document examiners make analyses from photocopies on a regular or at least an occasional basis. However, it is becoming more common for a client to request that an examination be made of an image stored as a digital file, with that image having been captured by some form of routine scanning. The client generally offers to attach the file to an e-mail and provide it to the document examiner in that form. In fact, in some cases this digital image may be the best evidence; that is, this may be the image that is closest to the original.

The question for the document examiner becomes:

If the original document has been lost, destroyed, or is not reasonably available for examination, can a reliable analysis of handwriting and/or a signature be made from an image captured by routine scanning (that is, scanning over which the document examiner has no input or control)?

This study was undertaken to address that basic

question. Other issues considered are:

1. When an image of the document has been made by scanning and/or photocopying, what is the best image for examination purposes? And how should it be submitted?
2. Clients often suggest that they send the "document" as a pdf or tif file, attached to an e-mail. Is this acceptable? Is anything significant lost in this process?
3. In some instances, the image has been received from a source by one recipient and stored; then it is forwarded on to a second recipient. Is anything lost by successive e-mail transmissions?
4. If a first-generation photocopy of the document was made prior to or in addition to scanning of the original, is the examination better made from the first generation photocopy or from the digital file?

Even though digital photographs and scans may be stored using various file storage formats, they are all bitmap graphic types. Essentially, the stored image is composed of tiny bits of information. That is, the images are composed of a grid of pixels, each of which represents a single color in the image. If one of these images is dramatically enlarged using imaging software, the presence of these pixels can be seen. In digital terms, the

word pixel is a contraction of the term “picture element,” and it is the smallest unit of a picture than can be controlled. Typically, the more pixels per inch used for capturing and storing the image, the more accurate the representation of the original will be. When an image is captured in a black-and-white system, each pixel is either black or white. In color image systems, a color is typically represented by three or four component intensities such as red, green, and blue (RGB), or cyan, magenta, yellow, and black (CMYK).

PPI vs. DPI

Both of these abbreviations refer to the resolution. PPI is an abbreviation for pixels per inch; DPI is an abbreviation for dots per inch. It is true that these two terms are often used interchangeably. However, in actual fact, ppi refers to image resolution when an image is saved and viewed on the screen; dpi refers to output resolution or the resolution of the image on the paper print.

File Storage Formats

There are a variety of formats available for storing digital images (i.e., those captured by the process of scanning or photographing with a digital camera). Some file formats are referred to as “lossy” while others are “lossless.” In a format that is “lossy,” such as jpeg, the file storage size is reduced because not all of the image information is actually stored. In “lossless” formats, all image information is stored. When documents are scanned for storage purposes, lossless formats are typically used.

Two commonly used storage formats are tif and pdf. Tagged Image File Format (abbreviated TIFF or with the file extension .tif) is a file format that is under the auspices of Adobe. Portable Document Format (PDF) is a file format created by Adobe Systems to be used for document exchange. A pdf file format image can be stored as a compressed file or as an uncompressed file. In conversations with bank personnel and with Harris County personnel, I learned that these entities scan files as black-and-white images and store files without compressing them. Documents stored by Harris County (Texas) are scanned as black-and-white images at a resolution of 200 pixels per inch, are stored as tif files, and are not compressed.

Methods and Materials

Samples of natural handwriting were prepared

using a variety of writing instruments and on two different kinds of paper. The same two pages of samples were used for all the experimentation.

The papers used were (1) plain white bond paper and (2) yellow ruled paper from a legal pad.

The writing instruments used on each page of sample writing were:

- A Uni-ball Vision Rollerball Pen with blue ink
- A Zebra F301 ballpoint pen with black ink
- A Zebra F301 ballpoint pen with red ink
- A Sanford Sharpie marker with brown ink
- A Sanford Sharpie marker with black ink

Each of these two pages was scanned using a Hewlett Packard Officejet Pro 8500 scanner. They were scanned using various resolutions and stored in two different formats (.pdf and .tif).

Scans were made at the resolutions and using the storage file formats listed in Table 1.

First-generation and second-generation photocopies were also made of each sample in question, using a typical electrostatic photocopier.

Transmission of the images was then undertaken. Because of size limitations in the files, the transmission was only conducted of the black-and-white images. The images were each attached to an e-mail by Sender 1 and sent from that e-mail address to another (referred to here as Recipient 1). Upon receipt by Recipient 1, the images were downloaded and stored.

