# Searching for discriminating features in B&W inkjet prints to individualize printers in a forensic setting

***Koen Herlaar (k.herlaar@nfi.minjus.nl), Jan de Koeijer, Mignonne Fakkel-Slothouwer, Henk Madhuizen; Netherlands Forensic Institute (NFI); The Hague, The Netherlands***


## Abstract

*In recent years several authors have demonstrated that it is possible to identify a certain printer model by the use of image texture analysis. However, these articles often describe only a limited number of printers (typical 10) and non realistic test targets.*

*In this study we search for discriminating features in B&W inkjet prints, not only to be able to exclude a specific printer as a possible source, but also for individualizing printers. The scope of the study is anonymous letters and forged (financial) documents. The study is performed using a realistic test document, which only incorporates text and symbols. This document is printed on the same printing paper on more than 75 different printers and on three sets of 10 identical printers, four samples on each printer. The samples will be analyzed with the use of the ImageXpert Full Motion system at more than 500 different locations on the document, measuring more than 50 different features, such as line spacing, edge quality, line thickness, area and overspray.*

*Advanced data analysis will be performed on the measured data to select the best (combination of) discriminating features.*

*This article will describe the scope and goal of the study, the design of the test document and the research strategy. Furthermore, the first preliminary results will be shown.*


## Introduction

### *Project background*

The last two decades inkjet printers became very common both in small and home offices and in households. As a result also, inkjet technology is used to counterfeit currency, identity documents, product labels, and so on. Furthermore, it is often used to produce documents such as anonymous letters (threatening letters), wills, contracts and other business associated documents. Therefore, the demand for forensic examination of documents produced by inkjet printers has evolved significantly.

This research project is focused around these documents and especially around anonymous letters. In some cases the threat posed in these documents is severe and urgent. In the first phase of such an investigation, any leads towards a possible suspect which can be found by the forensic document examiner are valuable. However, anonymous letters produced by an inkjet printer are usually very hard to trace back to a certain make and model of a printer.

Furthermore, in a later phase of the investigation when a suspect has been arrested, it is the job of the forensic document examiner to make an objective statement about the possible link between the anonymous letter and the seized inkjet printer of the suspect. This study is usually done by making *known* sample prints with the seized printer and by comparing features of the *known* sample prints with features of the *questioned* anonymous letter. Among others, features which can be compared are the inks, printhead dimensions, print swath directions and printing defects such as blocked nozzles or deflected droplets.

### *Evaluating forensic evidence*

In evaluating the forensic evidence of the questioned letter and the seized printer, two important questions need to be answered by the forensic document examiner:

What is the probability of the observed similarities and/or differences in the studied features, if the questioned letter is produced by this printer?

What is the probability of the observed similarities and/or differences in the studied features, if the questioned letter is produced by another random inkjet printer?

The ratio between these two probabilities (the odds) is called the Likelyhood Ratio (LR) and this is a measure for the strength of the evidence. The higher the LR, the stronger the evidence for the hypothesis that the questioned letter is produced by the seized printer.

However, currently it is only possible to reach a high LR (and therefore strong evidence) if there is a similarity in a characteristic printing defect between the *questioned* letter and the *known* sample prints, such as a blocked nozzle. The probability to find this similarity if the questioned document is produced by the seized printer is normally very high (close to 1), depending on the time difference between printing. However, this probability is very small if the questioned document is produced by another random inkjet printer, since it is quite unlikely to find exactly the same blocked nozzle. Hence, the ratio between these probabilities (the LR) is high and therefore the evidence is strong.

In the ‘forensic language’ such a feature is called an individual characteristic, in contrast to class characteristics. Examples of class characteristics are ink and printhead dimensions and similarities in these characteristics usually only lead to weak evidence. It is possible to combine similarities in different, independent, class characteristics which can lead to stronger evidence.

Unfortunately for the forensic document examiner, very few defects are found in questioned documents and therefore it is necessary to find new individual characteristics.

### *Goal of this research project*

Since individual characteristics are rarely found in studying inkjet printed questioned documents, this research project has the goal to find new characteristics in features of printed text of inkjet printed questioned documents. These features should be objective and qualitative. Furthermore, the LR's of these new characteristics should be determined in order to evaluate the strength of the evidence. There has been research in the past on this topic, however this was limited to a small number of printers and often specially designed testprints to study printing quality [1] – [5].

