High-Capacity Data Hiding in Text Documents

Aravind K. Mikkilineni[‡]
George T. C. Chiu[‡], Jan P. Allebach[†], Edward J. Delp[†]

[†]School of Electrical and Computer Engineering
[‡]School of Mechanical Engineering
Purdue University, West Lafayette, Indiana USA

ABSTRACT

In today's digital world securing different forms of content is very important in terms of protecting copyright and verifying authenticity. One example is watermarking of digital audio and images. We believe that a marking scheme analogous to digital watermarking but for documents is very important. In this paper we describe the use of laser amplitude modulation in electrophotographic printers to embed information in a text document. In particular we describe an embedding and detection process which has the capability to embed 14 bits into characters that have a left vertical edge. For a typical 12 point document this translates to approximately 12000 bits per page.

Keywords: document security, secure printing, printer identification

1. INTRODUCTION

In today's digital world securing different forms of content is very important in terms of protecting copyright and verifying authenticity[1–10]. One example is watermarking of digital audio and images. We believe that a marking scheme analogous to digital watermarking but for documents is very important[1]. Printed material is a direct accessory to many criminal and terrorist acts. Examples include forgery or alteration of documents used for purposes of identity, security, or recording transactions. In addition, printed material may be used in the course of conducting illicit or terrorist activities. Examples include instruction manuals, team rosters, meeting notes, and correspondence. In both cases, the ability to identify the device or type of device used to print the material in question would provide a valuable aid for law enforcement and intelligence agencies. We also believe that average users need to be able to print secure documents, for example boarding passes and bank transactions.

There currently exist techniques to secure documents such as bank notes using paper watermarks, security fibers, holograms, or special inks[11, 12]. The problem is that the use of these security techniques can be cost prohibitive. Most of these techniques either require special equipment to embed the security features, or are simply too expensive for an average consumer. Additionally, 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.

We propose to develop two strategies for printer identification based on examining a printed document. The first strategy is passive. It involves characterizing the printer by finding intrinsic features in the printed document that are characteristic of that particular printer, model, or manufacturer's products. We refer to this as the *intrinsic signature*. The intrinsic signature requires an understanding and modeling of the printer mechanism, and the development of analysis tools for the detection of the signature in a printed page with arbitrary content.

We have previously reported techniques that use the print quality defect known as *banding* in electrophotographic (EP) printers as an intrinsic signature to identify the model and manufacturer of the printer[13–15]. We showed that different printers have different sets of *banding frequencies* which are dependent upon brand and model. By using texture features to characterize this intrinsic signature, we have been able to correctly identify the source printer of printed text documents generated by both EP and inkjet printers[16–19].

This research was supported by a grant from the National Science Foundation, under Award Number 0524540.
Address all correspondence to E. J. Delp at ace@ecn.purdue.edu

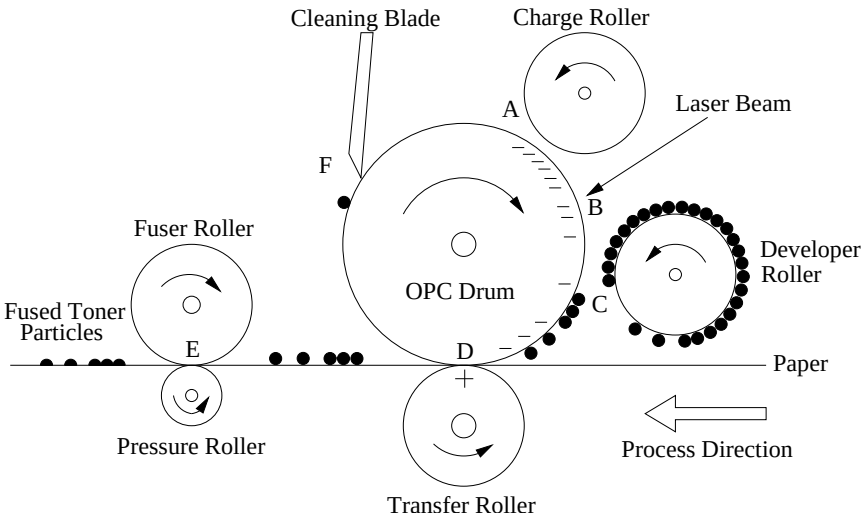

Figure 1. Diagram of the EP process: (A) charging, (B) exposure, (C) development, (D) transfer, (E) fusing, (F) cleaning

The second strategy is active. We embed an *extrinsic signature* in a printed page. This signature is generated by modulating the process parameters in the printer mechanism to encode identifying information such as the printer serial number and date of printing. To detect the extrinsic signature we use the tools developed for intrinsic signature detection. We have successfully been able to embed information into a document with electrophotographic (EP) printers by modulating the banding intrinsic feature[20].