In the next step, the images received and stored by Recipient 1 were opened and attached to an e-mail that was then transmitted back to Sender 1. Upon receipt, these images were downloaded and stored.

The following comparisons were made:

1. Each stored image (including color, grayscale, and black-and-white images) was compared to the relevant original writing sample page.
2. Each originally-stored black-and-white image was compared to the relevant image that had been transmitted once as an attachment to an e-mail and to that transmitted twice.
3. Each stored image was compared to the relevant first-generation and second-generation photocopies.
4. The first-generation and second-generation photocopies were each compared to the relevant original writing sample page.

Writing characteristics studied were (1) lightly-written connections or upstrokes, (2) stroke directions, (3) initial and terminal strokes in words, (4) diacritical marks and punctuation marks, (5) overall line quality, (6) letter designs, (7) internal proportions of letters, and (8) height and spacing relationships.

SCAN RESOLUTION	B&W/COLOR CAPTURE	FILE STORAGE FORMAT
100 PPI	BLACK AND WHITE	PDF (uncompressed)
100 PPI	BLACK AND WHITE	TIF
100 PPI	GRAYSCALE	PDF (uncompressed)
100 PPI	GRAYSCALE	TIF
100 PPI	COLOR (RGB)	PDF (uncompressed)
100 PPI	COLOR (RGB)	TIF
200 PPI	BLACK AND WHITE	PDF (uncompressed)
200 PPI	BLACK AND WHITE	TIF
200 PPI	GRAYSCALE	PDF (uncompressed)
200 PPI	GRAYSCALE	TIF
200 PPI	COLOR (RGB)	PDF (uncompressed)
200 PPI	COLOR (RGB)	TIF
300 PPI	BLACK AND WHITE	PDF (uncompressed)
300 PPI	BLACK AND WHITE	TIF
300 PPI	GRAYSCALE	PDF (uncompressed)
300 PPI	GRAYSCALE	TIF
300 PPI	COLOR (RGB)	PDF (uncompressed)
300 PPI	COLOR (RGB)	TIF

Table 1 Scan resolutions and file format storage used for samples.

Results

1. Features such as letter designs, internal proportions of letters, and height/spacing relationships were typically accurately depicted, even on scanned images at relatively low resolution. See Figures 1a and 1b on the next page.
2. Black-and-white scans often have deficiencies in their replication of subtle writing features, including lightly-written strokes, the presence of connections between letters, the presence and especially the length of initial and terminal strokes, the size of diacritical marks, and the line quality of the writing overall. The directions of strokes such as crossbars of letters “t” are often not clear on low-resolution black-and-white scanned images. See Figures 2a and 2b on the next page.
3. Features of the paper, such as ruled lines and folds in the paper, are often not fully-depicted in images scanned in black and white. See Figure 3 on page 45.
4. No differences in the replication of handwriting features were found between the images of handwriting stored as tif files and those stored as uncompressed pdf files when the two scans were made at the same settings (resolution and color choice).

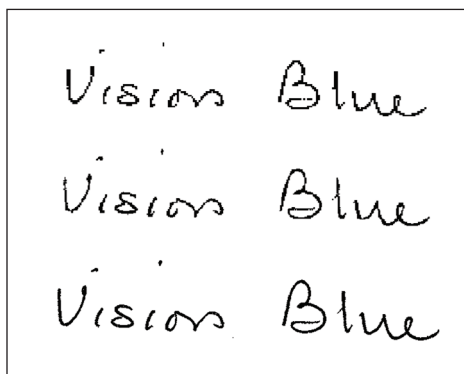


Figure 1a *Comparative scans at different resolutions* The top image is from a black-and-white scan, at 100 ppi; the middle image is from a black-and-white scan, at 200 ppi; the bottom image is from a black-and-white scan, at 300 ppi. Note that the image of writing at 100 ppi was enlarged to correspond to that of the same writing at 300 ppi, etc., but no change was made to the resolution.

