Identifiability of the Flatbed Scanner and Its Products (Graphics Files and Printed Results)

Richard A. Horton, BS, M.Ed.

Virginia Division of Forensic Science, 6600 Northside High School Road, Roanoke, VA 24019

This study addresses the ability to identify the specific original document, scanner or graphics file from which a scanned product (graphics file or printed result) was made. Contaminants in many types of paper are often reproduced by the scanner, allowing a correlation of the product to an original document. Further, Small Computer Systems Interface (SCSI) chain noise (interference) frequently produces a non-repeating series of marks allowing a graphics file to be correlated to a printed result. Photocopier-type defects from the scanner platen and internal components may also allow the correlation of a specific scanner to a printed result.

The opinions and assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Virginia State Laboratory system.

Introduction

The Collaborative Testing Services (CTS) provides semi-annual questioned document problems that many Forensic Document Examiners (FDEs) take on a regular basis. Currently, the CTS provides one handwriting proficiency test and one "other" proficiency test each year. The fall-1997 problem involved a counterfeit gift certificate and required its comparison to a genuine certificate. Participants were primarily being tested on their knowledge of printing processes (although it is likely examiners subjected the certificates to a battery of examinations). A series of small black marks were noted on the counterfeit certificate which could not easily be accounted for. This observation led to a small but interesting study involving the identifiability of the common flatbed scanner and its products, the graphics files and printed results. This paper is an overview of that study.

Overview of Scanning Technology

A scanner is a peripheral device used to digitize artwork, photographs, or other items from hard copy. In a sense, the scanner works like a pair of eyes for the computer. The human eyes see an image and translate it into electrical impulses that travel to, and are interpreted by, the brain. Similarly, a scanner captures images and converts them to digital data that travels to, and is interpreted by, the computer. A typical application requires that a picture be placed onto the scanner platen where it reads the material and converts it to a data file. The scanner divides the image into microscopic rows and columns by measuring how much light (or lack thereof) reflects from each individual intersection of the rows and columns. The intersection is recorded as a dot, or picture element (pixel). After the scanner collects the information from all of the dots, it compiles the result into a digital file on the computer. From there, the file can be manipulated as with any other data file.

There are a wide variety of scanners—flatbed, sheetfed, hand held, barcode, slide, etc.—that work in a number of different ways, but the technology behind them is essentially the same. They can usually scan in black and white, grayscale or color. Their technical capability, or optical resolution, is measured in dots per inch (dpi). The higher the dpi, the more detailed and sharp the on-screen printable image will appear. The reason for this is that the more dots that can be placed in the same area (i.e., square inch), the smoother and more detailed the overall image will look. Typical flatbed scanners range from 300dpi to over 3,200dpi.

Scanners can really stretch the limits of a computer system, both in operating capability and the storing of digital files. Saving scanned images can be extremely memory intensive. The amount of memory and storage necessary is directly related to the size and resolution of the file, the type of file it is saved as, and whether it is scanned as a black and white (1-bit), grayscale (8-bit) or color (24-36-bit) image. In its simplest terms, a grayscale file derived from the scan of a 1" square image at 300x600dpi produces 300x600x8-bits or 1,440,000 bits of information (or 180,000 bytes/ 180 kilobytes/180 kb/.18mb). An 8" x 10" color image (30-bit) may take up over 50mb of space on the computer's hard drive. The higher the resolution, the more space it will consume. For this reason, many files are cropped, scanned and saved in black and white mode to conserve hard drive memory. If a picture is not going to be enlarged for printing, the resolution is also frequently kept low. This is because most printers cannot take advantage of the higher resolution when they print (most printers have a resolution of between 300-1440dpi).

There are scanner and printer drivers and computer graphics software programs that can be used to manipulate the image before, during, and after the scanning process. Some of the more common types of manipulation (both unintentional and intentional) are discussed later in this paper.

Clearly, it is safe to say that scanning technology is an amazing process.