### *Project description*

This research project is performed in five different phases:

A. Become an expert in the ImageXpert Full Motion System
B. Design a testprint and collect printsamples
C. Collect measurement data and first analysis of this data
D. With the results, perform more specific analysis necessary to calculate LR's of the new characteristics
E. Evaluation of the project

Currently the project is running in phase C. This article will describe the used ImageXpert system, the design of the testprint and the strategy used in collecting the printsamples, the measurements performed and the first preliminary results.

## The used ImageXpert system

The ImageXpert Full Motion System is used in this research project. Our set-up consists of two digital B&W camera's with different optical systems above a computer (iMac) controlled X-Y motion table. With the ImageXpert software it is possible to control the X-Y motion table and to create automated measurements tasks for printsamples. The measurements which are performed are based upon image quality measurements.

The two B&W camera's have different fields of view (FOV) and different measurement resolutions. The large FOV camera (~ 30 x 25 cm) has a measurement resolution of 12.1 µm/pixel (~2100 dpi). The small FOV camera (~ 3.8 x 2.8 cm) has a measurement resolution of 3.7 µm/pixel (~ 6900 dpi).

Furthermore, the system is equipped with a calibration target plate. Measurements on this calibration target have shown that the ImageXpert system is stable and accurate over a long period of time [6].

## The design of the testprint

The design of the testprint needed to fulfill the following requirements:

- It should represent an anonymous letter, therefore text only: preliminary research showed that the 'modus' of a printer is influenced by the contents of the testprint. In this context 'modus' should be interpreted as the way a printhead moves across the document and fires its nozzles. In other words, if for instance one graphical element (such as a line) is part of the contents of a testprint, a printhead can move differently and can fire its nozzles differently than for a text only testprint.

- Different (textual) elements, each distributed over the entire page: depending on the measured characteristic, it is possible that this is influenced by the position of this characteristic on the printed page. In order to be able to study this influence, textual elements should be randomly distributed over the entire page. Therefore, the testprint consists of 35 lines of text.

- Different fonts and font sizes: ideally, the new characteristics should be font and font size independent. In order to be able to study this influence, three different fonts and font sizes are used on the testprint. The fonts used are the most commonly found in forensic cases in the Netherlands; Times New Roman, Arial and Calibri.

- Different characteristic (textual) marks: in order to automate the measurements of the different printsamples, it is important that the measurement system can locate certain characteristic marks on each document. Subsequently, the system uses these location for the correct automatic positioning of the X-Y motion table.

- interchangeable, fixed file format: the digital file, which is the input for the printing process, should be identical for each printer. Preliminary tests showed that a PDF document with embedded fonts is the best choice. This format does not change the 'modus' of a printer compared to an identical MS Word input file, and is therefore much more stable across different platforms and different software versions.

Due to the limited space, the testprint can not clearly be shown in this article. One can request a copy of the digital PDF file through the author.

## Collecting printsamples

Since the number of variables which can influence the result of a print is very large, we have chosen to keep at least one variable fixed: the substrate. We have distributed the same type of printing paper, general multipurpose printing paper, among more than 120 colleagues within our institute with the request to produce four printsamples with their printer at home. Furthermore, a questionnaire was added together with a small manual how to print the printsamples. The digital file was distributed via email.

The questionnaire requested information about the make and model of the printer, the operating system, the printer driver and the cartridges and inks used. In total we received printsamples from 106 colleagues, existing of 68 different inkjet printer models (and 18 laser/LED printer models). The 68 different inkjet models exist of 41 HP, 27 Canon, 11 Epson, 3 Dell, 3 Brother and 3 Lexmark printers. 77 different combinations of printer model/used cartridge were found, since in some cases the same make and model of a printer is used with different make and model cartridges.

With this collection of printsamples, the basic collection to study the 'between variation' between printers of different make and model is sufficient. However, we still need a collection of printsamples from a large number of different printers from an identical make and model to be able to study the 'within variation' of characteristics. For this purpose two specific printer models will be purchased in bulk.

## Measurements performed

An automatic measuring program has been written in ImageXpert to measure different features in the text of the collected printsamples. The most important features which are measured on different letters and symbols are summarized in Table 1. Figure 1 to Figure 5 show examples of these letters and symbols. ISO 13660 [7] is taken as guideline for measuring the features. However, most features are measured at multiple ways.

Table 1 also lists on how many different locations throughout one printsample letters and symbols are measured. These locations are spread throughout the printsample. In total, measurements are done on 2357 different locations on one printsample, measuring multiple features on one location. Measuring one printsample takes 32 minutes.

