Intrinsic and Extrinsic Signatures for Information Hiding and Secure Printing with Electrophotographic Devices

Gazi N. Ali, Aravind K. Mikkilineni, Jan P. Allebach, and Edward J. Delp
School of Electrical and Computer Engineering, Purdue University
West Lafayette, Indiana

Pei-Ju Chiang and George T. Chiu
School of Mechanical Engineering, Purdue University
West Lafayette, Indiana

Abstract

There are a number of applications in which it is desirable to be able to identify the technology, manufacturer, model, or even specific unit that was used to print a given document. Also, it may be desirable to embed additional information in the printed document, such as the date and time that it was printed. In previous work with laser electrophotographic printers, we developed strategies to reduce quasiperiodic banding artifacts that are characteristic of the specific print mechanism. The dominant process-direction spatial frequencies associated with these artifacts are determined by the parameters of the gear train in the print mechanism. Here we take a different view of these artifacts, treating them as a signature of the printer that can be identified by appropriate image analysis techniques. We refer to the characteristics of the native device as its intrinsic signature. By using the same strategies employed earlier to reduce the banding artifacts, we can amplify and modulate the banding to embed additional information, as discussed above. We refer to this as an extrinsic signature. In this paper, we describe our on-going research in this area. We present preliminary results from our effort to identify intrinsic printer signatures for a number of electrophotographic printers that are now on the market.

Introduction

Secure printing involves the idea that printer output, from here on called a *document*, is an effective means to identify some features of the printer. These features, which are printer specific, can be used for document security. For example, in case of a suspected forgery we ideally should be able to tell what type of printer was used to create the document. In this project the focus is on extracting features from the document and developing methods to securely hide information which can later be embedded in printer firmware.

We propose to develop two strategies for printer identification. The first strategy is passive. It involves characterizing the printer and finding intrinsic features in the printed output that are characteristic of that particular printer, model, or manufacturers products. We call this the *intrinsic signature*. Detecting the intrinsic signature requires an understanding and modelling of the printer mechanism. Development of image analysis tools for intrinsic signature detection from documents with arbitrary content can be realized by design based upon standard test patterns. The intrinsic signature is detected by scanning the printed pages with a high resolution drum scanner and applying low-level image analysis routines to extract features. These features are processed with a soft classifier to yield likelihoods at each level of a decision tree. The decision tree tells us what class a document belongs to, where the classes in this case are based on the printer that created it.

The second strategy is active. Here we embed an *extrinsic signature* in every printed page. Information which is hidden in the document can be thought of as a watermark. There are two basic types of watermarks which can be employed. First, there are robust watermarks, such as a logo printed at the top of a page or overlaid on top of the document. These watermarks are able to survive many attacks such as copying and are not easily removed. A more sophisticated robust watermarking scheme for text documents is developed in a paper by Brassil[1] where alternate text lines are shifted slightly up or down to encode bits of information.

Robust watermarks are useful for encoding data to trace a document to a particular person, but they cannot be used to determine whether a document has been altered or if it is the original (legal copy). Fragile watermarks can be used for these purposes since they are easily destroyed and can be used to determine if any part of the document was altered.

Currently only standard watermarking techniques have been used to encode information into documents. Techniques, such as those shown by Voloshynovskiy[2] among others, rely on designing a watermark which can survive the printing process. In contrast, we wish to design the watermark by modifying the print process itself by modulating the process parameters in the printer mechanism to encode identifying information, such as the printer serial number and date of printing, in every printed page. To detect the extrinsic signature, we again scan the printed pages, and process them using image analysis techniques; but in this case, our goal is to decode the signature to extract the information embedded in it. Development of the methodology for extrinsic signature embedding will build directly on our work with intrinsic signatures. We will use our knowledge of the printer mechanism models and the results of the printer characterization to determine the printer process parameters that can be modulated to encode the desired identifying information. The modulation of these parameters will require modification to the actual printer mechanism.

In this paper we briefly describe our laboratory setup for printer characterization. We analyze the banding characteristics of four different printer models using two different test patterns. Our initial results show that by using large gray patches we can get a strong banding signal which is printer dependent.

