Heuristic for Accelerating Visual Examination of Questioned Documents

ABSTRACT

A technique for preparing and examining questioned documents that have been damaged by photocopy, photostat, offset print, and other Nth-generation processes. The technique involves scanning texts and converting them into 3-D images that make subtle variations more visible. The visualization process is first demonstrated on documents of known origin, then tested on texts of questionable origin.

INTRODUCTION

In a classic discussion of procedures for examining questioned documents, Wilson Harrison describes the beginning of the process as, "Apart from a magnifying glass and possibly a metal rule, all that is required is a comfortable chair and a table of convenient height placed in good light, ample patience, and last but not least, a healthy skepticism" [1].

The implication is not that document examiners need few tools, but that first they need time and patience and ongoing doubt. Time and patience, however, may often be in short supply. People, left for too long without answers, may simply began manufacturing answers of their own.

Possibly most time is consumed and most patience is demanded in the lapse between when examiners sit down and when they begin to discover the subtle details leading to informed conclusions. Often the things that initially spring out are irrelevant once tested, and the important information shows up in details that become visible only after continued examination.

My goal

In recent months, I have developed interesting and perhaps valuable tools for examining photocopied (or otherwise obscured) questioned documents. I have created several heuristics that permit me to tease meaning from what may seem to be meaningless information. One of these heuristics is a process for scanning characters of interest and converting their many, subtle, and varied shades of gray into apparent levels of altitude – a cartographic map of sorts. The apparent levels of altitude often highlight characteristics in texts that might otherwise pass unnoticed. I have been unable to find a description of this process in the relevant literature, so for my purposes I have named it "3-D visualization." This paper demonstrates the 3-D visualization process by examining how it can be used to evaluate a number of conditions that exist in typed and printed documents of known provenance. However, it also presents insights and procedures applied to a questioned document – a photocopy of a memo purported to have been written in August of 1972 and signed by Jerry Killian.

Problems with photocopies of complex documents

I need not detail all of the standard practices for examining typed texts. They are available in any of a number of texts discussing the topic [e.g., 2,3]. In some cases for monotype typewriters, just examining vertical and horizontal alignment can be enough to identify a specific typewriter based on a page of typed text. On the other hand, leading and character spacing in some proportional typewriters can be changed character by character and line by line, and once the documents have been photocopied, traditional approaches become limited. Not only is it difficult to demonstrate which typewriter might have produced a document, it can be difficult to demonstrate that the document was even typed.

In 2004, CBS 60 Minutes presented a collection of memos exemplifying this problem. They were typed or printed, then photocopied numerous times and finally faxed (twice in one case) [6]. The documents were purported to have been written in 1972 and 1973, and implied that the then 1Lt George Bush shirked his Air National Guard duties. Fundamental criticism was the memos were digitally produced and not typed, and therefore were forgeries. Peter Tytell took on the task of identifying the physical source of the documents for the Blue Ribbon Panel headed by Dick Thornburg and Louis Boccardi. Given the condition of the documents, Tytell’s task approached impossible. With originals he would have known the answer instantly, but without tools for examining texts with this level of damaged and distortion, his final suggestions, published in “Appendix 4” of the report, were little more than educated guesses (although they were generally accepted as fact and became a part of American history). In hindsight, Tytell is probably wrong. Had he and the panel sufficient time to acquire better copies, his answer would almost certainly have been different. As Harrison said, “time . . . patience . . . skepticism. . . .”

In discussing this memo, questioned documents experts have argued that it is impossible to authenticate them. I agree. On the other hand, David Moore argues, “Document examiners are often asked to examine copies in lieu of originals. Although copies frequently limit the forensic examination, they sometimes provide evidence that greatly increases their investigative importance and prosecutorial value.” [8] In Forensic Signature Examination, [9] Steven Slyter concurs that “Copies need not be an unworkable limitation in a case.” While it may be difficult and often impossible to authenticate photocopied documents, they still have a great deal to tell us. It seems to me that, authentic or not, these memos are interesting and important, and we should try to eventually know everything they have to say – not so much so that the documents can be authenticated, but so that we (and history) can know what truths they may divulge in the end.

SAMPLE CHARACTERISTICS AND PROBLEMS

Although even poor photocopies might interfere with the collection of evidence, The tools I demonstrate, along with an understanding of typewriters (in this case, IBM Selectric typewriters) makes it possible to extract a surprising amount of information. In figure 1, below, I present a combination of two typed, overlapping characters: a dash and an “s.” The character depicted in Figure 1.1 was typed on an IBM Selectric III. The character in figure 1.2 was inkjet printed with an HP Photosmart 8450.