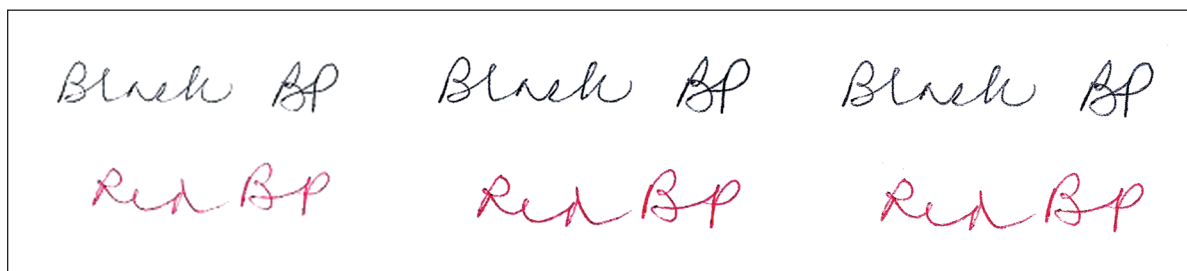


Figure 1b *Comparative scans at different resolutions* The image on the left is from a color scan, at 100 ppi; the middle image is from a color scan, at 200 ppi; the image on the right is from a color scan, at 300 ppi. Note that the image of writing at 100 ppi was enlarged to correspond to that of the same writing at 300 ppi, etc., but no change was made to the resolution.

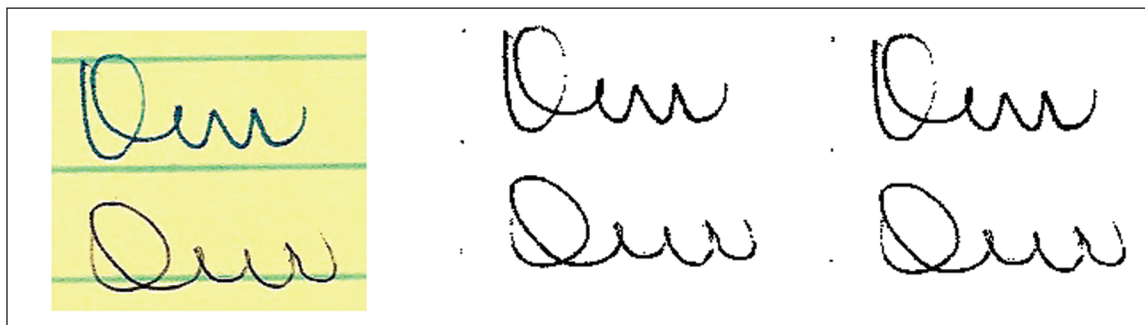


Figure 2a The image on the left is from a scan in full color at 200 ppi, with the image stored in tif format; the middle image is from a scan at 200 ppi in black and white, with the image stored in pdf format; the image on the right is from a scan at 200 ppi in black and white, with the image stored in tif format.

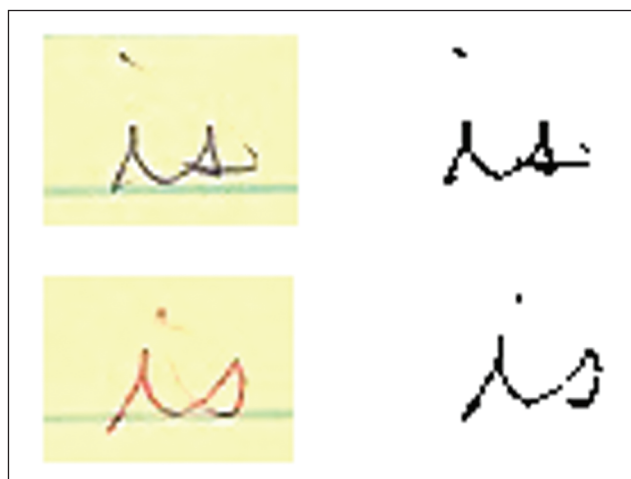


Figure 2b All four images are at a resolution of 100 ppi, with the two on the left being color scans while the two on the right are black-and-white scans.