Test Bed for Printer Analysis

We have set up a laboratory facility to analyze different characteristics of a printed page. This setup includes various printers, scanners, an image analysis system and image processing software. The *printer bank* has nine different models of monochrome electrophotographic (EP) printers. There are two printers for each model to verify test results across the same model. Table 1 provides the model numbers and specifications for the printers in the printer bank. These printers were chosen because

1. They are available in most consumer electronics retail outlets.
2. They are in the low-end of the price range for monochrome EP printers.
3. They have similar speed and resolution.

All but one model of the printers are connected to a single host computer via a USB hub. Only the LaserJet 4050 printers are accessed via the network using TCP/IP. Figure 1 shows the printer bank.

Figure 1. Laboratory setup for printer characterization

Table 1. Printers installed in the printer bank*

Maker	Printer Model	Writing Method	Resolution (dpi)	Speed (ppm)
Brother	HL-1440	laser	1200	15
HP	LJ 1000	laser	600	10
	LJ 1200	laser	600	15
	LJ 4050	laser	600	17
Lexmark	E320	laser	1200	16
Minolta	Pagepro 1250w	laser	1200	17
Okidata	Okipage 14e	LED	600	14
Samsung	ML-1430	laser	600	12
	ML-1450	Laser	1200	14

*There are two identical units for each different model.

The characterization process has a logical flow that starts with the test page design and ends with image processing, as shown in Figure 2. The output from the printer is scanned at high resolution to get maximum information for image processing. We have installed four different scanners (a) Saphir Ultra2 (1200 dpi)[f1], (b) HP Scanjet 4570C (2400 dpi)[f2], (c) HP Scanjet 8250 (4800 dpi)[f3], and (d) AZTEK Premier (8000 dpi)[f4]. The AZTEK Premier is a drum scanner, and is capable of providing high resolution scans for image analysis. The other scanners are used when low or medium resolution scans are sufficient. A new addition to the lab is the QEA IAS 1000 system[f5]. This system can provide a resolution of apprximately 4230 dpi with the bellows position set to minimum and focal ratio 1:1.[3] The IAS 1000 system has already been used for print quality assessment in electrophotography.[4]

f1Heidelberg USA, Inc., Kennesaw, GA 30144
f2Hewlett-Packard Company, Palo Alto, CA 94304-1185
f3Hewlett-Packard Company, Palo Alto, CA 94304-1185
f4Aztek, Inc., Lake Forest, CA 92630
f5Quality Engineering Associates, Inc., Burlington, MA 01803

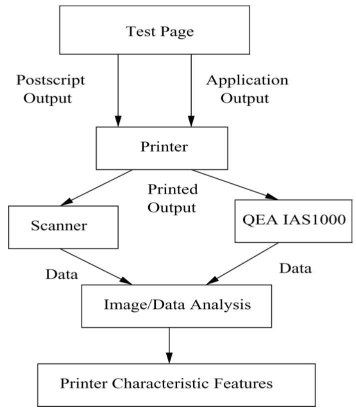


Figure 2. Steps followed in printer characterization

Overview of Intrinsic Features

As shown in Figure 3, the EP process can be divided into six main steps, namely charging, exposure, development, transferring, fusing, and cleaning. Due to electromechanical fluctuations or imperfections, we have print quality defects. The intrinsic features that characterize the printer are closely related to printer defects or print quality defects. These artifacts can be treated as a signature of the printer that can be identified by image analysis techniques. These artifacts are the result of wobble or imperfections of the polygon mirror, optical photoconductor (OPC) imperfections and speed fluctuation, and charge roller and fuser defects.

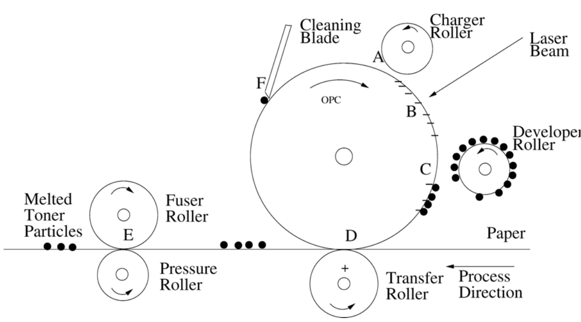
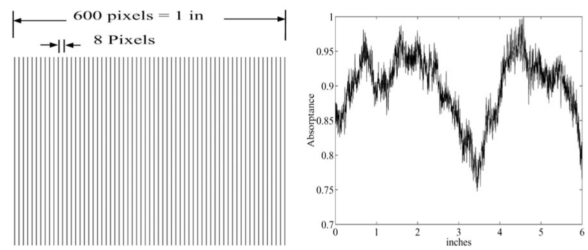


Figure 3. Electrophotographic process: cross sectional view of typical laser printer (A) Charging (B) exposure (C) development (D) transferring (E) Fusing and (F) cleaning.