**Table 1. Summary of measurements on printsamples**

| Which letter/symbol ? | What is measured ? | Comments |
|---|---|---|
| All letters and symbols (1834 locations) *(Figure 1)* | The top and bottom points of each character | This data can be used to measure the line to line distance, which is related to the paper transport mechanism of a printer |
| Squared brackets (35 left ‘ [ ‘ and 35 right ‘ ] ’ ) *(Figure 2)* | Line quality, line thickness, distance from left bracket to right bracket, … | Each line starts and ends with a squared bracket. The distance between these brackets is related to the movement of the printhead. |
| Dots (on i, j, end sentence) (205 locations) *(Figure 3)* | Area, radius, circumference, number of satellites, average grey value, … | Every text has dots. Therefore, if a characteristic feature is found in dots, this will be very valuable |
| Letter e (130 locations) *(Figure 4)* | (hole) area, (hole) circum-ference, number of satellites, average grey value, ... | The letter e is also very common in text. A discriminating feature will be very valuable |
| Letter n (115 locations) *(Figure 5)* | Area, line quality, line thickness, circumference, number of satellites, average grey value, … | The letter n is also very common. Another benefit of the letter n is the straight legs, on which line quality can be measured. |
| Corners of paper (3 locations) | The bottom left, top left and top right corner of the paper | The corners of the paper are measured to capture the orientation of the paper on the ImageXpert system |

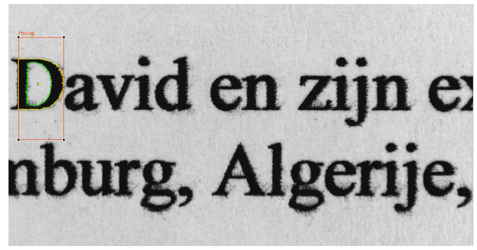


***Figure 1.*** *WFOV camera; top and bottom point measurements*

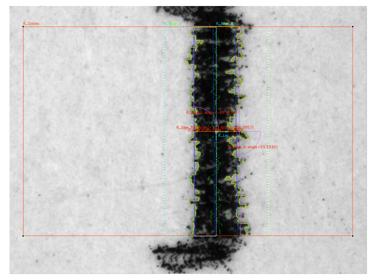

***Figure 2.*** *SFOV camera; measurements on (right) squared bracket*

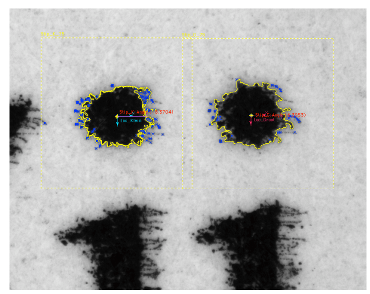

***Figure 3.*** *SFOV camera; measurements on dots*

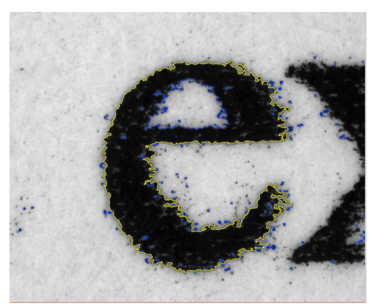

***Figure 4.*** *SFOV camera; measurements on the letter e*

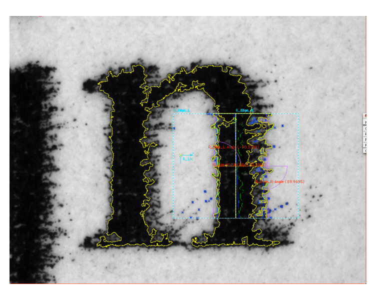

***Figure 5.*** *SFOV camera; measurements on the letter n*

## First preliminary results

Although no advanced data analysis has been done yet, already some remarkable characteristics are possibly found. The measurements from the top and bottom points lead to line to line distances. These line to line distances, related to the paper transport mechanisms of a printer, seem to be characteristic. Furthermore, features measured from the dots seem promising as well.

Another example is the data measured from the squared brackets at the beginning and end of each line. Amongst others, the width of each text line is measured from the center of the left bracket to the center of the right bracket. This characteristic is related to the movement of the printhead. In the design of the testprint, the choice is made that this distance is identical for each line. Figure 6 shows data from this characteristic for 21 printsamples of different printers.