Embedding extrinsic features into a document also requires knowledge of the specific print mechanism. During the past few years, many techniques to reduce banding artifacts have been successfully demonstrated by modulating various process parameters such as laser intensity/timing/pulse width (exposure modulation)[21], motor control[22], and laser beam steering[23]. These techniques can also be used to inject "artificial" banding signals that are not intrinsic to the printer. The technological understanding and capability to implement these techniques are not easily obtained meaning it will be difficult for someone to "hack" it.

One issue with embedding extrinsic signatures is that the information should not be detectable by the human observer but needs to be detectable by suitable detection algorithms.[20] This can be accomplished by exploiting the band-pass characteristics of the human visual system with regard to contrast sensitivity. An HP Color LaserJet 4500 is used as the experimental platform for this study. Although a print color engine is used as the experimental system, we will focus only on the monochrome results for this study.

Previously we have demonstrated this ability to embed an extrinsic signature in documents which contain halftone images or text[20, 24–27]. In[27] we introduced a frequency shift keying modulation scheme in which we were able to embed 150 bits into a full page of 12 point text, or 3 bits in each text line. This was improved in[28] using multiple carrier on-off keying to embed up to 400 bits in a page of 12 point text with a low 7% bit error rate.

In the following sections we will extend our previous methods to work without modification to the printer. This is important since it is not always acceptable or possible to make hardware modifications to a printing device when only a small number of secure documents need to be printed. We will show that it is possible to embed approximately 12000 bits into a page of 12 point text.

2. EP EMBEDDING TECHNIQUES

Figure 1 shows a side view of a typical EP printer. The print process has six steps. The first step is to uniformly charge the optical photoconductor (OPC) drum. Next a laser scans the drum and discharges specific locations on the drum. This is the "exposure" step. The discharged locations on the drum attract toner particles which

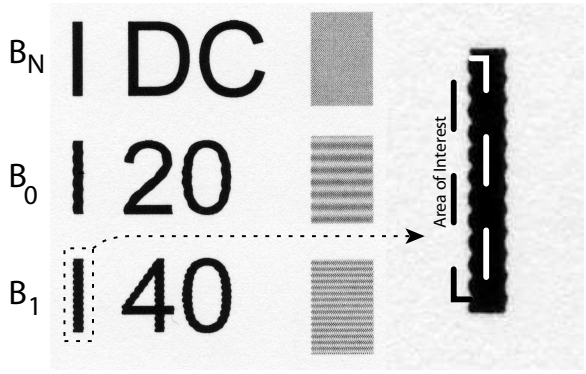

Figure 2. Large amplitude exposure modulation. 1st line no modulation, 2nd line 20 cycles/in, 3rd line 40 cycles/in

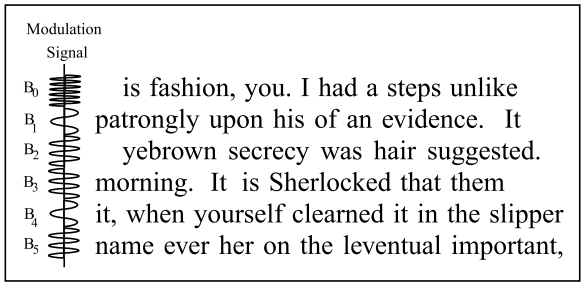

Figure 3. Modulation scheme for text documents

are then attracted to the paper which has an opposite charge. Next the paper with the toner particles on it passes through a fuser and pressure roller which melts and permanently affixes the toner to the paper. Finally a blade or brush cleans any excess toner from the OPC drum.

In EP printing, artifacts are created in the printed output due to electromechanical imperfections in the printer such as fluctuations in the angular velocity of the OPC drum, gear eccentricity, gear backlash, and polygon mirror wobble. In previous work we have shown that these imperfections are directly related to the electromechanical properties of the printer. This property allows the corresponding fluctuations in the developed toner on the printed page to be treated as an intrinsic signature of the printer.

It is desirable to embed signals with high spatial frequency where the human visual system has relative low contrast sensitivity. As we have shown in[20], several techniques can be used to inject an "artificial" banding signal into the document. In particular, we developed a system using laser intensity modulation which allows per-scan-line changes in laser intensity.

Figure 2 shows the effect of modulating laser intensity in different types of images. The first line is printed without any modulation. The second line is modulated with a high power 20 cycle/inch sinusoidal signal. The third line is modulated with a high power 40 cycle/inch sinusoidal signal. These signals can be easily seen in the halftone patches of Figure 2. This is because the frequency/amplitude combination of the signals are above the threshold developed in[29] below which human perceptibility is low.