Banding is an especially prominent PQ artifact for laser printers. It usually appears as nonuniform light and dark lines across a printed page perpendicular to the paper process direction (the direction in which the paper passes through the printer). One of the main causes of banding is the non uniform scan line spacing due to variations in the OPC drum velocity. It can be shown that the absorptance fluctuations are linearly related to the variation in line spacing.[5] The highly nonlinear OPC charge-to-developed toner mass characteristics amplifies the affect of non-uniform line spacing.[6] Much research has been done to characterize and reduce banding.[5,7,8] The absorptance fluctuation can be attenuated with a closed loop system in which variations in OPC drum speed are detected and used to compensate laser pulse-width,[5] or the vertical position of each scan line,[7] or to drive the motor controller to attenuate the specific banding frequencies.[8] A goal of our research is to determine how these mechanisms may be used to embed an extrinsic banding signature in the printed page.



(a) (b)

Figure 4. (a) Test pattern for 12.5% fill gray level patch and (b) 1-D horizontal projection of the printed and scanned test pattern. Each row of the test pattern consists of a repeating pattern of a one followed by seven zeros. The test pattern was scanned at 600 dpi.

Preliminary Experiments and Results

We used four different models of EP printers. We also tested two separate units for each model. The tested printers are the HP Laserjet 4050[f6], HP Laserjet 1200, Samsung ML-1450[f7], and HP Laserjet 1000. The printed outputs were scanned using the Saphir Ultra2 and the HP Scanjet 4570C scanners. Each test pattern was printed and scanned at least three times to verify consistency of the results.

We used two different test patterns. In the first pattern, vertical lines were used to create different grey level patches. Line patterns are used to avoid the influence of the halftoning algorithm used to render the test pattern. The test pattern is sent to the printer in postscript format. We used 12.5%, 25%, 33%, and 50% fill patterns. A sample test pattern and the corresponding 1-D horizontal (scan direction) projection are shown in Fig. 1. The projected signal was Fourier transformed to get the banding frequencies. In Table 2, the results for the four different printers are summarized.

f6Hewlett-Packard Company, Palo Alto, CA 94304-1185

f7Samsung Electronics America, Inc. (SEA), Ridgefield Park, NJ

Table 2. **Principal Banding Frequencies for Four Different Printer Models Generated using Postscript Files**

Printer Model	Principal Banding Frequencies (cycle/inch) for Greylevel Patches			
	12.5%	25%	33%	50%
LJ 1000	27, 69	27, 68	27, 69	27, 69
LJ 4050	51	51	51	51, 100
ML-1450	16, 100, 106	16, 32, 100, 106	16, 32, 100, 106	16, 32, 100, 106
LJ 1200	69	69	69	69

In the second test pattern, we used three different test patches. The first patch had forty five 0.25 point lines; the second patch had fifteen 0.75 point lines; and the third patch had fifteen 1.5 point lines. All the patches were 6 inches long. These patches were printed directly from an application program, i.e. Microsoft Powerpoint. The first patch should give us the strongest banding signal. The other two patches were used to check to see if we can identify the banding frequencies even with sparse lines and lower graylevel patches. These patches were also analyzed by taking the Fourier transform of the 1-D horizontally projected signal. The results for the second test pattern are summarized in Table 3.

Table 3. **Principal Banding Frequencies for Four Different Printer Models Generated using Powerpoint Files**

Printer Model	Banding Freq. From Patch 1 (cycles/inch)	Banding Freq. From Patch 2 (cycles/inch)	Banding Freq. From Patch 3 (cycles/inch)
LJ 1000	27, 69	27, 68	27, 69
LJ 4050	51	51	51, 100
ML-1450	16, 100, 106	16, 32, 100, 106	16, 32, 100, 106
LJ 1200	69	69	69

The results obtained from the two different test patterns match very closely. The banding frequencies for the print ers that have the same model number are almost identical; and the the variation of banding frequencies among different models is clearly visible. From these banding frequencies, we can uniquely identify most printer models. The LJ 1000 and LJ 1200 appear to have very similar print engines, based on the fact that they both have banding at 69 cycles/inch. On the other hand, the LJ 4050 and the ML-1450 use different print engines; and they have very different banding frequencies. Previously 50% graylevel patches were used to determine banding frequencies. Our results show that banding frequencies can be obtained even from 12.5% grey level patches.