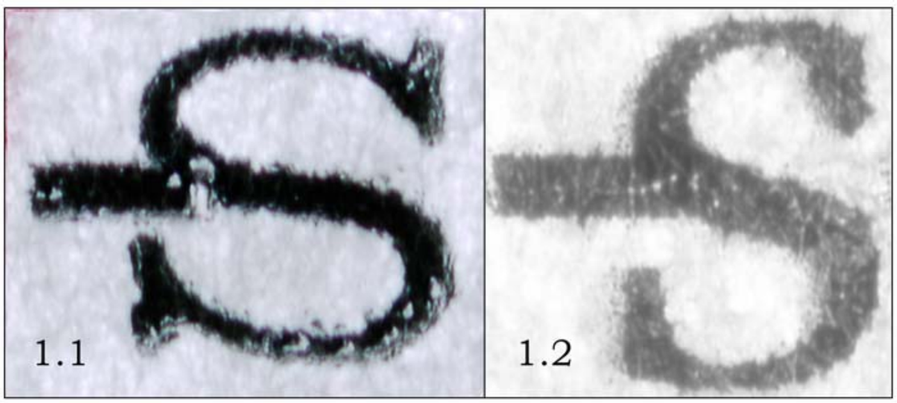

Figure 1: Image 1.1 (above left) is scanned normally. Image 1.2 (above right) is scanned as film negative. Image 1.3 (bottom) is scanned from an inkjet printed document.

The differences are immediately obvious if we examine how the characters interact. Less obvious but nonetheless important is the possibility that a careless examiners looking at Selectric generated type may leap to wrong conclusions. In figure 1.1, the dash appears to be on top of the “s,” (implying it was typed last) although clearly, it was typed first. The illusion is caused by the unprinted space after the dash. The space penetrates the “s” creating a confusion of their chronological order. In contrast, the image in figure 1.2 exemplifies a digital, typewriter typeface (Passport) but manifests no comparable confusion of chronology. To the extent we can see the left edge of the “s,” it appears (as it should) to be in front of the dash.

Second sample – more subtle problem

The “am” combination (figure 2) provides a similar though more subtle example.

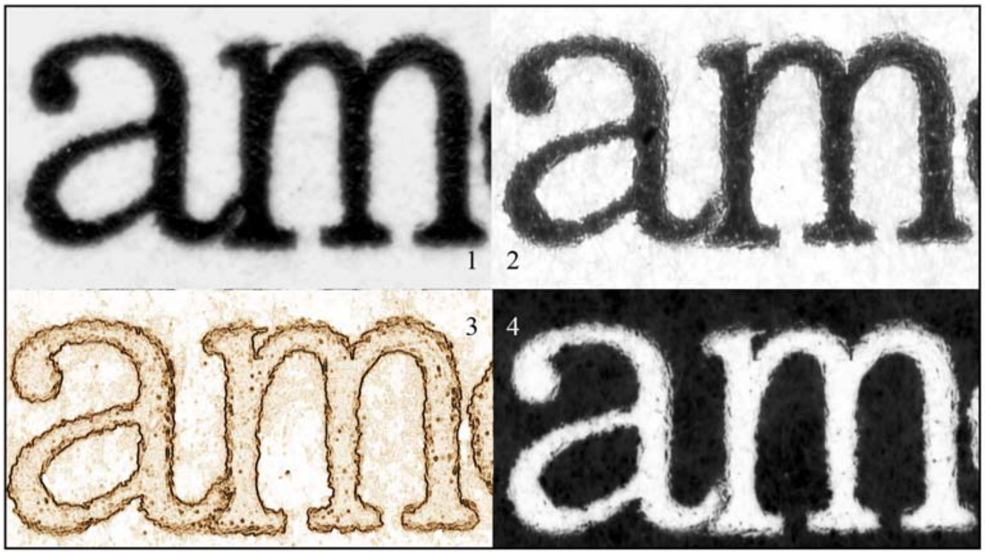

Figure 2: “am” typed with an IBM Selectric II. Scans are (1) normal, (2) positive rear projection, (3) normal with 3-d visualization, and (4) negative rear projection.

In this case, the terminal on the “a” is surrounded by a faint halo of lighter gray. In the figure below, Figure 1.1 is a direct scan at 9,600 DPI. Although it is possible to see the halo between the “a” and the “m” in this image, the relationship is subtle and could easily be missed by a document examiner. Moreover, exactly what is happening between the characters might be difficult to demonstrate to a lay person.

Figures 2.2 and 2.4 are alternate approaches to examining the characters. In these cases, “spoofing” the scanner into rear-projecting its light source through the page (as film positive and as film negative) makes the relationship between the “a” and “m” much more immediate. If one understands the physics of typing, sufficiently enlarging either image makes it possible to make a strong case that the “a” was typed first, and the “m” was typed on top of it.

Finally, Figure 2.3 was produced by converting image 2.1 using a 3-D visualizing technique. Note that image 2.3 is every bit as descriptive as images 2.2 and 2.4 for showing the interaction between the characters. Because the rear projection capabilities of scanners are usually limited to slide- or negative-sized spaces, creating images 2 and 4 required identifying specific characters I wanted to examine and focusing the scanner on them. Image 2.3, however, was created by converting an entire page for examination. With this tool, an examiner can scan a page at 9,600 DPI, save a copy, and convert a copy into a format that makes oddities such as the one above particularly easy to spot.