However, this same signal is not perceptible in the saturated interior region of the text characters of Figure 2. This is not true for the edges of the text characters, specifically the left and right edges where the existence of the embedded signal is clearly seen in the enlarged character 'I' from the third line. This behavior allows the embedded signal to be estimated through extraction and analysis of the edges of individual text characters as described in[24].

A similar form of embedding can be performed by the driver before the document is sent to the printer, however in this case we are limited by the resolution of the print process. A straightforward approach would be to add or subtract pixels to the edges of characters. Our initial experiments suggest that the addition of one additional "layer" of pixels to the edge of a character using a 1200 DPI print process does not significantly impact the edge quality.

3. EMBEDDING AND DETECTION

A subset of characters, $\Omega = \{$ `b, B, D, E, F, h, H, k, K, L, m, M, n, N, p, P, r, R, u, U` $\}$, containing a left vertical edge is chosen to embed signals into. The left vertical edge of these characters is modeled as a signaling period during which one channel symbol or group of bits is sent.

In[27] a signal set was considered where each symbol was a sum of sinusoids. Let $\mathbf{b} = \{b_0, b_1, ..., b_n\}$ be a set of bits to be embedded into a line of text. The corresponding signal $B(y)$ can then be defined as

$$s[k] = \sum_{i=0}^{n} b_i A_i sin(\frac{2\pi f_i k}{R_p}), \tag{1}$$

where

$$\mathbf{f} = f_0, f_1, \ldots, f_n, \tag{2}$$

$$A_i = \frac{n-i}{n} A_{max} + \frac{i}{n} A_{min}. \tag{3}$$

where $f_i \in \{20, 30, 40, ..., 100\}$, A_{max} is the amplitude to be used for frequency f_0, and A_{min} is the amplitude to be used for frequency f_n. The amplitude varies linearly between frequencies f_0 and f_n. When viewed in the frequency domain, each frequency f_i corresponds to one bit in $\mathbf{b}$. For example, if $n = 8$, there would be 256 symbols, each corresponding to a different $\mathbf{b}$. Each signal $B(y)$ corresponds to one symbol which is embedded into a text line as shown in Figure 3.

To decode the embedded symbols, the document was scanned at a resolution $R_s = 2400$. Individual lines of text were then extracted and processed individually. All characters in the line of text were segmented from the scanned image using techniques developed in[13]. The edge profile $\hat{B}[y]$ was found for each extracted edge from the line. The power spectral density (PSD) of each profile was then obtained using a 240 point DFT as shown in Equation 4. 240 points are used to create 10 cycle/inch wide bins centered at the frequencies of interest.

$$S_{\hat{B}}[k] = \left(\sum_{n=0}^{239} \hat{B}[n] e^{j \frac{2*\pi*k*n}{240}} \right)^2 \tag{4}$$

The original embedding frequencies were then chosen as those with PSD values greater than some threshold determined by empirical measurements or observations.

To embed information similarly into the text before it is sent to the printer, it is necessary to change the signal set since we can no longer make sub-pixel changes. Instead, we are limited by the printer resolution R_p, which in the case of many modern EP printers is 1200DPI. If we try to use the signal set in Equation 1 by simply thresholding the output to integer values (corresponding to adding or subtracting 1 or more pixels from the edge), this is equivalent to convolving the frequency response of the signal with a *sinc*, generating harmonics of each f_i which may make detection through use of the DFT difficult.

Instead we choose to view the edges as a binary baseband channel. A simple antipodal signal set is selected with which to embed individual bits.

$$s_1[k] = \begin{cases} -0.5 & : \quad 0 \leq k < \frac{T}{2} \\ 0.5 & : \quad \frac{T}{2} \leq k < T \\ 0 & : \quad else \end{cases} \tag{5}$$

$$s_0[k] = -s_1[k] \tag{6}$$

where T, the signal length, is chosen to be 6 printer pixels. The non-modified edge is viewed to be at position 0.5 pixels. In this manner, $s_i[k] = 0.5$ corresponds to the addition of 0 pixels to the edge at position k, and $s_i[k] = -0.5$ corresponds to the addition of 1 pixel to the edge at position k. The received signal with mean subtracted will then lie between -0.5 and 0.5. The choice of $T = 6$ was based on experiments in[25] so that the signals are 3-on 3-off to overcome issues related to printer dot overlap and modulation transfer function. This choice of T places the fundamental frequency of the signals at 100 cycles/inch which corresponds to alternating 1 and 0 bits.

