# Document Verification and Traceability Through Image Quality Analysis

*David Wolin*
*ImageXpert Inc.*
*Nashua, New Hampshire*

## Abstract

The age of digital printing has complicated the task of forensic document examination in two significant ways. In the case of document falsification and verification, the widespread availability of sophisticated hardware and software tools (and knowledge to use them) for imaging, manipulation, and duplication has made it difficult if not impossible to detect counterfeiting without equally powerful equipment and know-how. Where a document has been identified as a fraud, or where the pedigree of a valid one is required, traceability to the printing source is critical for successful prosecution. The equivalent of a fingerprint for the printer must be identified, and matched to the source from either a database or the suspect device.

Objective, quantifiable image quality analysis provides a means to achieve these objectives. Powerful image analysis algorithms combined with flexible imaging hardware and instrumentation offer the tools required to analyze documents and establish defensible *curricula vitae*.

## Introduction

Document examination and verification have always been important in the fields of justice and commerce. Historically, if a document existed, an imitation could (and would) be created, that would pass all but the most intense scrutiny by trained inspectors using the current state of the art technology. As improvements in copying and printing have been introduced to the legitimate user market, they have been adopted and adapted quickly and efficiently, increasing the sophistication of the false product and the difficulty in identifying it as such. Unfortunately, as in many fields, many of the early adopters are less than legitimate, and those responsible for countermeasures have to catch up, reacting to the existing situation instead of proactively staying ahead.

The converse situation, identifying the actual source of a document, real or counterfeit, also suffers from the advancement and proliferation of imaging technology. Here the problem can be divided into two distinct issues: specifying the technology that was used to create the document; and identifying the actual piece of equipment that produced it. While the evolution in the different printing techniques has resulted in readily identifiable features allowing relatively easy classification, actual correlation of a document with its source is made more difficult due to the plethora of similar products available. The task is further complicated by the interchangeability and replaceability of printer components and consumables.

## Digital Printing Technologies

There are three major categories of digital printing, each with several variations and multiple manufacturers. These are inkjet, electrophotographic, and thermal. For each of these, the quality of the printing is increasing and the cost of using them is decreasing. The ultimate goal is the quality, speed and cost of traditional analog processes, such as offset or lithography.

Perhaps the most affordable and available, and therefore the most ubiquitous, is inkjet. For low volume, high quality output there are two primary technologies, known collectively as drop on demand. They are thermal and piezo, named for the mechanism used to produce the ink drops. There are many manufacturers worldwide, producing models for both global and local consumption. In the quest for customers, these manufacturers are pushing the performance of this traditionally low-end technology to levels that compete with higher quality imaging systems. Part of this improvement in image quality is being achieved through the use of higher technology consumables.

In the field of electrophotographic (EP) printing, two of the major techniques are laser and LED. Because of their higher cost (and speed) these are predominantly business machines. The image quality is usually better than acceptable while printing on quite a variety of inexpensive and unsophisticated substrates. And while traditionally black and white, these technologies are rapidly implementing color printing as a standard capability.

Thermal transfer printing offers a range of options. At the low end, thermal wax printers are a low cost, low resolution method. High quality thermal printing uses a technology called dye diffusion. This creates continuous tone images of photographic quality using special substrates. The respective markets for these has been primarily labels and marking at the low end, and digital photo printing for the high end.

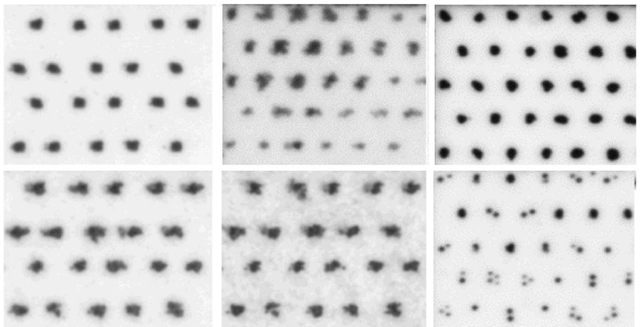

*Figure 1. Variation in dot quality*

## Identification of Printer Technology

In spite of the convergence in image quality between the different digital printing technologies, it is still fairly easy to differentiate their output using optical techniques. The chemistry and delivery of the marking media, the sophistication of the substrate, and the technology of the print engine combine to produce hardcopy output with unique features that enable identification of the method used to create it. Examination of certain text characters, straight features like lines, and solidly filled print areas can yield the data critical to the determination of printing technology, manufacturer, and model.