Third sample – complicated problem and value of 3-D visualization

Sometimes, interaction between characters is particularly difficult to understand (Figure 3) My third example is a strikeover involving a “c” and a “v.” The most difficult question that arises is, “Which character was typed first?” The answer is immediately clear if I point out that the word “mevhanical” is corrected to “mechanical.” On the other hand, how would a document examiner know that the strikeover is a correction and not a contrivance? How does the examiner know the "v" was actually typed first?

Figure 3: "c" "v" combination. These characters are typed in 1985 with a well used IBM Selectric II typewriter and scanned with an Epson Perfection 4180 scanner using default settings at 9,600 DPI.

After a bit of time it might be possible to see that the ball of the "c" in figure 3 seems to be in front of the right serif of the "v," but the crotch of the "v" seems to be in front of the bottom of the "c." Surely a contradiction; they cannot both have come first. To make matters more confusing, the left serif of the "v" appears to be in front of the "c," and at the left stem of the "v" the characters simply blend. In some cases, interactions are visible under a microscope, but in some cases they have to be teased out using more proactive visualization methods.

The characters above are made up of shades of gray evolving from very dark to almost white. By identifying the shades of gray in the characters and remapping them as planes of altitude, it becomes possible to disclose their different interactions. All of the many textures in the image in figure 4 result from the largely invisible complexities of values of gray that actually exist in the characters.

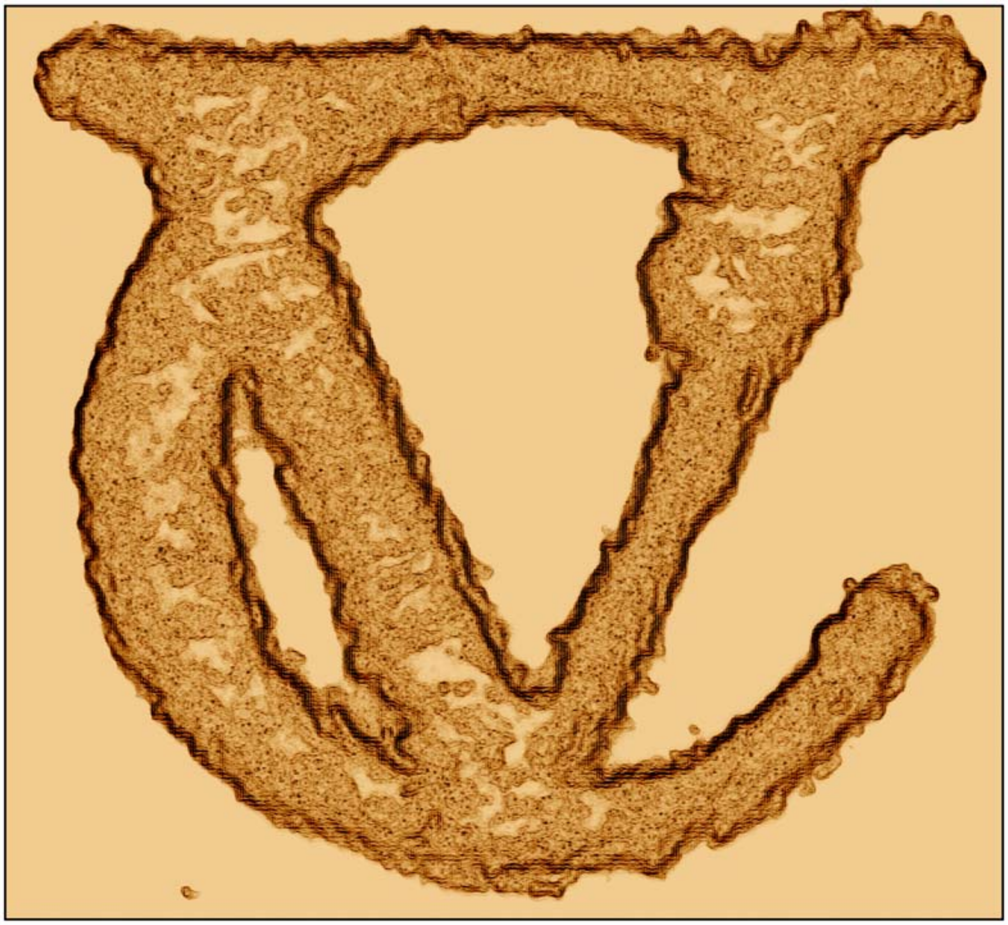

Figure 4: The textures in the above characters are caused by the many shades of gray hidden within the apparent blackness. For the "educated" eye, even interactions between paper fibers and ink become apparent.

Now, a careful examination of the left serif on the "v" shows its entire shape. The crotch of the "v" is even more visible. More importantly, with careful examination, the left stem of the "v" can be extracted from the stroke of the "c." Finally, as I said earlier, the ball on the "c" seems to be in front of the right serif of the "v." In short, all of the information one might eventually parse out of the image in figure 3 is much more readily visible in figure 4, and new information is available to us. Nonetheless, it is possible to parse even more details.