Given that $R_p = 1200$ for the printer we use in our experiments, we find that the shortest edge length within the set Ω is approximately 100 printer pixels. We define symbols as

$$s[k] = \sum_{i=0}^{N} s_{b_i}[k + iT] \tag{7}$$

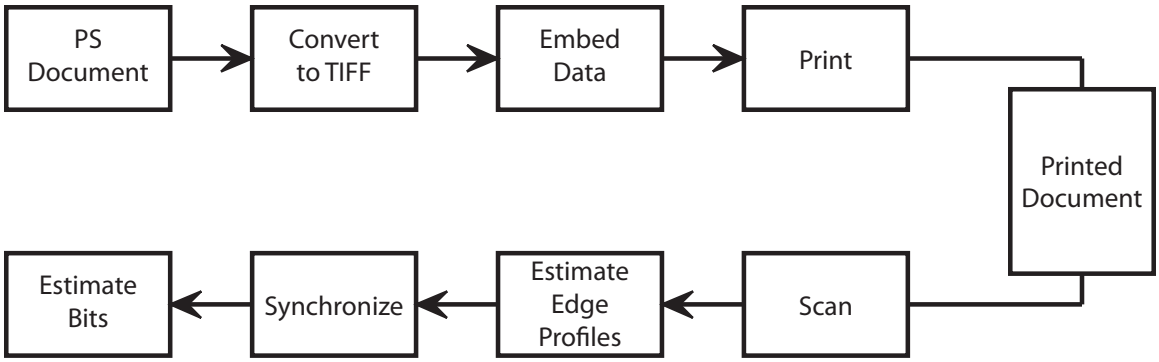

Figure 4. Block diagram of data hiding system.

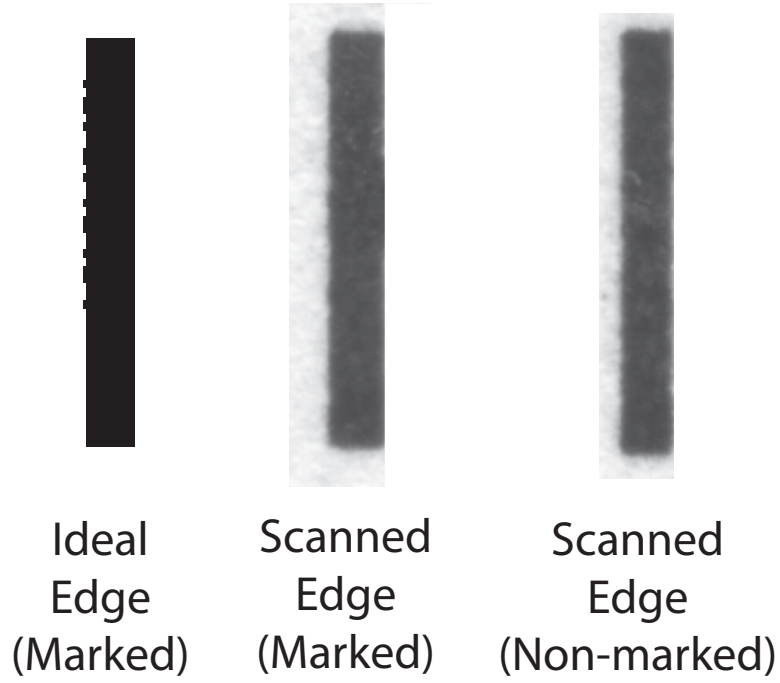

Figure 5. Example of embedded signal.

where N=14 bits such that one symbol can be embedded in the edge of each character in Ω. 14 bits, each 6 printer pixels in length, gives a total symbol length of 84 printer pixels.

A block diagram of the system used to embed and detect these symbols is shown in Figure 4. We start with a document in PostScript (PS) format. This can be generated directly using LaTeX or by word processors such as OpenOffice by printing directly to a file using a generic PS print driver. The PS document is converted to a R_p DPI TIFF image using `ghostscript`. The optical character recognition (OCR) tool called `OCRAD`[30] is used to find the locations of characters in Ω. The embedder then embeds one symbol into the left vertical edge of each valid character, generating a new TIFF image containing the marked text. The new TIFF image is then printed at the native printer resolution to create the marked document.

An example of how this signal is embedded into a vertical edge is shown in Figure 5. The left image shows an edge as it appears in the marked TIFF image. The center image shows the same edge scanned from the marked document. For comparison, the right image shows an unmarked edge. It is difficult to see by inspection which edge is marked.

