GEOMETRIC DISTORTION SIGNATURES FOR PRINTER IDENTIFICATION

Orhan Bulan, Junwen Mao, Gaurav Sharma

ECE Dept., University of Rochester, Rochester, NY, 14627-0126
{bulan, jumao, gsharma}@ece.rochester.edu

ABSTRACT

We present a forensic technique for analyzing a printed image in order to trace the originating printer. Our method, which is applicable for commonly used electrophotographic (EP) printers, operates by exploiting the geometric distortion that these devices inevitably introduce in the printing process. In the proposed method, first a geometric distortion signature is estimated for an EP printer. This estimate is obtained using only the images printed on the printer and without access to the internal printer controls. Once a database of printer signatures is available, the printer utilized to print a test image is identified by computing the geometric distortion signature from test image and correlating the computes signatures against the printer signatures in the database. Experiments conducted over a corpus of EP printers demonstrate that the geometric distortion signatures of test documents exhibit high correlation with the corresponding printer signatures and a low correlation with other printer signatures. The method is therefore quite promising for forensic printer identification applications. We highlight several of the capabilities and challenges for the method.

Index Terms— Printer forensics, printer identification, hardcopy document forensics, geometric distortion signature

1. INTRODUCTION

Printed documents continue to carry important data for financial, legal, and governmental transactions. Banknotes, driver's licenses, passports, tickets, and contracts are examples of printed documents that play a crucial role in our day to day lives. For these applications, it is vital that the documents be secure in the sense that they are tamper-proof, not readily replicable, machine/human verifiable, etc. Various methods have therefore been developed (and others are under development) to address these security concerns [1–5]. In addition to these security techniques, forensic tools for the analysis of printed documents are also of significant interest because they can provide valuable information to law enforcement and intelligence agencies [6–9]. In this paper, we focus on forensic analysis aimed at identifying the printer used for producing a given document, or more generally at identifying which of a group of documents were produced on the same printer. Of the many possible applications that could benefit from such a forensic tool, we note a couple of motivating examples: Forensic analysis that can recognize which documents were produced on a single printer can be invaluable for intelligence agencies because links among printed documents found at different sites of illicit activity can reveal the latent connections between groups operating at the different sites. Likewise, such analysis can also be invaluable in identifying and tracking counterfeit and contraband documents such as forged currency, intelligence reports, and design documents for controlled armaments such as nuclear weapons.

In our work, we address a particular instance of the printer identification problem focusing specifically on the identification of electrophotographic (EP) printers from printed images. Electrophotographic printers, commonly referred to as laser printers, are widely deployed in office environment and due to their higher speeds and cost efficiency are the largest source of printed pages on the desktop, despite having a smaller installed base than inkjet printers. Our proposed method operates by exploiting the locally varying geometric distortion in the printing process encountered during EP printing, wherein, due to limitations and inaccuracies in the components and assembly, the ideal rectilinear digital grid of pixels that forms the "input" to the printer is warped in the process of printing resulting in a geometrically distorted version in print. This geometric distortion provides a useful signature for printer identification because: a) some distortion is inevitable in the printing process, b) the distortion is distinctive since it varies across both printer models and individual printers, and c) an estimate of the distortion can be obtained from a high resolution scan of the printed image without requiring access to the printer internals or specialized equipment. We demonstrate that we can reliably estimate the geometric distortion signature from printed halftone images by estimating the variations in the positions of halftone dots in printed images. Once the geometric distortion estimate is available, a simple correlation based measure can be utilized to establish the presence/absence of a signature in the print. We demonstrate experimentally that the proposed printer identification method can be used to identify the EP printers reliably.

The rest of this paper is organized as follows. Section 2 gives an overview of physical structure of EP printing mechanism and illustrates the source of the geometric distortion. In Section 3, we describe geometric distortion signatures and correlation based identification algorithm. We present experimental results in Section 4. Section 5 concludes the paper by summarizing the key aspects of our method.

2. GEOMETRIC DISTORTIONS IN EP PRINTING

Geometric distortion caused by an EP printer is illustrated in Figure 1. The solid blue lines show a regular rectangular grid on the 2-D array of pixels forming the input to the printer and the dashed blue lines indicate the corresponding grid obtained on a printed sheet at the output of the printer[1]. Due to non-uniform geometric distortions caused by the printer, the input and output grids do not match

This work was supported in part by the Xerox Foundation, by the Air Force Office of Scientific Research (AFOSR) under grant number FA9550-07-1-0017, and by a matching grant from New York State Office of Science, Technology and Academic Research (NYSTAR) through the Center for Electronic Imaging Systems (CEIS).