Rear projection and 3-D visualization combined

By their natures, characters typed with carbon ribbons are especially dense and reflect little light and so can hide subtleties even at microscopic levels. Still, these weaknesses tend to be translucent when back lighted. This translucency can be exploited by rear projecting light through the page. The translucent elements become obvious, and the characters become much easier to understand.

There are four images in figure 5. Image 5.1 results from treating the page as a film negative. Image 5.2 comes from treating the page as a film positive. In image 5.1 translucent weaknesses in the characters are shown as ragged black lines passing through them. Film positive weaknesses, of course, show up as pale breaks in the characters. Also, interactions between the characters show up as ghostly shadows. Still, in both cases it is possible to make their interactions even more clear, by adding 3-D visualization:

In image 5.3 the exact shape of the "v's" serifs are now highly visible. The interaction between the bottom of the "v" and the bottom of the "c" is similarly visible. It becomes easier to pick out the pale shadow along the left edge of the "v," where it interacts with the thick stroke of the "c."

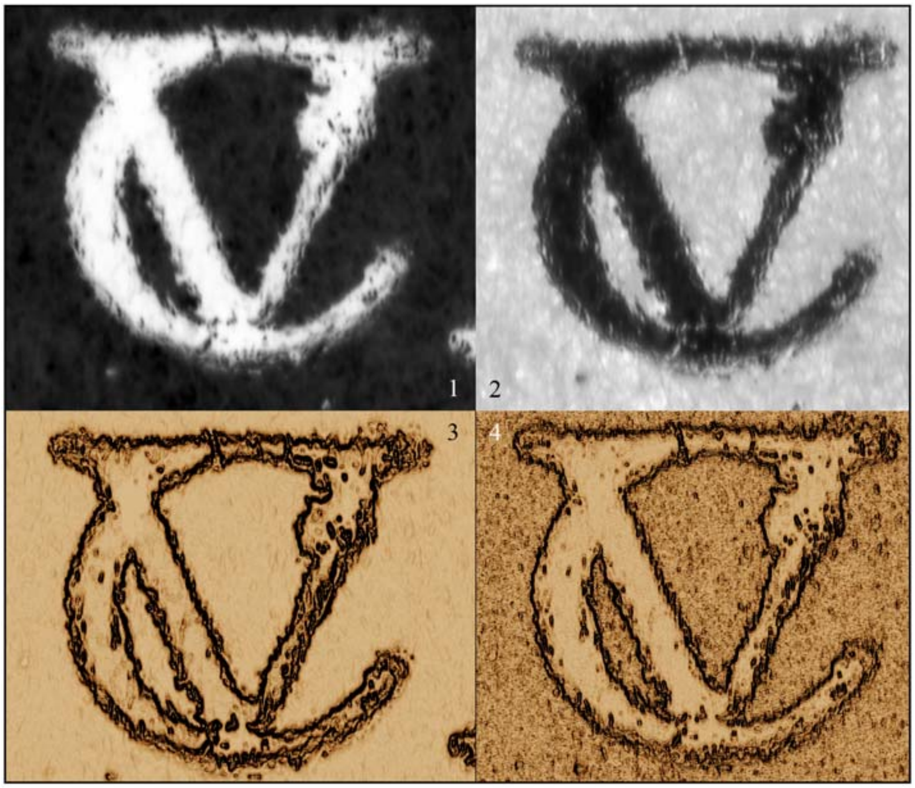

Figure 5: Four views of the "cv" chimera: 5.1 and 5.2 use rear projection with negative and positive film settings; 5.3 (5.1 modified) presents the characteristics of the character best, while image 5.4 (5.2 modified) presents the characteristics of the paper best.

Apart from the ambiguity between the right serif of the "v" and the serif of the "c" the "v" seems to be in front of the "c." Although it seems to contradict logic, that the "v" seems to be in front makes the case for it having been typed first. It represents a phenomenon where IBM Selectric typewriters frequently fail to type new characters over old impressions. The space we saw in the previously described dash "s" exemplifies the phenomenon, the "am" exemplifies it, and the strikeover exemplifies it. Since the "v" was already in place when the "c" was typed, large segments of the "c" simply failed to print. In addition to being unable to print on the "v," segments of the "c" failed to print in the space <u>near</u> the "v." For example, the top of the "c" fails to print near either of the "v's" serifs leaving gaps in both places. At the bottom of the "c" there is a gap on either side of the point of the "v." With the above image, the "c" appears to be behind the "v" precisely because it is in front of the "v." The "v" was typed first, and the "c" was typed over it, but it failed to print in most of the already printed places – leaving the impression the "c" was typed first and is hidden behind the "v." This characteristic with overprinted texts is (to the best of my knowledge) unique to typewritten pages.

Fourth sample – when subtle clues are destroyed

Photocopying may remove a great deal of subtle information from a typed page, but photostatting and offset printing processes remove significantly more. Nonetheless, they do not necessarily remove it all. In some cases, even with photostated and offset printed pages, characters that fail to overlap, interact uniquely and can be used to close the argument for having been typeset with a typewriter. The following examples were typeset using a typewriter, then offset printed as a book (San Francisco: The Urbanized Estuary). On any page of the book it is possible to find examples of characters that are typed near each other without actually overlapping. In many cases (because typesetting with a typewriter is sequential), one character will change the outline of the previous or next character. If the second character impacts close to the first character and deforms the paper or pops loose carbon from the impression, the first character may be deformed. These deformities whether they occur on the first or subsequent characters are indications of sequential impact printing.