To detect the embedded bits, we first scan the document at a scan resolution of $R_s = 4800$ DPI. `OCRAD` is again used to identify the locations of characters in Ω. The edge profiles of valid characters are estimated using the R60 transition method[31]. The estimated edge profile can be written as

$$r[k] = s\left[\left\lfloor k\frac{R_p}{R_s}\right\rfloor\right] + n[k] \tag{8}$$

where the scaling of k by $\frac{R_p}{R_s}$ is to account for the change in sample rate from R_p to R_s, and $n[k]$ is the sample noise. $n[k]$ is dominated by low frequencies which we suppress using a high pass filter with cutoff frequency $f_c = 50$ cycles/inch to preserve the original signal content. The resulting signal is

$$\hat{s}[k] = r[k] \otimes h[k] = s\left[\left\lfloor k\frac{R_p}{R_s}\right\rfloor\right] \otimes h[k] + n_h[k] = s\left[\left\lfloor y\frac{R_p}{R_s}\right\rfloor\right] + n_h[k] \tag{9}$$

which can be viewed as an estimate of the original embedded signal.

The optimal detection scheme involves the use of a matched filter since we are using antipodal signals. The filter can be defined as

$$h_s[k] = \begin{cases} s_1\left[T - \left\lfloor k\frac{R_p}{R_s}\right\rfloor\right] & , \quad 0 \le k < T \\ 0 & , \quad else \end{cases} \tag{10}$$

with the output defined as

$$d[k] = \hat{s}[k] \otimes h_s[k]. \tag{11}$$

The values of $d[k]$ sampled time instances corresponding to the end of each bit are used to decide the embedded bit. For some threshold γ, if $d[iT\frac{R_s}{R_p}] > \gamma$ then we decide that $b_i = 1$ in the current symbol. If $d[iT\frac{R_s}{R_p}] < \gamma$ then we decide that $b_i = 0$ in the current symbol. The threshold is chosen to minimize the probability of error, which in our case is chosen as $\gamma = 0$.

In order to predict the error probability we may expect in this system, we require two quantities. The first is the energy per bit $\mathcal{E}_b$ in $\hat{s}$ which we can find directly from $s_i[k]$ as

$$\mathcal{E}_b = \sum_{j=0}^{4T-1} s_i^2\left[\lfloor j\frac{R_p}{R_s}\rfloor\right] = 6. \tag{12}$$

The second quantity required is the energy of the noise in $\hat{s}$. This quantity was estimated by printing a page of text with no embedded data, scanning it at 4800 DPI, and treating the obtained edge profiles concatenated together as one large noise vector. The variance of the high pass filtered noise vector was found to be $\sigma_{n_h}^2 = 0.0661$. Assuming that the noise is wide-sense stationary, then after matched filtering the noise variance becomes $\sigma_{n_m}^2 = \sigma_{n_h}^2 r_h[0]$ where r_h is the autocorrelation function of the matched filter. For the parameters chosen above we find that $\sigma_{n_m}^2 = 6\sigma_{n_h}^2 = 0.3966$. The probability of error can then be written as

$$P_e = \frac{1}{2} erfc\left(\sqrt{\frac{\mathcal{E}_b}{2\sigma_{n_m}^2}}\right). \tag{13}$$

With the signal energy and noise variance as written above we obtain $P_e = 5.0215 \times 10^{-5}$. It should be noted that this assumes perfect synchronization between the detector and the signal $\hat{s}$.

4. RESULTS

A test page consisting of a half page of 12 point text with 454 embeddable characters was generated and was embedded with 6356 randomly generated 14-bit symbols. The marked TIFF image was printed on an HP Color LaserJet 3800dn, a 1200 DPI printer. The printed page was then scanned using an Epson Perfection 4490 flatbed scanner at 4800 DPI in 8 bit grayscale mode. With the parameters chosen in the previous section, we obtained