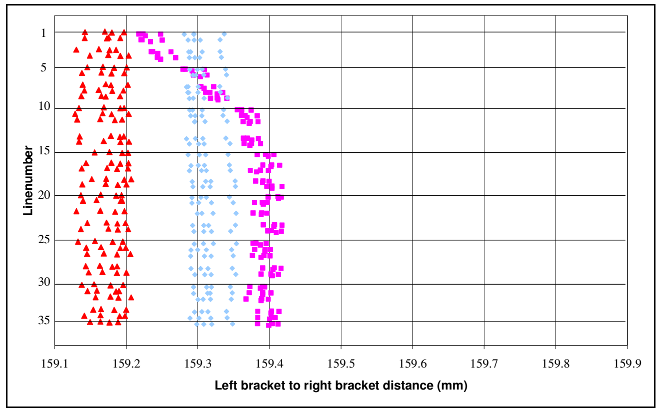


***Figure 7.*** *Width of the 35 text lines for 12 printsamples of 3 different printers.*

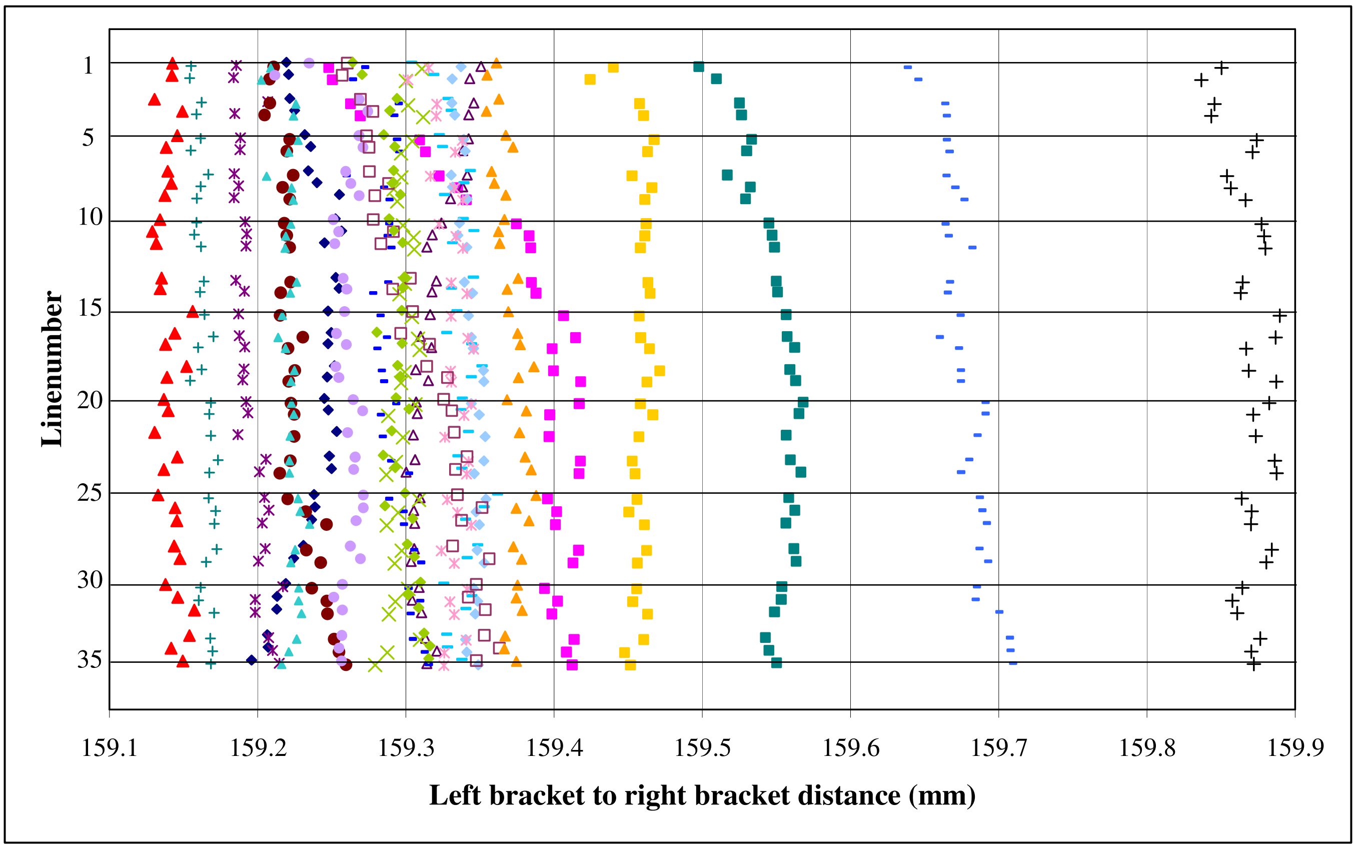


***Figure 6.*** *Width of the 35 text lines for 21 printsamples of 21 different printers.*

Figure 6 clearly shows that the data is different for all printers. Furthermore, also the shape of a datagroup ( the variation with linenumber) seems to be a characteristic of a printer. This is illustrated again in Figure 7. This figure shows data from 3 different printers, 4 printsamples of each printer. Although there is variation in the data within different printsamples of one printer, the general shape of the datagroup is constant.