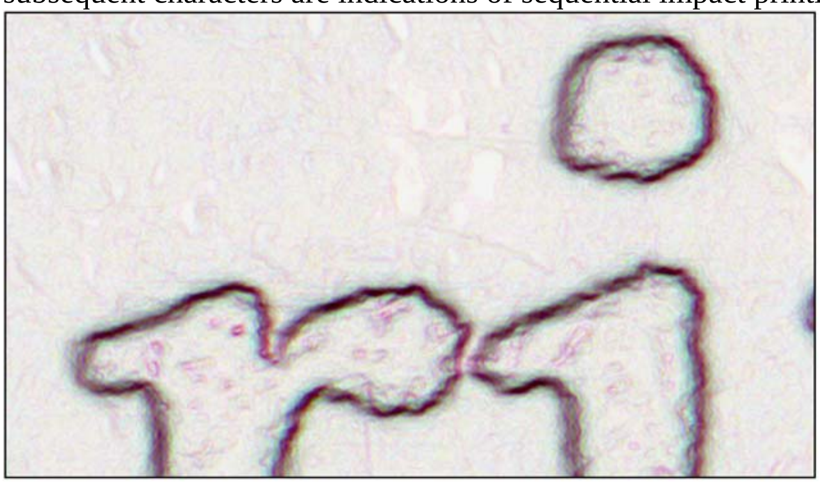

Figure 6: Characters taken from the book *San Francisco: The Urbanized Estuary.* Typeset with an IBM Composer typewriter.

Although the "r" in figure 6 was typed first, the impact of the "i" distorted its terminal ball. Curves such as the one in the loop of the "g" (figure 7) show an identical (and common) tendency to distort.

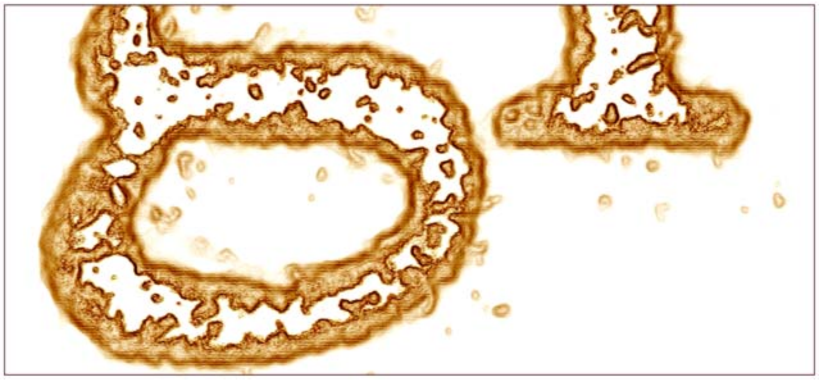

Figure 7: Top row: "gr" taken from a page typeset with a Composer typewriter. Bottom row: "gr" from top row, plus two additional samples where the "r's" are far enough from the "g's" not to impact them. From *San Francisco: The Urbanized Estuary.*

Still, the second character can be just as likely to show the same kind of curved distortion. I believe the cause, however, is different. The "g" seems to be distorted by the impact of the next character. I believe that the "4," below, is distorted because it is unable to print on the outer edge of the impression made by the dash, and so marks the boundaries of the dash's impression. This is an example of characters behaving like the characters in figures 1-3. These interactions are rarer and more difficult to spot, but still occur on every page making it possible to demonstrate numerous examples of sequential impacts.

Figure 8: The leading edge of the "4" suffers from near impact with the dash. Other "4s," with no nearby characters suffer no such deformities. From *San Francisco: The Urbanized Estuary.*

Fifth Sample -- Interaction With Ball Point Pen

Ball point pens sometimes interact with typed texts. At the very least, the pen will sometimes skip immediately after passing over an impression. But under the best of conditions, the pen will actually track impression long enough to give away the direction of its stroke.

The sample in Figure 9 shows how a clickable ballpoint pen behaves as it crosses a diagonal, scoured line comparable to a typed impression. In a series of tests, a lightly held pen consistently moves slightly to the left as it drops into the diagonal. Sometimes the pen then follows the impression to the right until pressure from the downward stroke forces it back onto its original track (see below). In all other cases, the leftward pressure overrides the tendency and the pen comes out of the groove somewhat to the left of its entry point. In either case, interaction between pen and groove is identifiable.