Case Scenario

Microscopic examination of the CTS counterfeit certificate revealed it to be a four-color copy (excluding the serial number) made by a four-color inkjet process (printer). The reproduction of numerous plate defects indicate it was derived from an original certificate, although not necessarily the one provided for comparison. The most likely means of creation was by computer and digital scanning device (hand held or flatbed scanner, video capture board, etc). The confusion began with the discovery of tiny black marks in the background of the counterfeit which could not be correlated to plate defects on the original. If one accepts that the counterfeit was scanned and copied, there are only four possible sources for each of these marks: they can be contaminants or defects in the original document that were reproduced in the scanning process; they can be photocopier-type defects caused by dust, scratches, etc., on the scanner platen, lens or mirror; they can be defects that were made by the associated equipment (computer, scanner or printer) during the scanning process; or they can be contaminants or defects in the (counterfeit) paper. Based on this logic, several test "cases" were prepared and various theories were tested in an attempt to account for the background marks.

Details of the Study

Two scanners were available and used for this study. A Hewlett Packard ScanJet 4C flatbed scanner was available in the document section of our laboratory in Roanoke. It is capable of scanning in black-and-white, grayscale, or full color (24-bit). Optical resolution is 600x600dpi, with a maximum of 2400dpi via interpolation (enhancement). A privately-owned Envisions ENV8800S scanner was also available. It will scan in modes similar to the HP 4C, except the color is 30-bit. Optical resolution is 400x800dpi, with a maximum of 6400dpi via interpolation. The computer systems that are attached to the scanners are typical Pentium systems: a P-200 with 32mb of RAM and a P-90 with 16mb of RAM.

Scanned results were printed either directly from the scanner, or from saved graphics files using various graphics programs and digital file formats. Four printers were used: a Hewlett Packard (HP) LaserJet 4, capable of 600x600dpi printing; a HP DeskJet 692C, four-color inkjet, capable of 600x600dpi printing; an NEC Superscript 100C, three-color inkjet, capable of 600x300dpi printing; and an Okidata OL600e LED, capable of 600dpi-class printing.

The test-problem scenarios were intentionally kept simple. Various types of white paper (xerographic bond, inkjet and lined tablet paper) containing a hand-drawn letter within a 2-inch border (for reference) were scanned and printed, then examined and compared to the resulting graphics files and printed results. Specifically, attempts were made to correlate background marks on the printed results to contaminants and defects in the paper. Next, sheets of blank paper were scanned and their printed results were examined in this same manner. (One test involved another examiner scanning 10 sheets of blank paper, marking each copy, and requiring them to be matched up to their original.) Then, office letterheads were scanned and printed as "forgeries" in an attempt to create a realistic case. Finally, copies of the CTS original certificates and counterfeit certificates were examined and compared. The scan passes used for this study were mostly black and white (1-bit), which not only duplicates the typical cut-and-paste scanner forgery, but also produces the most identifiable results.

It appears both scanners were effective at reproducing contaminants and defects in almost all of the test papers. Bits of wood, black specks of apparent dirt in the paper, pinholes and shadows, and dust particles trapped on the surface of the papers between paper fibers produced marks in most of the printed results on all types of paper with all four printers. Close examination of the graphics files revealed the reproduced marks consisted of small groups of pixels, usually between one and six. Depending on the printer, the results could appear as either black marks, random colored marks, or three-color black marks. However, the vast majority of marks appeared as black (probably due to the 1-bit scans).

Surprisingly, there were more defects in most of the papers than the author suspected would be found. In fact, there were a minimum of five (5) contaminants in any of the test areas on the first set of (lettered) papers while some had many more. All copies of these test sheets could be matched to their respective originals. Scans of entire sheets of paper sometimes revealed a dozen or more identifiable contaminants, pin holes and shadows. The copies of the 10 sheets of blank paper could easily be matched to their respective originals.

These types of contaminants and defects, and their reproduction by the scanner and printer, would seemingly allow an original document to be matched to either: a saved graphics file; results printed from a saved graphics file (a counterfeit); or results printed directly from the scan (a counterfeit). In other words, one could take a graphics file located on a suspect computer and either print out a copy or view it on the monitor, then correlate many of the background marks to contaminants or defects in an original document and determine the graphics file was derived from a direct scan of that document. One could also take the counterfeit (as printed from a scanner or graphics file) and correlate many of the background marks to contaminants or defects in the original.

