Printer differentiation using position-dependent character attributes

Introduction

The purpose of this study was to determine the possibility of differentiating similar printers and identifying specific printers through characterization of text using image quality analysis. Nominally identical printers, of the same make and model, can exhibit individual differences due to causes ranging from mechanical tolerance variations to interchange of consumables and components. These differences, sometimes not detectable or quantifiable by the human eye, can provide circumstantial if not positive correlation of a document to a specific printer.

Methodology

The study was performed using three Hewlett-Packard LaserJet 5Si laser printers, and analyzing five identical printed pages from each. The test page consisted of a pattern of characters repeated over the entire page, in order to determine whether consistent positional variations could be used as a kind of fingerprint. The letters “e” and “t” were used to provide features which included straight and curved segments, horizontal and vertical edges, and a combination of solid area and holes. These letters were chosen because of their ubiquitous nature in any textual printed material. It would be relatively easy to find one of these on any area of a printed page. Format of 12 point Times, also fairly common, was selected as the font type.

Figure 1: image of letters

The pair of letters was characterized in five locations on the page, the four corners and the center, to get the maximum variation in print quality due to printer differences. The metrics used to evaluate the letters include measure of the printed area of each, the area of the hole in the "e", the lengths of the outer letter perimeters, and the widths of the straight linear sections of each letter (horizontal for the "e" and vertical for the "t").

The data was analyzed in two ways. The values of the measurements were plotted against the printer unit, and against the location on the page. This enables immediate discrimination between printers for several of the metrics. For the metric vs. the printer unit, the values for each of the five pages from each machine are plotted to indicate the intrinsic range for a given printer. On the plot of the metric vs. the position on the page, the five values for each printer are averaged.

Results

Analysis of the results in this preliminary study demonstrates that there are significant differences in the measurements of the features of interest between the three printers. The metrics over the whole page show relatively consistent values on all pages for each printer. Take the printed areas of the "e" and the "t" for example. It can be seen that the range of values over the five samples from each printer are clearly different, but at a level that would be indiscernible to the human eye.