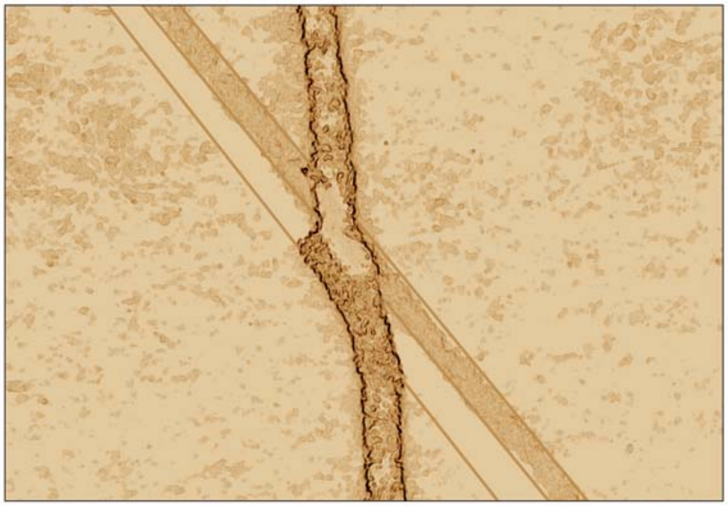

Figure 9: 3-D visualization of a sample vertical line drawn with a ballpoint pen through scribed diagonal (parameters of diagonal score are highlighted for clarity).

A real world application

In July 2005 I was given a copy of a memo alleged to have been written on 01 August 1972 and signed by Lt. Col. Jerry Killian. The copy I examined was photocopied but not yet faxed. In this copy, I saw examples of the character defects and interactions I have described above, but perhaps more telling is the signature block. It is possible to examine the signature block for symptoms of interaction between signature and characters. The kinds of interactions I describe above might manifest themselves if the characters were typed and signed while the indentations were still fresh. These are completely different from laser or inkjet printed documents, which offer little or no resistance. In these cases, there might be interaction between inks, but there is no reason for the pen to change course as it passes over a character.

The left image in figure 10 shows the signature in question at approximately 4X actual size. The center image shows a single character (an “R”) from that signature block with lines passing from top to bottom and from bottom to top. The final image shows the same thing using 3-D visualization to make interactions more obvious.

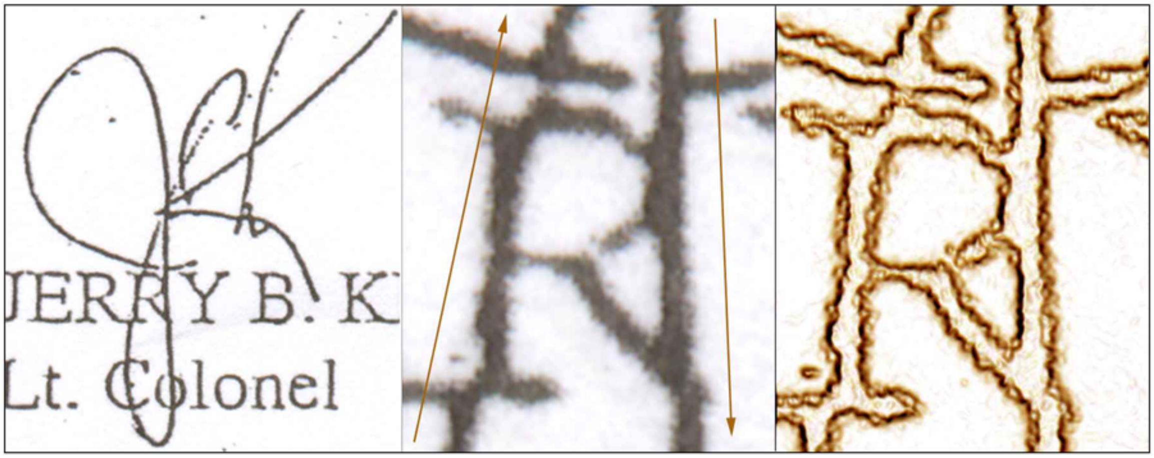

Figure 10: Left: Killian signature block taken from a questioned 01 August 1972 memo at approximately 6X. Center: "R" with accompanying line alleged to be from Killian's signature blown up to permit microscopic examination. Note: brown arrows indicate direction of signature strokes. Right: Same view of the "R" and line using 3D visualization.

Following the line from top down (Right side of the "R") presents interesting conclusions. First, the line passes close to the right edge of the "R." The line is clearly interacting with the bowl of the "R" beginning at the shoulder and finishing just before the juncture with the leg. Next, however, the line passes into the leg of the same "R." In this case the line swerves slightly to the left, then shifts to the right, tracking the leg all the way to its conclusion before continuing a downward path.

With the upward stroke of the signature (left side of the "R") the line behaves similarly. It crosses the bottom serif of the "R" and skips for perhaps one-hundredth of an inch then tracks up the stem, jumping out the top (skipping several times) and continuing its upward stroke. In short, the behavior of the pen interacting with this text seem to be unique and identifiable as a clickable ball point interacting with a typed text. I make no claims about the authenticity of the signature, however. But I do claim that it is written on a typed signature box. Others should have no difficulty achieving similar results with similar tests.

In the literature, there is occasional discussion of how to tell whether a pen is written over or under a typed word. [10] With this process, it may be possible to identify how the pen is interacting with the text, but it may also be possible to infer the age of the text when it was overwritten. An old text may have insufficient depth to impact the stroke of a pen.