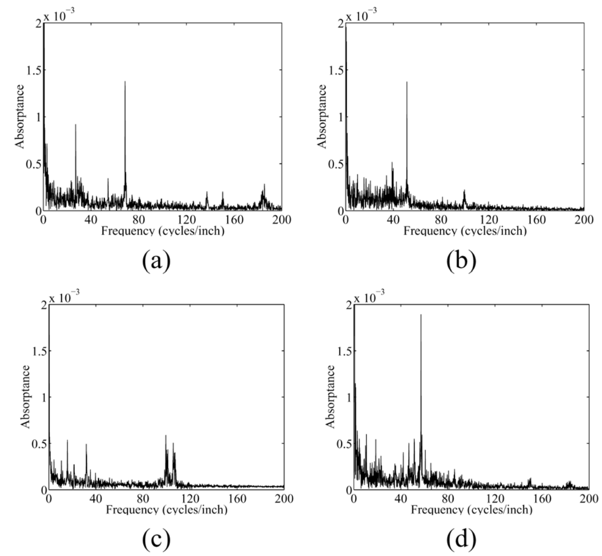


Figure 5. Spectra of projected absorptance for (a) LJ 1000, (b) LJ 4050, (c) ML-1450, and (d) LJ 1200

Conclusion

Large midtone graylevel patches provide a good banding signal that can be used as an effective tool for characterizing banding with EP printers. The banding frequency among the units of the same model of printers is stable enough to be used as an intrinsic feature. We can get reliable results with 12.5%-50% graylevel patches; but text-only documents will require more sophisticated processing to yield banding frequencies. Our future goals are to develop methods to extract the banding signature from text-only documents, explore additional intrinsic features of the printed documents, and finally to embed watermarks for document security.

References

1. J. T. Brassil, S. Low, N. F. Maxemchuk, "Copyright Protection for the Electronic Distribution of Text Documents," in *Proceedings of the IEEE,* Vol. 87, No. 7, July 1999, pp. 1181-1196.
2. S. Voloshynovskiy, F. Deguillaume, S. Pereira, T. Pun, "Optimal Adaptive Diversity Watermarking with Channel State Estimation," in *SPIE's 13th Annual Symposium,* Electronic Imaging 2001: Security and Watermarking of Multimedia Content III, SPIE Proceedings, San Jose, California, USA, 23-27 January 2001.
3. Quality Engineering Associates, Inc., "IAS-1000 Image Analysis System, Calibration Data".
4. M. Tse, David J. Forrest and King Y. She, "Use of an Automated Print Quality Evaluation System as a Failure Analysis Tool in Electrophotography," in *IS&T Eleventh Congress on Advances in Non-Impact Printing Technologies,* Nov(1995).

5. G. Lin, J. Grice, J. Allebach, G. Chiu, W. Bradburn, and J. Weaver, "Banding artifact reduction in electrophotographic printers by using pulse width modulation," in *J. Imaging. Sci. and Tech,* vol. 46, no. 4, pg. 326-337, (2002).
6. D. Kacker, T. Camis, J. P. Allebach, "Electrophotographic Process embedded in Direct Binary Search," in *IEEE Trans. On Image Processing, 11,* pg. 234-257,(2002).
7. C. Chen and G. Chiu,"Banding reduction in electrophotographic process," in *Proc. IEEE/ASME Int. Conf. Advanced Intelligent Mechatronics,* pg. 81,July (2001).
8. M. Ewe, J. Grice, G. Chiu, and J. Allebach, "Banding reduction in electrophotographic process using piezoelectric actuated laser beam deflection device," in *Proc. IS&T's NIP16,:International Conference on Digital Printing Technologies,* pg. 276,November (2000).

Biography

Gazi Naser Ali received the B.S. degree in electrical engineering from Bangladesh University of Engineering and Technology,in 1998 He is currently pursuing his Ph.D. degree in the School of Electrical and Computer Engineering, Purdue University, West Lafayette. He is a student member of IEEE and IS&T.