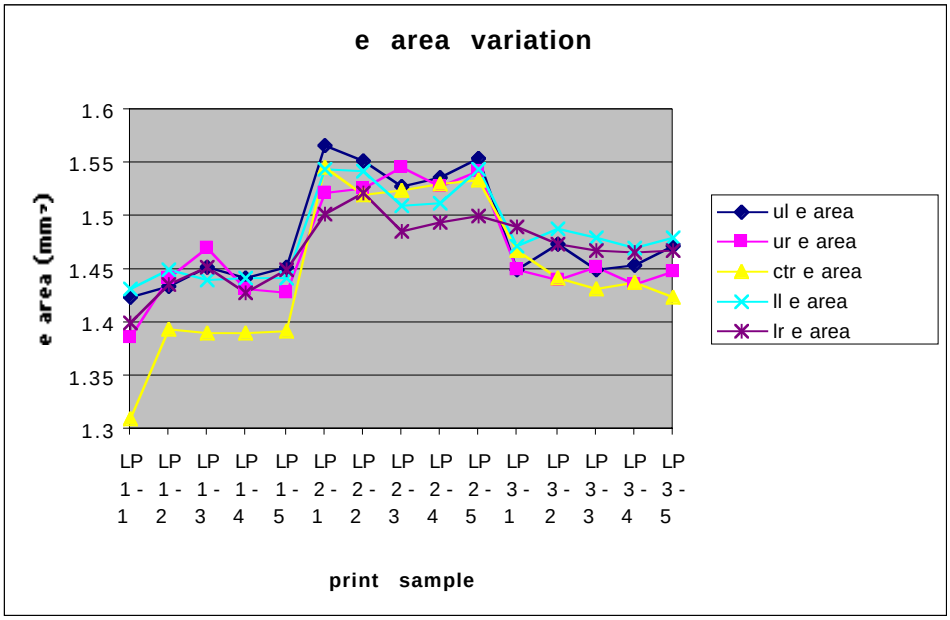

Figure 2: letter "e" area vs. printer

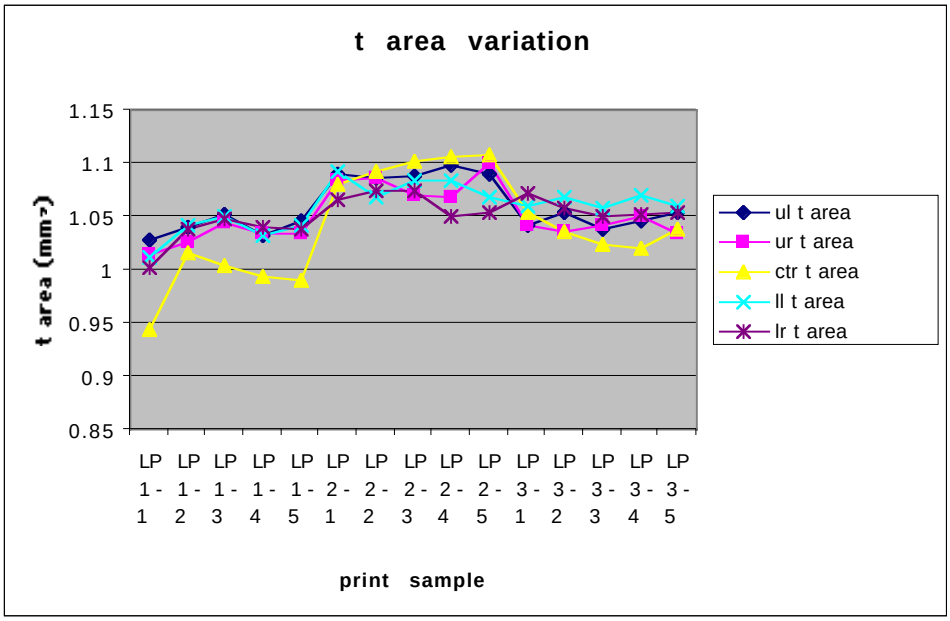

Figure 3: letter "t" area vs. printer

While the overall average variation of a feature such as the area of a letter can vary noticeably between printers, the range in the actual values can overlap making an absolute identification questionable. On the other hand, position-dependent artifacts are more supportable as positive identification. The data obtained in this study shows measurable spatially correlated variation on top of the differences in measured value due to the printer as a whole.

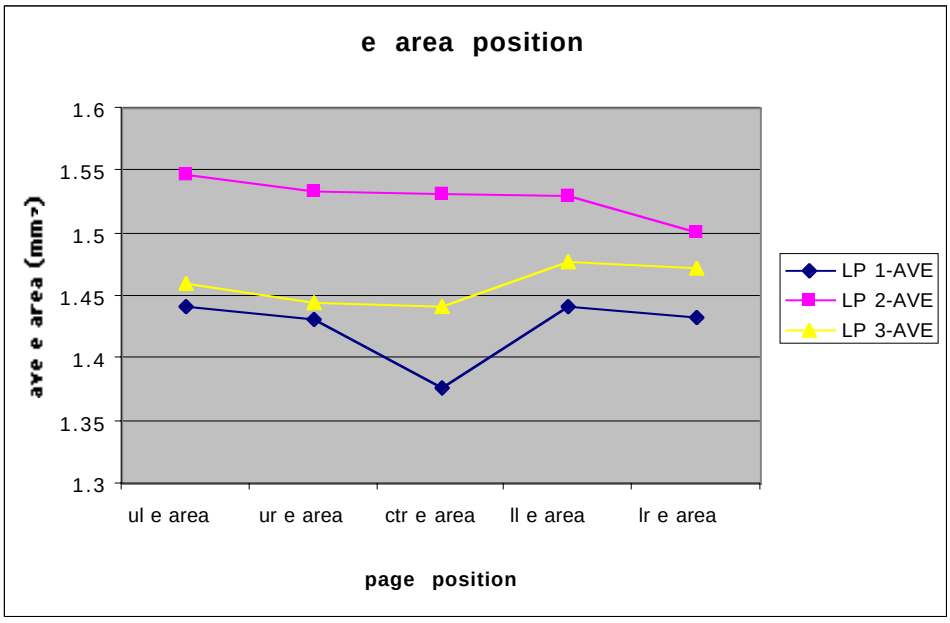

Figure 4: dependence of "e" area on page position

Figure 5: dependence of "t" area on page position

A similar metric, namely the area of the hole in the "e" character, can be equally discriminating in identifying a printer. In fact, sometimes the holes in characters have more variability than the printed areas, due to the physical and chemical properties of the printing process.