Defective characters

Defects in characters become particularly visible when subjected to 3-D visualization. In this case (figure 11), I scanned the entire page at 9,600 DPI and mapped the entire text to 3-D. In PhotoShop it becomes possible to very quickly scour a high-resolution and magnified segment of text for weaknesses, and should an examiner find what appears to be a weakness, he or she can quickly examine for comparable weaknesses in the same character throughout the document. On this page, the edge in the "w" quickly displays a weak right-stem at the crotch. Once the potential weakness is identified, an analyst can quickly check other "w's" for similar problems. In this case, the problem persists throughout all "w's" in the document.

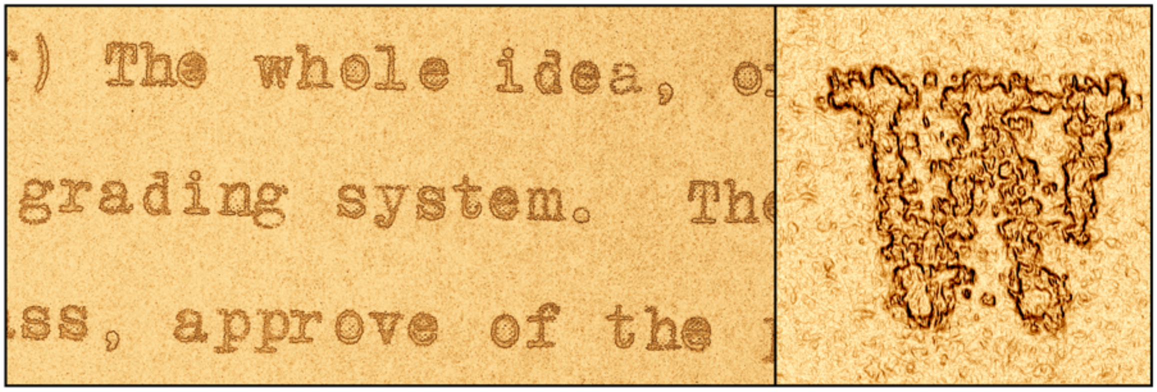

Figure 11: Example of characters taken from text typed on 1923 (circa) manual typewriter.

It is also possible to demonstrate that defects are artificial. The "e's" in figure 12 are produced on an inkjet printer using a font designed to mimic typewritten texts -- including defects common in typewritten texts.

Figure 12: Selection from a document typed in a font called "Passport" designed to mimic typed texts, including defects.

The thing that gives the "e" away as digitally produced is that even insignificant damage is both persistent and consistent. There is very little difference between characters. In typed texts, damage is more likely to appear persistently but inconsistently.

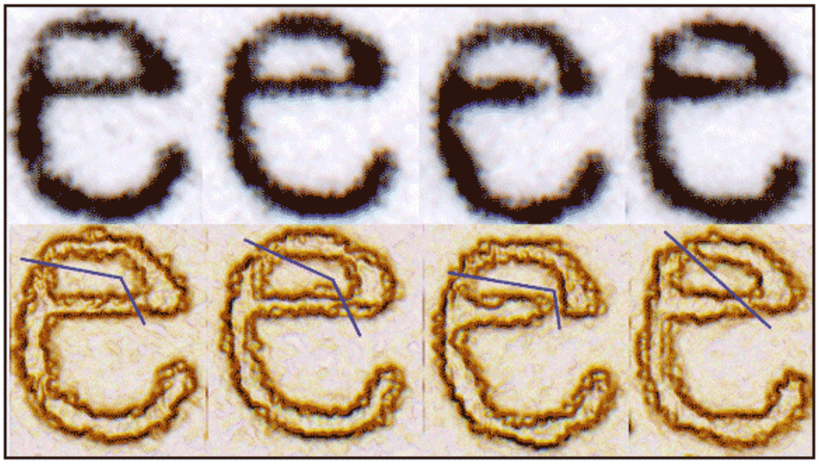

Figure 13: From a photocopy sample probably typed with a defective typewriter (extracted from the 01 August memo).

The apparent damage in the images above is persistent through all “e’s” in the document but inconsistent in its specifics, suggesting they were typed. This may not be conclusive on its own but might be used to add to the weight of other evidence.

THE PROCESS

The process for producing the 3-D visualization is simple but requires some calibration and some understanding of exactly what is happening to the texts as they are being manipulated. Depending on what I want to look at, I scan at between 4,800 and 96,000 DPI while changing brightness and contrast to highlight specific elements. For example, if I want to examine the nature of the paper, I underexpose to make the shadows of the fibers more visible. If I wish to examine the nature of the characters, I overexpose to make various values in the characters more visible. The difference in exposure is largely driven by the nature of the text. Offset, laser, letterpress, and inkjet print is denser and requires more overexposure than typed characters (which often need none). The idea is to make their various shades of gray more visible while not destroying the nature of the character. Anybody with a good scanner can do this.