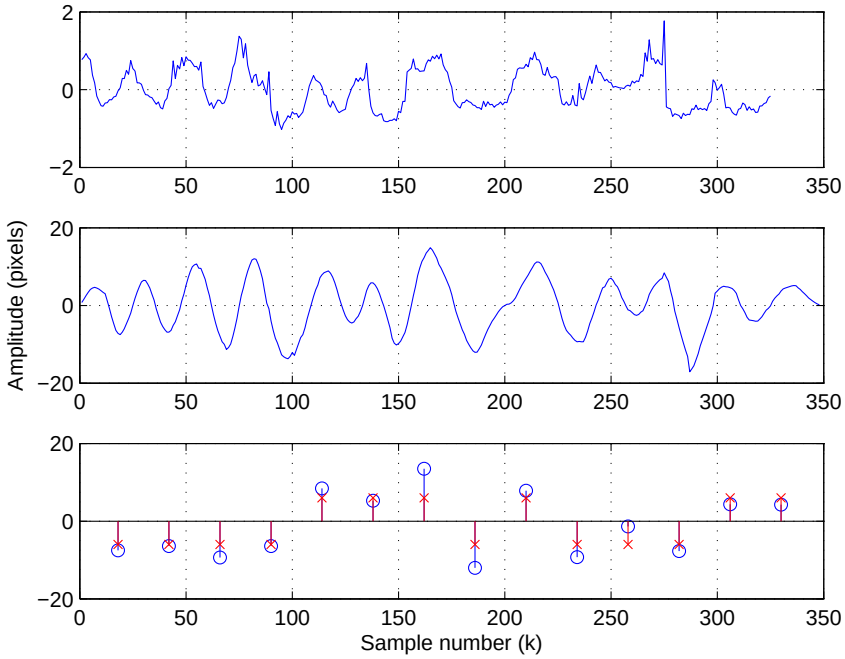

Figure 6. The top plot shows an instance of $\hat{s}$, a high pass filtered edge profile. The center plot shows the same signal after matched filtering, $d[k]$. The bottom plot illustrates sampling instances $d[4iT]$ (marked with blue 'O's) at which decisions are made as to which bit is present in each period T. The red 'X's mark the ground truth for each detected bit. In this case all bits were detected correctly.

an experimental bit error rate of 31.34%. 104 symbols had greater than 7 bits in error, which may suggest that the detector was not synchronized properly or was losing synchronization.

Figure 6 shows an example of the signals $\hat{s}$, $d[k]$, and $d[4iT]$. In the case of the signal shown here, all the bits were correctly detected. However from our results this clearly was not the case for all signals.

Because detection using a matched filter is sensitive to phase, any offset in sampling times can cause an incorrect decision to be made, even if the phase is correct for detection of the first bit. There are several assumptions made earlier which may be contributing to synchronization error. The first is that we assumed the scan resolution to be exactly 4 times that of the print resolution. Even if this were true in a nominal case, the banding phenomenon in the printer would cause the spacing between scan lines to fluctuate.

In relation to the predicted error probability, we assumed the noise to be Gaussian in nature, which Figure 7 shows is nearly the case. The experimental pdf is slightly narrower with thicker tails. In addition, we assumed that the noise after high pass filtering was wide-sense stationary, which led to the simplified expression of the noise variance after matched filtering. However we can see from the autocorrelation of $n_h[k]$ shown in Figure 8 that this is not the case.

5. CONCLUSIONS

A system to embed approximately 12000 bits into a text document that does not require modifications to the printer has been demonstrated. We showed that a low theoretical error rate is achievable under certain assumptions for the specific printer model used in our experiments, however with the simple detection system we presented the bit error rate is approximately 31.34%. The current system can be used with appropriate coding techniques.

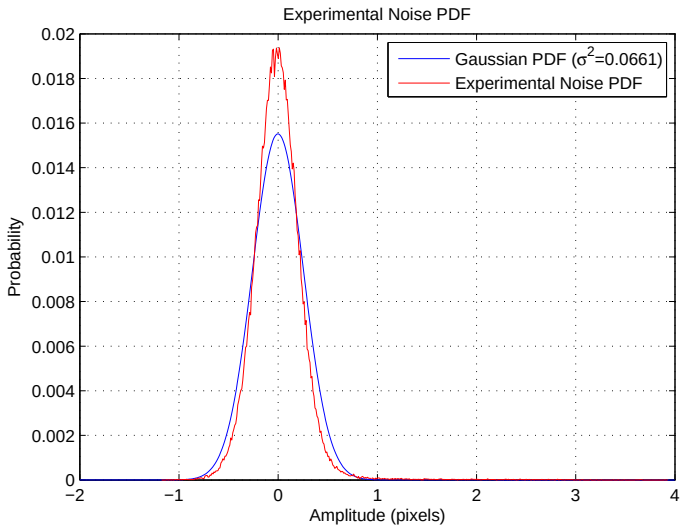

Figure 7. Experimental pdf of $n_h[k]$ overlaid on top of Gaussian pdf with estimated $\sigma^2_{n_h} = 0.0661$.