## Discussion

Although this research project is still at the beginning, first preliminary results are very promising. However, a lot of work still needs to be done. Before a certain characteristic can be used within the forensic setting, several variations need to be addressed. The most important variations will be discussed.

The first variation is the variation of a characteristic between different printsamples of the same identical printer and between printsamples of different printers. Figure 6 and 7 have already shown an example of this variation. In order to find a 'successful' characteristic, the variation within printsamples of the same printer needs to be small compared to the variation between printsamples of different printers. Furthermore, an comparison algorithm has to be developed to calculate comparison scores between a *known* characteristic and a *questioned* characteristic.

The second variation which needs be studied is the variation of a characteristic with time. In other words, the stability of the characteristic with time needs to be addressed. Therefore, a limited number of colleagues will be requested to print new printsamples over a certain period of time.

The third variation is the variation of a characteristic with the location of this characteristic on the page of the printsamples. This variation is also important since it will be positioned differently in a real case than on the studied printsamples.

The next variation which needs to be addressed is the variation of the characteristic with replacement of the cartridge. Characteristics related to print quality will probably vary a lot with replacement of the cartridge. However, characteristics related to the movement of the printhead or to the paper transport mechanism will possibly vary a lot less.

The last variation discussed here, is the variation of the characteristic with the used printing paper. Most characteristics will be influenced by the printing paper used. However, in comparing a printer with a questioned document, it is often possible to print 'testlines' on blank parts of the questioned document. In that way, one can eliminate the influence of varying paper. In this phase of the research project, we are using only one type of paper. At a later stage, variations in printing paper will be studied.

## Conclusion

By taking a different approach in the design of the testprint and in collecting printsamples, a search is started for new (individual) characteristics in inkjet printed, text only, documents. Although lots of questions still need to be addressed, first results are promising. Especially the results from measurements on the width of text lines are interesting. Especially since this characteristic is a distance measurement, related to the movement of the printhead. Possibly, the variation of this characteristic with varying paper and/or cartridge is small. Depending on the used printer, we have already seen that there can be a relative large variation of this characteristic with the line number of the printsample. However, this variation is not necessarily a bad thing since it is a stable variation across multiple printsamples. Therefore, this variation could possibly be used as a characteristic itself.

The research project will continue to run for at least one more year.

## References


[1] Fleming, P.D., Cawthome, J.E., Mehta, F, Halwawala, S., and Joyce, M.K., "Interpretation of Dot Fidelity of InkJet Dots Based on Image Analysis", Jour. Imaging Sci. and Tech., Vol. 47, No. 5, pp. 394-399 (2003).

[2] Oliver, J.O. and Chen, J., "Use of Signature Analysis to Discriminate Digital Printing Technologies", 18th NIP, San Diego, CA (2002).

[3] Parker, J.L.,"An Instance of InkJet Printer Identification", Jour. of ASQDE, Vol. 5, No. 1, pp. 5-10 (2002)

[4] Wolin, D., "Document Verification and Traceability Through Image Quality Analysis. 18th NIP, San Diego, CA (2002).

[5] LaPorte, G.M., "Modern Approaches to the Forensic Analysis of InkJet Printing - Physical and Chemical Examinations", Jour. of ASQDE, pg. 22-36 (2004).

[6] Eberhardt, T., Ross, J., and Virgin, G., "Validating the ImageXpert™ Full Motion System: A Foundation to Support Future Research into Document Authentication", 59th Annual Meeting of the AAFS, San Antonio, Texas (2007).

[7] ISO/IEC 13660, Information technology, Office equipment, "Measurement of image quality attributes for hardcopy output, Binary monochrome text and graphic images", 1st edition (2002)


## Author Biography


*Koen Herlaar received his MS degree in Applied Physics in 2002 from the Technical University of Delft, specializing in Optics. Currently he is working as a forensic scientist at the Netherlands Forensic Institute (NFI) in The Hague, in the field of questioned document examination. His focus is on R&D projects related to optics, imaging and image analysis.*

*The NFI uses state-of-the-art technology and science to provide high-quality forensic services and is one of the world's leading forensic laboratories.*