Both scanners also produced a consistent pattern of marks that appear to have come from the scanner platen or internal components. If two scans were made without moving the document, recurring marks would appear on multiple graphics files (or printed results) that did not correlate to defects in the paper. These marks are the result of a photocopier-type defect, i.e., scratch, dirt, or opaquing fluid on the platen (or some other defect in the scanner), and could aid in an identification of the scanner. As with the last type of marks, these were black in color (due to the 1-bit scans), but could be color or three-color black depending on the type of scan or printer. Close examination of graphics files revealed the reproduced marks consisted of small groups of pixels, usually only one or two, however the author's personal scanner had a scratch on the platen which left a nice (large) defect on every scan. Note that the controlled nature of this study contributed to the likelihood these type of marks would appear on multiple copies.

The third type of marks were totally random in nature, and are apparently the result of interference caused by poor shielding of the computer, scanner, printer and cables. (This interference is known as "SCSI chain noise" because it was originally believed to have originated within the SCSI chain: card, cables and peripherals. In fact, both of the scanners cited in this study did use a SCSI interface.) Virtually all of the graphics files contained some of these random marks that could not be correlated to the evidence under any circumstances and did not repeat with each subsequent scan. Depending on the printer, the results appear as black marks or random color marks. Further, because of the random nature of these marks, light or dark marks may appear within a printed area. At first, this phenomenon would appear to be a nuisance, but it is actually quite important because it can allow a differentiation between two successive scans saved as different graphics files. (This is because consecutive scans may duplicate the first two types of marks—paper contaminants and scanner defects/dust—but should never duplicate the random noise pattern.) Thus, the graphics file may have its own "fingerprint." In other words, a specific graphics file located on a computer could be correlated to a printed result (counterfeit) if it contained the same (or near exact) noise pattern. Unlike the paper contaminants and equipment defects, close examination of the individual noise pattern marks in the graphics files revealed they frequently consisted of only one pixel. Subsequent examination of these marks on the scanned copies allowed them to be matched to the specific graphics file that created them (Appendix A).

It should be noted the LED printer complicated the search for identifiable marks by camouflaging them. The nature of many LED/laser printers to leave a "scatter pattern" of small marks in the background of the papers made it difficult to tell the difference between them and the SCSI noise marks and machine defects. The correlation could still be made, especially between the paper defects and some scanner defects, however some marks were harder to verify than others. This problem was not encountered with the inkjet printers or HP LaserJet. The inkjet printers produced the cleanest copies and were easiest to work with.

Another problem that could arise during an actual examination is that one could not easily differentiate between SCSI noise marks and equipment defects (scanner platen defects/dust), especially if only one questioned document was available. Although it should be quite easy to match the results of a scan (graphics file to original, or printed results to graphics file and/or original) it could be very difficult to match the graphics file or printed results to the scanner. Even with multiple questioned documents, the variation in noise pattern could make it difficult to find the few marks caused by equipment defects or dust unless the mark was a scratch or larger platen defect. Multiple prints from the same graphics file would be needed to differentiate one from another (to correlate marks on the graphics file to the copy).

Application of the Study to the CTS Problem

The results of the aforementioned study allow for some interesting conclusions, as well as expectations from the CTS problem. First, it should be able to be determined if the counterfeit the author examined was derived from the specific genuine certificate provided as a control in the test. Second, if the CTS used the same graphics file to print all counterfeits for this problem, each counterfeit should contain similar marks as derived from four sources (plate defects, paper contaminants/defects from the genuine document, SCSI noise pattern, and scanner defects). Third, if the genuine certificate used to make the counterfeits was retained by CTS, it should be able to be determined that it was the specific genuine document used to make the graphics file.

The genuine certificate provided in my CTS problem was re-examined and the defects were compared to the marks on the counterfeit. It was easily determined that this was not the genuine document used to make my counterfeit, due to the presence (and absence) of relevant paper contaminants and defects.

It was next determined that CTS used the same graphics file to print out all of the counterfeits. This was due to the presence of the same noise pattern (and/or scanner defect pattern) indicating the same scan pass, and was verified by an examination of several original "counterfeits" as well as additional ones made by the author with the original graphics file.

Finally, the genuine certificate used by CTS to make the graphics file was obtained and examined, and it was easily determined that it was the specific genuine document scanned to make the graphics file. This was due to the correlation of paper contaminants present on the genuine certificate to the marks on the counterfeits.