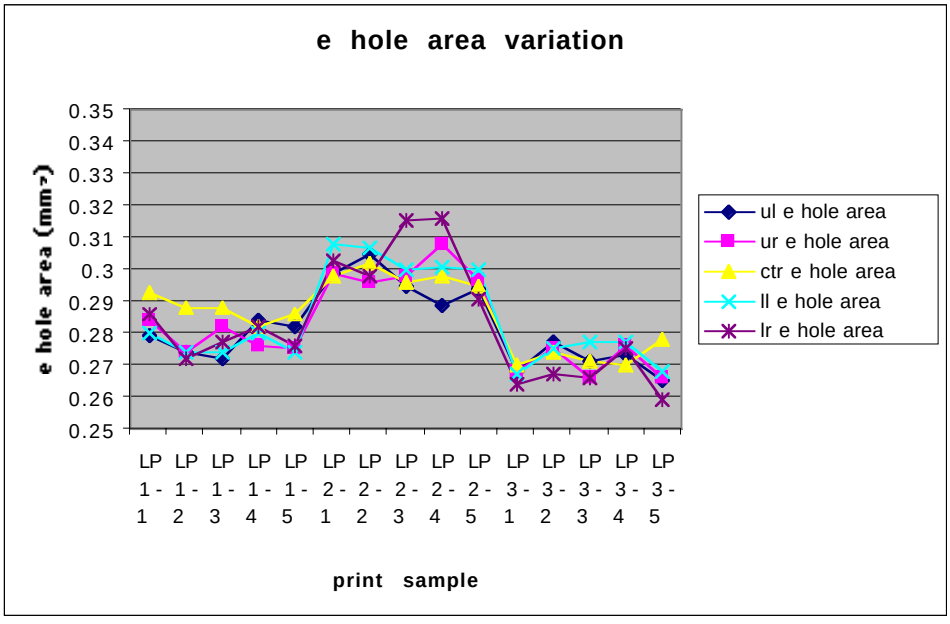

Figure 6: letter "e" hole area vs. printer

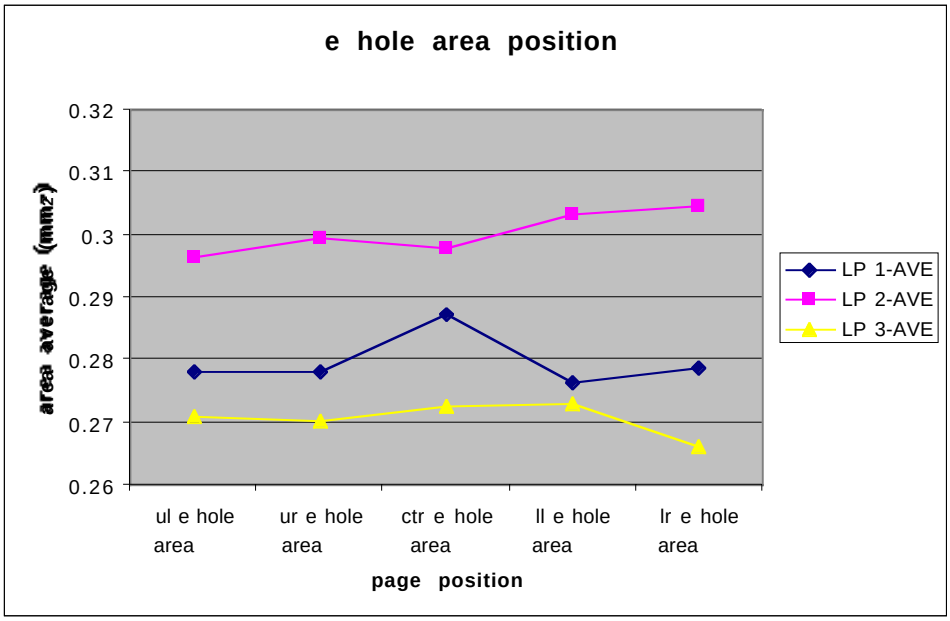

Figure 7: dependence of "e" hole area on page position

The width of a line, using the horizontal segment of the letter "e" again, also provides a metric that can have significant variation from printer to printer. However, it should be noted that, at least for the examples here, there was less significant variation related to the position on the page. This would make it slightly more difficult to use as a fingerprint feature due to the lack of correlation to position-specific anomalies.