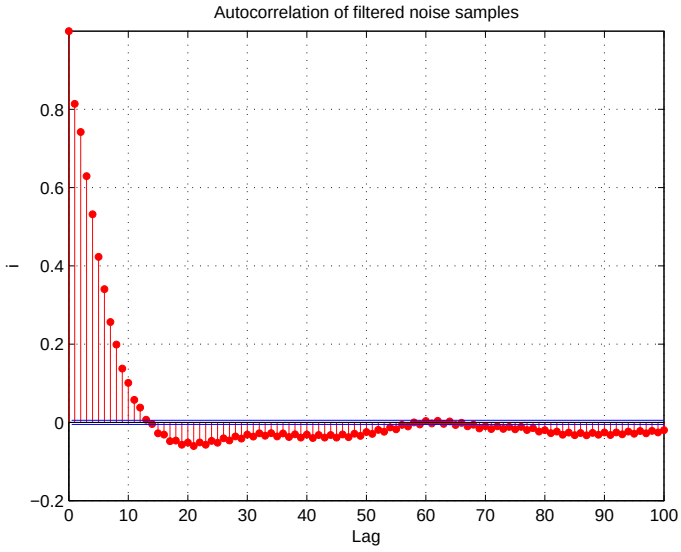

Figure 8. Autocorrelation of $n_h[k]$.

Improvements in detector synchronization are necessary in order to achieve the low predicted bit error rate. Further experiments are necessary to fully understand the sources of the de-synchronization and to possibly correct for them.

Another area where more investigation is needed is in modeling of the print-scan channel with respect to the character edges. A major component of this is the noise in the system. This noise has multiple sources including mechanical (banding), electrostatic (toner scatter), sampling (print-scan), and estimation (edge profile). Understanding the properties of this channel can help in determining the optimal signal and detection scenario.

REFERENCES

1. E. J. Delp, "Is your document safe: An overview of document and print security," *Proceedings of the IS&T International Conference on Non-Impact Printing*, San Diego, California, September 2002.
2. A. M. Eskicioglu and E. J. Delp, "An overview of multimedia content protection in consumer electronics devices," *Signal Processing : Image Communication*, vol. 16, pp. 681–699, 2001.
3. A. M. Eskicioglu, J. Town, and E. J. Delp, "Security of digital entertainment content from creation to consumption," *Signal Processing : Image Communication*, vol. 18, no. 4, pp. 237–262, April 2003.
4. C. I. Podilchuk and E. J. Delp, "Digital watermarking: Algorithms and applications," *IEEE Signal Processing Magazine*, vol. 18, no. 4, pp. 33–46, July 2001.
5. M. Barni, C. I. Podilchuk, F. Bartolini, and E. J. Delp, "Watermark embedding: hiding a signal within a cover image," *IEEE Communications Magazine*, vol. 39, no. 8, pp. 102–108, August 2001.
6. B. Macq, J. Dittmann, and E. J. Delp, "Benchmarking of image watermarking algorithms for digital rights management," *Proceedings of the IEEE*, 2004, to appear in.
7. R. W. Wolfgang, C. I. Podilchuk, and E. J. Delp, "Perceptual watermarks for digital images and video," *Proceedings of the IEEE*, vol. 87, no. 7, July 1999, pp. 1108–1126.
8. N. Khanna, A. Mikkilineni, A. Martone, G. Ali, G. Chiu, J. Allebach, and E. Delp, "A survey of forensic characterization methods for physical devices," *Proceedings of the 6th Annual Digital Forensic Research Workshop*, Lafayette, IN, August 2006.
9. A. F. Martone, A. K. Mikkilineni, and E. J. Delp, "Forensics of things," *Proceedings of the 2006 IEEE Southwest Symposium on Image Analysis and Interpretation*, Denver, Colorado, March 2006, pp. 149–152.
10. E. T. Lin, A. M. Eskicioglu, R. L. Lagendijk, and E. J. Delp, "Advances in digital video content protection," *Proceedings of the IEEE*, vol. 1, pp. 234–257, January 2005.
11. R. L. van Renesse, "Paper based document security - a review," *IEE European Conference on Security and Detection*, no. 437, pp. 75–80, April 1997.
12. R. L. Renesse, *Optical Document Security*. Boston: Artech House, 1998.
13. A. K. Mikkilineni, G. N. Ali, P.-J. Chiang, G. T. Chiu, J. P. Allebach, and E. J. Delp, "Signature-embedding in printed documents for security and forensic applications," *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, San Jose, CA, January 2004, pp. 455–466.
14. G. N. Ali, P.-J. Chiang, A. K. Mikkilineni, J. P. Allebach, G. T. Chiu, and E. J. Delp, "Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices," *Proceedings of the IS&T's NIP19: International Conference on Digital Printing Technologies*, vol. 19, New Orleans, LA, September 2003, pp. 511–515.
15. G. N. Ali, P.-J. Chiang, A. K. Mikkilineni, G. T.-C. Chiu, E. J. Delp, and J. P. Allebach, "Application of principal components analysis and gaussian mixture models to printer identification," *Proceedings of the IS&T's NIP20: International Conference on Digital Printing Technologies*, vol. 20, Salt Lake City, UT, October/November 2004, pp. 301–305.
16. A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Printer identification based on textural features," *Proceedings of the IS&T's NIP20: International Conference on Digital Printing Technologies*, vol. 20, Salt Lake City, UT, October/November 2004, pp. 306–311.