As a bonus, while examining the genuine certificate (used to make the counterfeits) it was determined that the graphics file used to print the counterfeits was probably converted at some point from a black and white file to the final color version. This was due to the fact that many small plate defects, which should have been printed in color, were instead printed as black marks. Further, more of the SCSI noise marks should have randomly been printed as color. The only explanation is that the file was converted from a black and white file (possibly to save disk space and/or time), then changed to color after-the-fact. A discussion with the CTS representative confirmed this theory. Out of concern that the scanner might assign multiple-value (graduated) colors rather than two-color values to the red and blue certificate, the representative scanned the original directly into a paint program as a black and white file then converted it to red and blue using a "fill" function. Because the defects were so small, and unattached to the larger print, they were overlooked and not converted to color.

Component Alteration and Intentional Manipulation

It is recognized that the results of this study are optimal and rest on best-case scenarios. Most of the studies were done on typical white xerographic bond which probably contains more contaminants and defects than some more expensive specialty papers, thus contributing to the quantity of defects. The scans were made in a controlled setting, with well-maintained and state-of-the-art equipment. Further, the test designs were uncomplicated and did not take up a lot of the paper space, also contributing to the quantity of defects. But in the opinion of the author the results are significant. This type of examination will be most successful if a perpetrator is not computer "smart" and uses default settings, or is confident enough that manipulation of the scanned image is not deemed necessary.

There are many ways to unintentionally manipulate the results of the scan pass, graphics file and printed result during or after the scanning process. The optical resolution of the scanner is almost always intentionally or unintentionally altered (frequently reduced) by the scanner software, printer software, or a graphics program. For example, to save on disk space many people scan in a 300dpi (or less) mode rather than 600dpi. Reduction in the resolution had a dramatic effect on the recording and printing of various defects. Some defects disappeared and did not even show up in the graphics file or printed result when scanned at 300dpi. Conversely, the resolution may also be enhanced during a scan to increase detail (regardless of the printer's ability to reproduce the detail). This can occur with the scanner software, or done after-the-fact in a graphics program. These alterations affect the number of pixels per square inch, sometimes reducing and at other times creating them, and may cause single pixels to disappear completely or become odd pixel groups. Regardless of the scan rate, 100dpi or 3200dpi, the printed result can only be as good as the printer's capability. For example, the printer may print at only 300dpi. Additionally, many newer programs employ color correction or other filters that pick up "stray" pixels and eliminate or average them into the background color.

More importantly, there are ways to intentionally manipulate the results during or after the scanning process. Some of the common alterations include enlargement or reduction in size; brightness/contrast settings; color settings; filter and edge routines, such as sharpen, blur, despeckle; interpolation for graphics resolution; mirror, reverse, flip, stretch; color conversion and manipulation; touch-up and drawing routines; and so on. As an example, the "despeckle" filter removes virtually all unattached single pixels, and thus almost totally eliminates SCSI noise marks. Because of the quantity of processes that can be used to manipulate, and their complicated nature, no attempt is made here to discuss how the manipulations take place. However, note that a computer literate person could manipulate the results at almost any step of the process, and make it difficult to reach a productive conclusion.

Summary

The results of this study can be summarized as follows:

- If a specific document was scanned and the results printed out directly to a printer, it could usually be determined that the printed result was derived from that specific document (due to contaminants in the paper).

- If a specific document was scanned and saved as a graphics file, it could usually be determined that the file was derived from that specific document (due to the contaminants in the paper).

- If a specific graphics file was used to print a document, it could frequently be determined that it was made from that specific file or a derivative file (due to the "noise" pattern in the graphics file). The incidence of noise marks was dependent on the system and its location.

- If multiple scans of different documents contain the same marks, there are defects within the scanner (platen, lens, etc.) such as scratches or dirt that may allow it to be identified.

Acknowledgments

The author would like to thank Collaborative Testing Services, and their representative, Gordon Menzies, for providing information and materials for this study.