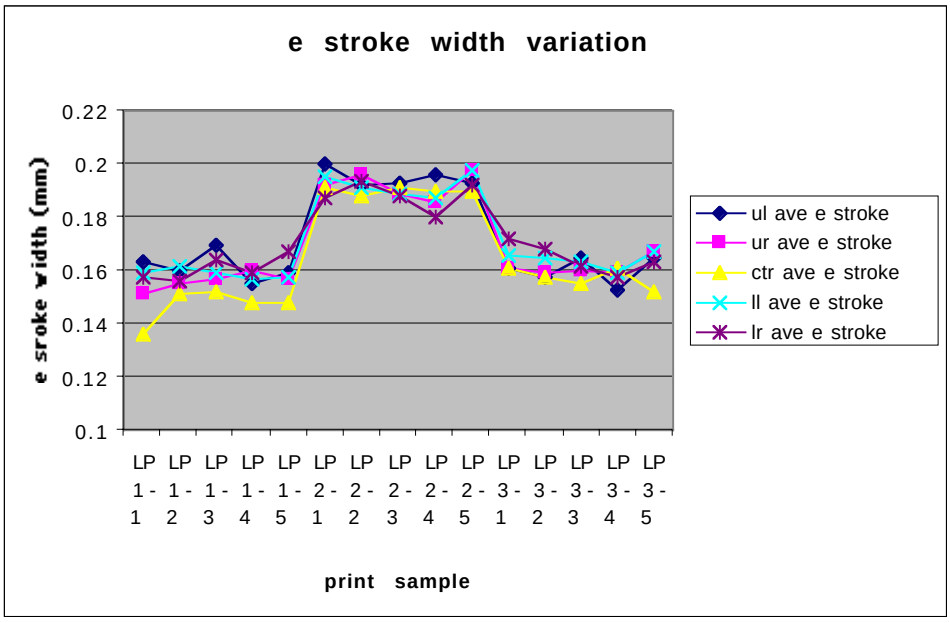

Figure 8: horizontal line width vs. printer

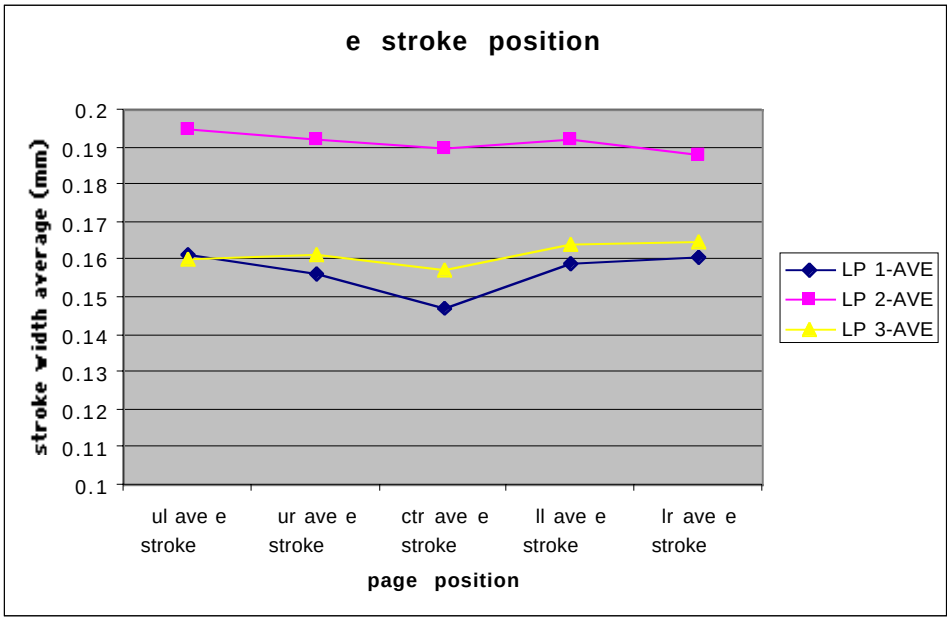

Figure 9: variation of horizontal line width vs. position on page

Conclusions

It is clear that some of the metrics chosen indicate measurable and reproducible differences between printers that might be useful in identification of a specific machine as the source for a particular document. Further work incorporating additional metrics on a larger, more statistically more significant sample of printers could yield dramatic advances in establishing the definition of fingerprints for printers. Accumulating a database, independently or through the cooperation of the manufacturers, would accelerate the process. While the results presented here are encouraging, they should be taken as a preliminary step in a potentially promising direction for the forensic examination of documents.

For more information, contact ImageXpert at info@imagexpert.com or visit our website at www.imagexpert.com .