By spoofing the scanner into scanning the text as film negative or film positive, I have even more options. Because most scanners with rear projected lighting have relatively small light sources, I generally use this only when I already know what I am testing for. For example, earlier in this paper I mention that lift-off correction sometimes requires multiple restrikes. Suppose I see a character I suspect has been corrected. By scanning as film negative, I can see subtle breaks in the character that will identify the original character. On the other hand, by scanning as film positive, I get a much better picture of the nature of the paper around the character making normally invisible bits of carbon obvious. This is also effective for examining photocopied characters that may have been typed with carbon ribbons (particularly texts that may have gone through the offset printing process) or laser printed.

Step two: producing 3-D visualization

Once I have imported the selected text into PhotoShop, I copy the entire text and paste it as a new layer. I convert the new layer with the "Find edges" tool (**Filter>Stylize>Find edges**). Photoshop sees all of the different values in the characters and maps their edges with contour lines. In effect, the character becomes a contour map. I use PhotoShop's "Hue/saturation" tool to reduce colors that may confuse the image (**Image>Adjustments>Hue/saturation**) and their "Brightness/contrast" tool to increase and decrease brightness and contrast to tease out nearly invisible information. Finally, I change the color to terra cotta to make everything more meaningful (**Image>Adjustments>Color balance**) and presentable. Obviously, the point is to find as much information as possible without contaminating the original image.

Step three: Examination

Keeping the text in PhotoShop, I go to full size. On a 21" monitor at 1600X1200 resolution, individual characters will be about 4" tall. I adjust character size down to the size that maximizes my ability to examine large blocks of text while still able to see all characteristics in the individual characters.

If I see something odd in the modified layer, I can make it invisible and examine the original. Usually, once I have seen the oddity in the 3-D layer, I usually have no difficulty finding it in the original text. So the 3-D visualization does not normally replace looking at the originals, it simply flags oddities that can be more carefully examined in the original, although (as in the case of the strikeovers I discuss earlier in this paper) the 3-D visualization sometimes makes sense of what would otherwise be nonsense.

Step four: Presentation

I do not testify in court, so I use the techniques for my own academic pursuits. The 3-D visualization makes subtleties much easier for me to see, to present, and clarify for myself as well as others. On the other hand, it also makes misrepresentation possible, so it is important to understand what is going on, not only because this may be a useful tool but because it can also be intentionally or accidentally misused, and we should have a defense for that.

CONCLUSION

The 3-D visualization and rear projection techniques I have presented here need more testing and developing. Nonetheless, I have found that I can quickly prepare a text for examination, making the apparently invisible spring into clear view and high resolution. And I have found that I very quickly begin finding the oddities in the text that make its characteristics unique.

Except in one case, I demonstrated these tools on texts with provenance that I already knew. I applied them to typed, photocopied, and photostated copies alongside offset, inkjet, laser jet, and letter press printed documents. My effort was to identify unique characteristics in most of the known, mechanical, printing processes, which characteristics should also occur in unknown documents. Having tested the tools, I applied them to a document (the Killian memo dated 01 August 1972) that has already been generally dismissed as fraudulent. It was not my intent to demonstrate that the memo I examined is authentic, but that it is typed. I believe the heuristics I developed have demonstrated this contention. More importantly, I believe the heuristics have demonstrated their potentials as forensic tools.

Additional research

I think there are more uses for 3D visualization than I have managed to present here. Others, more informed and creative, will no doubt have already considered their own options for the process. For example, because the process also highlights specific characteristics in ribbons, others might use the process to demonstrate (even after photocopying) that although two documents were produced on the same machine, different ribbons were used to produce them. Or 17th and 18th century, printed documents may be especially susceptible to careful examination to reveal printing processes, type inventory, typographic processes, and so forth (my interest).

In short, I suggest this tool opens the door for both quick and detailed examination of texts produced using a variety of mechanical processes. I suggest that once calibrated and understood more fully, it might become important for solving obscure problems in document examination (e.g., the "Killian memos") that until now have apparently remained unsolvable.

NOTES

[1] Harrison, W. Suspect Documents: Their Scientific Examination. Frederick Prager Publisher, New York: 1958 (p. 28).

[2] Ellen, D. The Scientific Examination of Documents: Methods and Techniques. Taylor and Francis, London: 1997 (p. 75, 76).

[3] Hilton, O. "The Evolution of Questioned Document Examination in the Last 50 Years." Journal of Forensic Science. Vol. 33/6 (1310 – 1318).

[4] Hicks, AF. "Electronic Typewriter Grids." Journal of Forensic Science. Vol. 44/1 (187-188).

[5] Crown DA, "The Differentiation of Electrostatic Photocopy Machines." Journal of Forensic Science. Vol.34/1 (142-155).

[6] Rather, D. "For the Record." 60 Minutes II. 8 September 2004.

[7] Thornburgh, D. and L Boccardi. "Appendix 4." Report of the Independent Panel. 5 January 2005.

[8] Moore, DS. "The Identification of an Office Machine Copy of a Printed Copy of a photographic Copy of an Original Sales Receipt." Journal of Forensic Science. Vol, 27/1 (169-177)

[9] Slyter SA. Forensic Signature Examination. Charles C Thomas, Publisher, 1995, (p. 68).