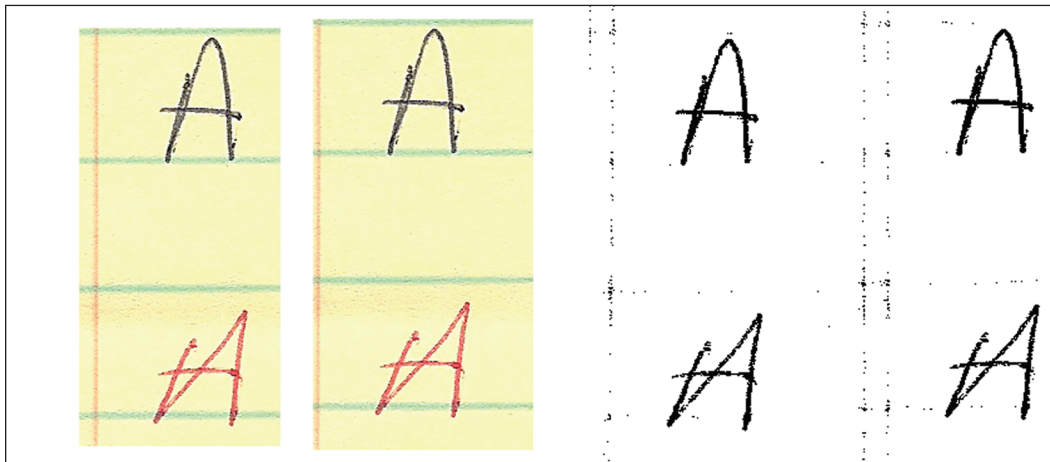


Figure 3 All four images are at a resolution of 300 ppi, with the two on the left being color scans while the two on the right are black-and-white scans. The color scan to the far left is stored in pdf format while the color scan second from the left is stored in tif format; the black-and-white image on the far right is stored in tif format while the black-and-white image second from the right is stored in pdf format. Note the difference between the color and black-and-white scans in the replication of the ruled lines and of the fold in the paper.

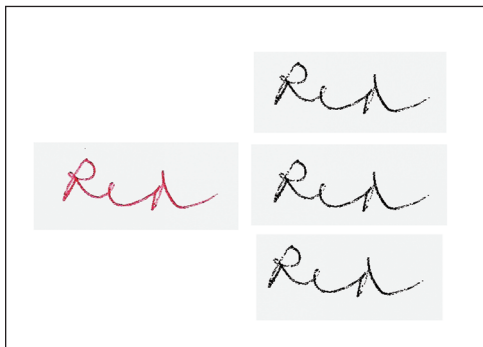


Figure 4 The image on the left is from a color scan at 300 ppi, stored in pdf format. On the right are images of the same handwritten word, from a black-and-white scan at 300 ppi, stored in pdf format. The top right image is directly from the scanned file; the second image is after one transmission as an attachment to e-mail; the third or bottom image is after two transmissions as an attachment to e-mail.

5. No deterioration was found in the scanned tif and pdf files as a result of sending them as attachments to e-mail messages. See Figure 4.
6. The lower the resolution, the worse the image. When scanned at the same resolution, color images are consistently and significantly better than black-and-white images; documents scanned as grayscale images (continuous tone) are better than black-and-white images but not as good as color images. See Figures 5 and 6 on the next page.
7. A first-generation photocopy provides significantly more information about the handwriting than does a low-resolution, black-and-white scanned image. See Figures 7 and 8 on the next page.
8. In a scanned image with low resolution, the image of the writing may appear to be tremulous when, in fact, the writing on the original document is spontaneous and natural.
9. When the scanning resolution is low, espe-

cially when the image is scanned at low resolution in black-and-white mode, the document examiner can often make reliable judgments about the form or design characteristics of the writing but be unable to reliably assess the writing quality and some subtle features such as stroke directions. That is, evidence of non-genuineness based on fundamental differences in letter designs, proportions, and similar writing features is often apparent, but slow writing movement, subtle breaks in the writing line, and other evidence of poor line quality (especially if this evidence is subtle) will often be concealed.

Conclusions and Recommendations

The original document is always the best evidence.