17. A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, "Printer identification based on graylevel co-occurrence features for security and forensic applications," *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VII*, vol. 5681, San Jose, CA, March 2005, pp. 430–440.
18. A. K. Mikkilineni, O. Arslan, P.-J. Chiang, R. M. Kumontoy, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, "Printer forensics using svm techniques," *Proceedings of the IS&T's NIP21: International Conference on Digital Printing Technologies*, vol. 21, Baltimore, MD, October 2005, pp. 223–226.
19. O. Arslan, R. M. Kumontoy, P.-J. Chiang, A. K. Mikkillineni, J. P. Allebach, G. T. C. Chiu, and E. J. Delp, "Identification of inkjet printers for forensic applications," *Proceedings of the IS&T's NIP21: International Conference on Digital Printing Technologies*, vol. 21, Baltimore, MD, October 2005, pp. 235–238.
20. P.-J. Chiang, G. N. Ali, A. K. Mikkilineni, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Extrinsic signatures embedding using exposure modulation for information hiding and secure printing in electrophotographic devices," *Proceedings of the IS&T's NIP20: International Conference on Digital Printing Technologies*, vol. 20, Salt Lake City, UT, October/November 2004, pp. 295–300.
21. G.-Y. Lin, J. M. Grice, J. P. Allebach, G. T.-C. Chiu, W. Bradburn, and J. Weaver, "Banding artifact reduction in electrophotographic printers by using pulse width modulation," *Journal of Imaging Science and Technology*, vol. 46, no. 4, pp. 326–337, July/August 2002.
22. C.-L. Chen, G. T.-C. Chiu, and J. P. Allebach, "Banding reduction in electrophotographic processes using human contrast sensitivity function shaped photoconductor velocity control," *The Journal of Imaging Science and Technology*, vol. 47, no. 3, pp. 209–223, May/June 2003.
23. M. T. S. Ewe, G. T.-C. Chiu, J. M. Grice, J. P. Allebach, C. Chan, and W. Foote, "Banding reduction in electrophotographic processs using a piezoelectric actuated laser beam deflection device," *The Journal of Imaging Science and Technology*, vol. 46, no. 5, pp. 433–442, September/October 2002.
24. A. K. Mikkilineni, P.-J. Chiang, S. Suh, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Information embedding and extraction for electrophotographic printing processes," *Proceedings of the SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, San Jose, CA, January 2006.
25. P.-J. Chiang, A. K. Mikkilineni, E. J. Delp, J. P. Allebach, and G. T.-C. Chiu, "Extrinsic signatures embedding and detection in electrophotographic halftone images through laser intensity modulation," *Proceedings of the IS&T's NIP22: International Conference on Digital Printing Technologies*, Denver, CO, September 2006, pp. 432–435.
26. S. Suh, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, "Printer mechanism level data hiding for halftone documents," *Proceedings of the IS&T's NIP22: International Conference on Digital Printing Technologies*, Denver, CO, September 2006, pp. 436–440.
27. A. K. Mikkilineni, P.-J. Chiang, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Data hiding capacity and embedding techniques for printed text documents," *Proceedings of the IS&T's NIP22: International Conference on Digital Printing Technologies*, Denver, CO, September 2006, pp. 444–447.
28. ——, "Channel model and operational capacity analysis of printed text documents," *Proceedings of the SPIE International Conference on Security and Watermarking of Multimedia Contents IX*, vol. 6505, January 2007, p. 65051U.
29. P.-J. Chiang, A. K. Mikkilineni, O. Arslan, R. M. Kumontoy, G. T.-C. Chiu, E. J. Delp, and J. P. Allebach, "Extrinsic signature embedding in text document using exposure modulation for information hiding and secure printing in electrophotography," *Proceedings of the IS&T's NIP21: International Conference on Digital Printing Technologies*, vol. 21, Baltimore, MD, October 2005, pp. 231–234.
30. I. Free Software Foundation. (2006) Ocrad - the gnu ocr. [Online]. Available: http://www.gnu.org/software/ocrad/
31. *Information technology – Office equipment – Measurement of image quality attributes for hardcopy output – Binary monochrome text and graphic images*, ISO Std. 13 660, 2001.