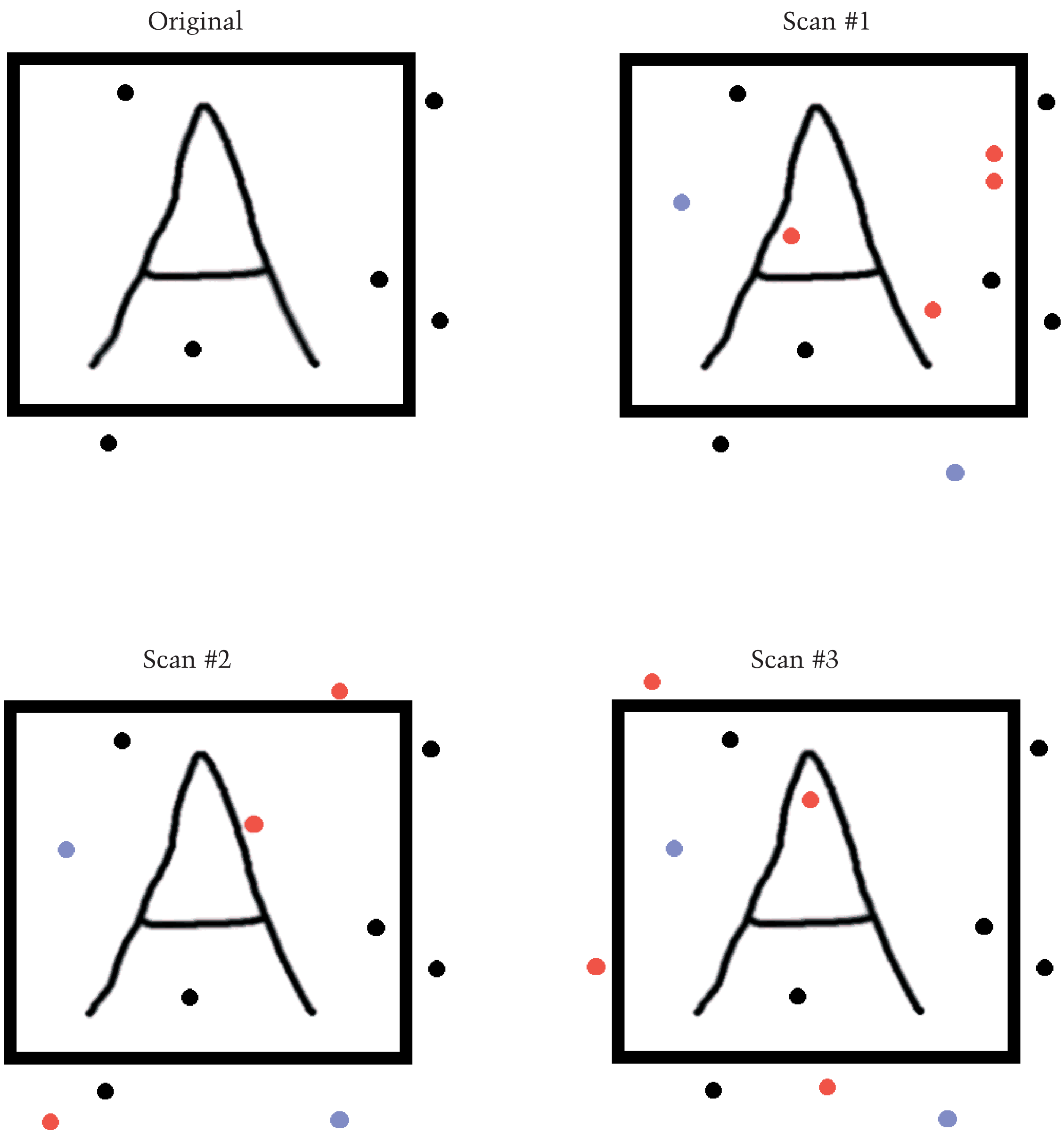
The author would also like to thank Joe Costa, technical expert, UMAX Technologies Inc., Fremont, CA, for providing information.

References

O'Hara, S., et al (1997). *Using PCS*, Cue Corp., Indianapolis, IN.

Baxes, G. A. (1984). *Digital Image Processing- A Practical Primer*, Cascade Press, Denver, CO.

Appendix A



These drawings depict the types of contaminants, defects, and SCSI noise observed during this study:

- Defects in the paper (dust and other contaminants)
- Copier-type defects, from scanner platen (scratches, dust, etc.)
- SCSI noise (interference)