If the original is not available, but a photocopy was made directly from the original, make an attempt to obtain it. It will likely be better than

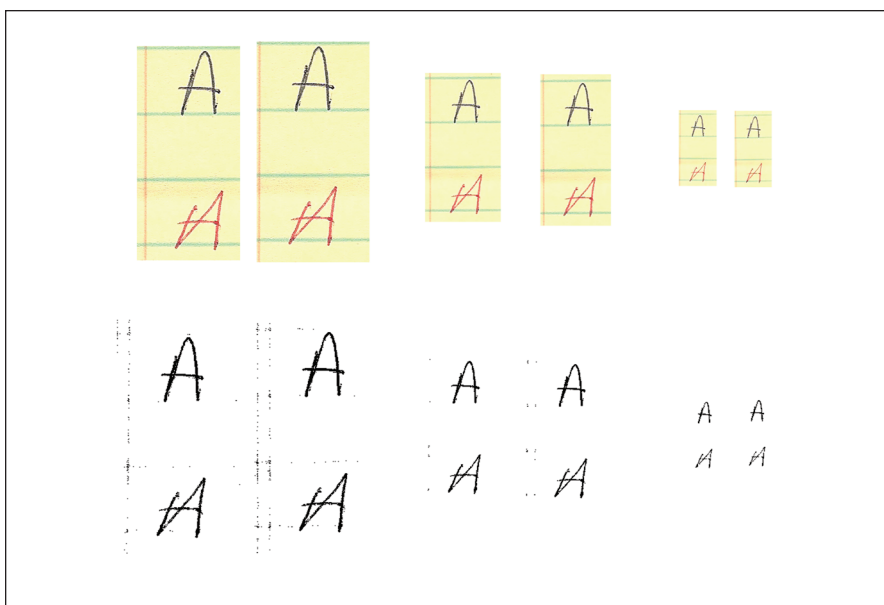


Figure 5 Images of handwriting without any adjustment made to their size. The top set of images are in color, with the two on the left being at 300 ppi, the two middle images are at 200 ppi, and the two on the right are at 100 ppi. The bottom set of images are from black-and-white scans, at 300 ppi (on the left), 200 ppi (in the middle), and 100 ppi (on the right).

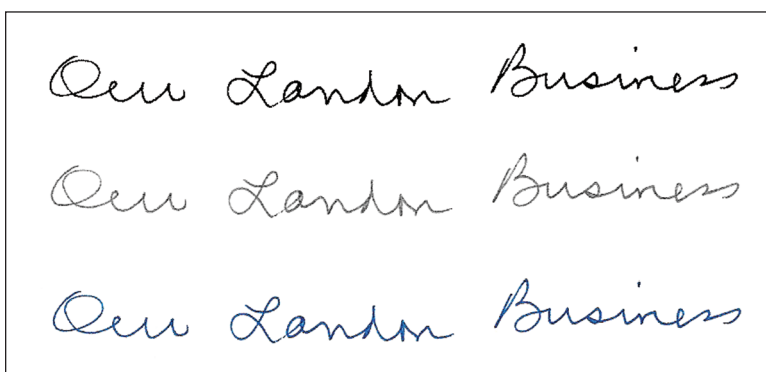


Figure 6 Images of text at the same resolution (200 ppi) but with different color selections. The top image is a black-and-white scan, the middle image is a scan in grayscale, and the bottom image is a color scan. All three images are stored as tif files.

a black-and-white scanned image.

The closer in generation the document is to the original, the better. Receiving the digital image as an e-mail attachment often provides more information than does a print made from that digital image.

While there are clearly limitations to the examination of handwriting on low-resolution scanned images, digital images of documents captured by scanning often provide useful information about the handwriting and cannot be disregarded when the original document has been destroyed. In some instances, scanned images (especially images scanned in color and higher-resolution

scans) are sufficient for a reliable evaluation of the handwriting. It is important for a document examiner to carefully assess the handwriting features clearly depicted in a scanned image as well as to consider what features may be missing or distorted.