There are several characteristic signatures of inkjet printers. The mechanisms of drop production and projection result in specific features that are readily attributable to this process. Variations in the detailed measurements of these features are indicators pointing to individual makes, and even models, of printer. Ink chemistry and the fluid dynamics of the interaction with the inkjet nozzle combine to produce not only dots of characteristic size and shape, but also in most cases a small "satellite" or trailing droplet. Since this satellite dot always follows the main intentionally formed drop, inkjet prints show a distinct difference between the leading and trailing edge of a character. Existence of satellites confirms that inkjet was the method used, and analysis of the size, shape and distribution of these satellites indicates the make and model of the printer

The resolution, or more accurately the addressability, of an inkjet printer is also an indicator of its pedigree. There are two classes of printers, one defined in multiples of 300 dots per inch (dpi) and the other in multiple of 360 dpi. As a rule, manufacturers choose one of these and stick to it throughout their product line. For example, a company such as Hewlett-Packard which chooses the 300 dpi standard may produce models at 600, 1200 and even 2400 dpi for different market segments. Alternatively, others including Epson choose 360 dpi and offer printers at 720, 1440 and 2880 dpi. Frequency analysis of the periodic variation in edges can identify the addressability of the printer.

A complicating factor is that in the quest for higher image quality, especially in tone reproduction, the resolution is no longer a function of the specified addressability of the printer. The "dots" are made up of "micro-dots" which are a fraction the size of the single dot. Different combinations of these micro-dots are used to produce more accurate tone scales. Additionally, some printers even produce multiple sizes of dots in order to achieve the same effect. However, here high magnification analysis of the dot size distribution, along with the shape and position, can uniquely determine the printer.

Electrophotographic printers have several unique features associated with their output which make them easy to identify. Due to the chemistry of the toner and the mechanics of the printing process, the printed foreground remains on top of the substrate, enabling documents to be printed on a wide range of stock. This results in both optical and textural characteristics. The printed areas are generally more glossy than the substrate, unless a coated medium or shiny material was used. In addition, the printed areas will extend above the surface of the substrate, which will show up clearly using surface profiling apparatus. Since the toner resides on the top surface of the paper, it is susceptible to being absent, either through the lack of proper fixation to the surface, or to subsequent removal due to abrasion. Therefore, irregular deletions in an otherwise solid printed area is a signature of the EP process.

Another feature is due to the electrostatic nature of the printing process. While the center of a character or solid area will appear darker on an inkjet print due to ink buildup, the optical density of a similar region from and EP printer decreases from the edge to the interior due to charge buildup. Therefore a profile of the reflectance across a printed feature will provide an indication of the process. In addition, there should be very little difference between the leading and trailing edges of characters because there is no satellite formation associated with EP printing. On the other hand, again due to the nature of charge distribution, there is usually a relatively isometric (spatially non-specific) pattern of toner scatter around characters. This may look similar to satellites, but with a uniform distribution surrounding the printed character. For models of inkjet printers that print bi-directionally, satellites may appear on two sides of a character and seem to be distributed in a similar way to toner scatter, but closer examination will reveal that in fact the distribution is not really uniform. There will be no satellites on the sides of a character orthogonal to the direction of printhead travel (in other words parallel to the paper transport). Additionally, there will be a background of very small toner particles over the entire unprinted region due to the difficulty of the physical nature of toner, and the difficulty in eliminating them from the EP printing process. No such background noise is produced in inkjet printing, providing yet another discriminant.

Both inkjet and EP prints can have periodic defects, though the cause and therefore their characteristics are considerably different. Missing or misdirected jets, as well as misaligned printheads and paper advance errors in inkjet printers can produce periodic linear defects known as banding or streaking. These tend to extend most if not all of the way across the printout. EP printers are also subject to periodic defects due to the fact that the process involves cylindrical components whose circumference is generally a fraction of the length of the print. In contrast to the inkjet defects, these artifacts can be (and usually are) smaller in extent, being caused by latent or induced damage or defects in the cylinder. Frequency analysis of any such periodic defects can identify the technology, make and model because all components have specific mechanical dimensions which can be correlated to such frequencies. In fact, such defects can also form the basis for the actual signature or fingerprint of a specific printer, in much the same way that mechanical typewriters used to be identified.

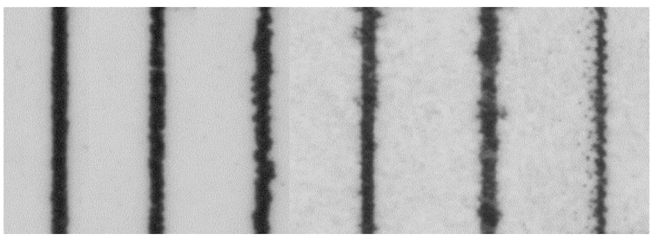

*Figure 2. Variation in line quality*

## Difficulties in Differentiation

The proliferation of printers produces two types of challenge to classification and identification. Improvements in printer and media technology are leading to convergence in image quality at the high end, both between different manufacturers of the same type of printer, and between the different types of printers. This means any differences are harder to find, and therefore require more sophisticated equipment and analysis.

On the other hand, the sheer number of different printers available makes correlation a daunting task in the absence of the actual source printer. The variety of different printers available is constantly increasing due to three factors. First, the quality and reliability of the products is improving, leading to longer useful lifetimes. Second, third party manufacturers are extending these lifetimes even more by offering replacement parts and consumables after the original manufacturer has abandoned a model. Third, new models are constantly being developed and produced in the quest for market share and differentiation.

The challenge is to assemble a database, which includes relevant characterization results for all printers produced. Variation between production lots and differences due to distributed manufacturing must be taken into account. This is further complicated by the potential variation caused by replacement and substitution of consumables and components, such as ink and toner cartridges, and receiving substrates. Even the replacement of a printer cartridge can change the appearance of a print and the fingerprint of the printer. The original manufacturer can have variation in quality from lot to lot or between locations, and aftermarket suppliers may produce substitutes with very different characteristics.

The contribution of the marking materials and the receiving substrates to the overall image quality of digital printing has been increasing recently. This means that many of the characteristics that would normally be used to distinguish and identify printers, such as dot size and shape, edge profile, and print darkness or density, are dependent on the particular combination of printer, paper and ink or toner used to create a specific document. Chemical analysis should be done on suspect documents in order to compare results with the appropriate combin-ations. This is necessary whether comparing to archived results, or reproducing a document with a suspect source printer (or type of printer).

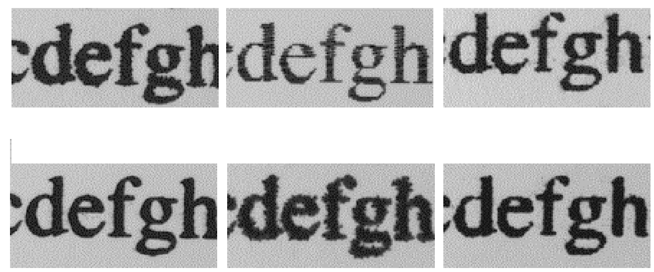


*Figure 3. Variation in text quality*

## Imaging and Analysis Techniques

There are two steps to analysing a suspect document. The first is to acquire an image with the appropriate instrumentation. The second is to apply mathematical analysis to that image. Depending on whether there are any security features and what they are (present in an original, missing or modified in a counterfeit), different optical technology is useful.