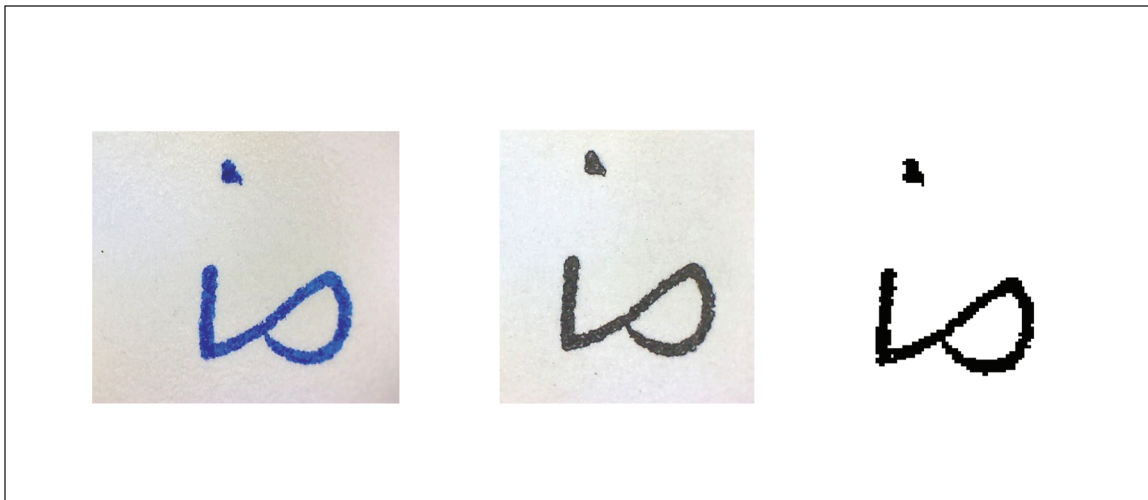


Figure 7 Comparison of three images of the same handwritten word. The image on the left is a digital photograph; in the middle is a first generation photocopy; on the right is an image scanned at 300 ppi in black-and-white mode and stored as a tif file.

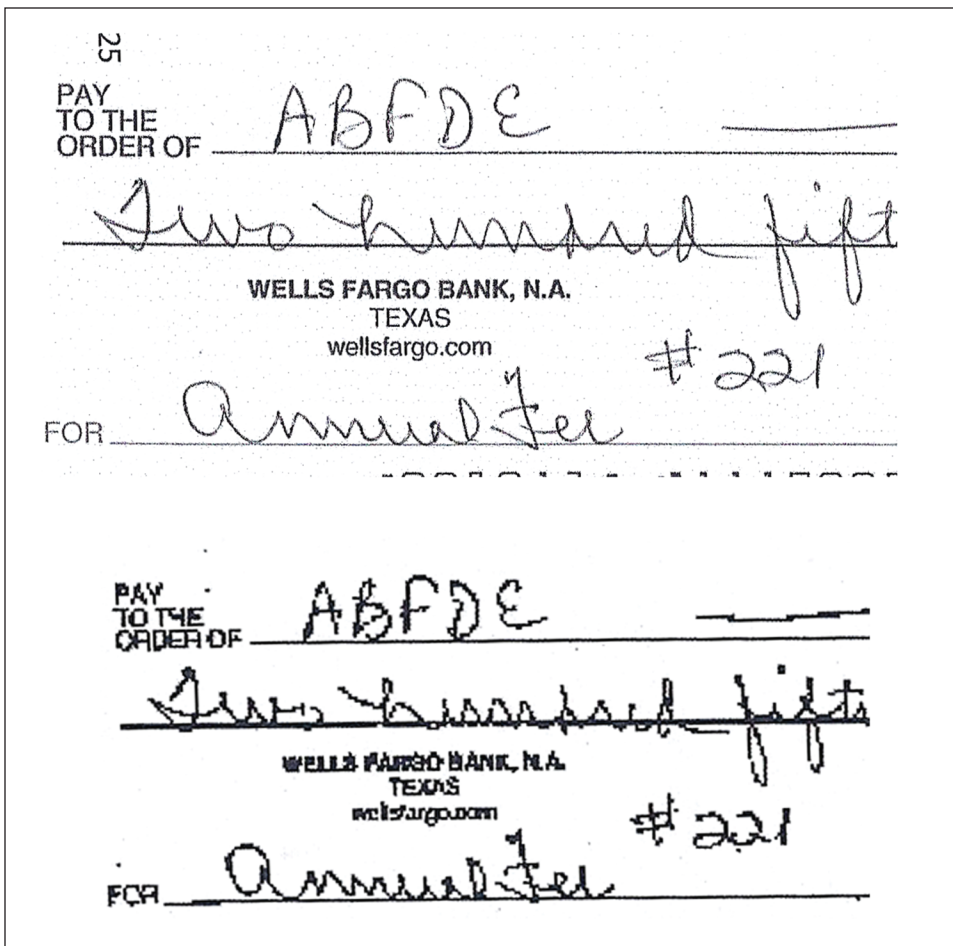


Figure 8 Comparison of a first generation photocopy (top) with an image from a bank (bottom). The image received from the bank was enlarged to approximate the size of the image on the first generation photocopy.