The basic requirement is to image the document at the correct resolution, providing a large enough field of view to include features of interest, and high enough magnification to allow adequate inspection of those features. Since color can be a discriminant, both from its spectral content and its print quality artifacts, adequate color separation and analysis is required. Cameras with spectral filters for the imaging elements, or bandpass filters for the illumination, along with color corrected lenses, are mandatory for accurate characterisation of color dependent print quality. Spectrophotometers are necessary for positive identification of colorant (dye and pigment) sets.

When more sophisticated technology is used to prevent duplication, such as the inclusion of markers, then the imaging system must include the appropriate instrumentation. Extensions of the spectral band into the infrared and ultraviolet can be implemented into either the detection (camera, spectrophotometer) or illumination (light source, monochrometer) function. Embedded tracers and watermarks (digital or analog) can be examined using transmission instead of reflection.

Since digital images are composed of dots and lines, analysis should focus on the features and artifacts that are directly related to these print components. The most common things printed on documents are characters, which provide lines and solid areas. Measurements can be made of line and edge quality, ancillary marks (ink satellites and toner scatter), background, and solid area coverage and density. Solid areas also provide opportunities for analysis of color as a discriminant. Other fairly common print features are graphical or shaded areas, which provide regions where individual dots are printed and can be analysed for size, shape, morphology and position.

## Instrumentation Configurations

The forensic environment and requirements will determine the necessary and appropriate equipment configuration. Currently available products enable the application of any combination of technologies to the forensic examination of documents. Starting with the analysis software, image processing and analysis algorithms are available with the power to extract all the relevant information from an image. The differences between various options are generally the ease of use and the flexibility. Unfortunately these can be somewhat mutually exclusive. While graphical user interfaces are standard, there is still a tradeoff between the constraints of limited preprogrammed tests vs. the open architecture of user definable analysis.

Incorporation of the wide variety of instrumentation hardware into and integrated analytical tool requires extension of the software image processing to include capabilities for image acquisition from cameras, data acquisition from other instrumentation, and robotic control to automate the examination of the documents in question. The appropriate type of software application is machine vision, which incorporates all of these elements and was specifically developed for automation of inspection.

Once the software engine has been chosen, a hardware configuration appropriate to the specific forensic environment must be defined. The range of options available to the end user allows selection of a complete system tailored to the needs of the moment, with the flexibility to adapt and upgrade as requirements evolve.

For field work, where portability is essential, compact imaging systems attached to laptop computers running preset or adaptable test sets are the optimal solution. In a laboratory or office situation, the number of similar samples is usually small and maximum flexibility in test definition is desirable. Here, a fixed optical bench with the capability to integrate multiple optical configurations comprised of various cameras, lenses and light sources, as well as additional instrumentation such as spectro-photometers and lasers would be the system of choice. Where larger volumes of similar documents are being examined optically, a scanner-based system with an automatic document feeder would be most efficient. For maximum capability, a full motion system that can robotically incorporate any combination of optical and other instrumentation, and which can be programmed to run automatically or used manually, is the answer.

## Summary and Conclusions

The advancement and proliferation of digital printing technology has created challenges to tracing and identifying the sources of suspect documents. Even in cases where the alleged source is known and accessible, variation in components and consumables can complicate the task of incontrovertible correlation and identification.

While it is clear that attempts to produce and use illicit documents will continue, and in fact get more sophisticated as printing technology advances, the tools exist to counter this activity. In addition, tracing the origin of documents, both counterfeit and legitimate, is made possible using these same tools. It is in the best interest of those in the field of forensic document examination to take advantage of the available knowledge and expertise currently utilised in the development and production of the very printing technology used to create these documents. There is no reason that the legitimate community can’t take the lead in the application of image quality analysis to thwart those efforts to use state of the art printing technology for purposes deleterious to the wellbeing of our society.

## Biography

Dave Wolin is Vice President of Business Development at ImageXpert Inc. After receiving his Bachelor's degree in Physics from Cornell University, he has spent more than twenty years working in the field of imaging. He has been involved in the development and production of imaging sensors and systems for a variety of applications. His work at ImageXpert has included metric development for image quality analysis of printers and media.