[1] The plot of Figure 1 represents an actual estimate obtained from a printer using the methods that we describe subsequently

with each other. Understanding of the source of the geometric distortion requires some knowledge about EP printing process. Hence, we will briefly describe EP printing next.

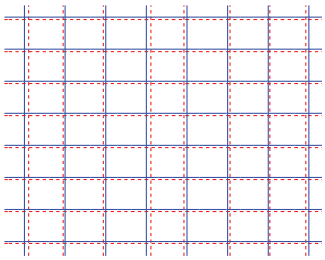


Fig. 1. Geometric distortion introduced by an EP printer

Figure 2 illustrates six steps involved in a typical EP printing process. In the charging step, the charge roller applies a negative charge to the optical photoconductor (OPC) drum. The charged drum is then exposed by a laser beam directed by a polygon mirror whose rotation causes the beam to scan across the OPC drum in a *raster line* (see Fig. 3). As the laser scans across the OPC drum it writes out a line of input pixels corresponding to the raster line. This writing is accomplished by turning the laser on and off so as to expose the drum in the regions in which no printing is desired. The OPC drum then rotates to position the next scanline along the laser's scanning path and the procedure is repeated for the following raster line. The negative charges on the OPC drum are neutralized in the exposed regions. In the development step, positively charged toner is dispersed over the OPC. The negatively charges locations on the drum attract toner particles. In the transfer step, the image formed on the OPC drum by the toner particles is transferred onto the paper. The fuser then melts the toner into the paper by using heat and pressure. In the last step, a cleaning blade removes any excess toner from the OPC drum.

Geometric distortion introduced by EP printers are mainly due to variations in the OPC drum and polygon mirror as illustrated in Fig. 3. As the laser beam is swept across a scanline on the OPC drum, its velocity on the drum is not constant. As a result the sizes of pixels and inter-pixel spacing along a scanline exhibit a corresponding variation. This results in the geometric distortion of the type illustrated in Fig. 1. In some systems, corrective mechanisms are used to minimize this geometric distortion [10, 11], however, residual distortion is still present after correction and can be detected in printed documents. Similar to the geometric distortion caused by the variations in laser scanning speed over a scanline, variations in the velocity of the OPC drum causes non-uniform spacing between raster lines. This is also a form of geometric distortion that is manifested as banding in the output print, a feature that has also been used for forensics by others [9].

3. GEOMETRIC DISTORTION SIGNATURES

Most printers use halftoning algorithms to binarize the contone image (8 bit/pixel) before "physical printing" such that halftone image gives the same visual perception from a typical viewing distance. Clustered dot halftones are commonly used in EP printers due to their stability and reproducibility. Clustered dot halftones are also called as AM halftones in the sense that they are periodic and generate different levels of gray by varying the size of halftone spots as illustrated in Figure 4 (a). Geometric distortions described in Section 2 causes small local variations in dot positions. In our work, we extract a geometric signature by comparing dot positions extracted from the printed image and estimated dot positions before printing.

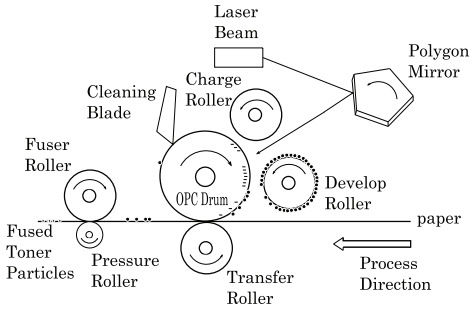


Fig. 2. EP printing process overview

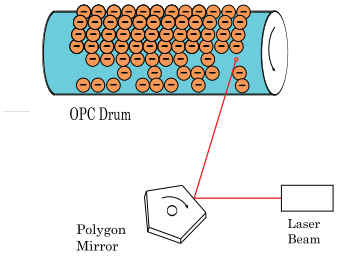
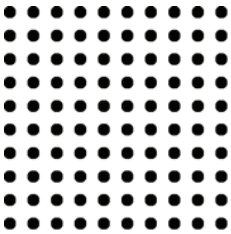
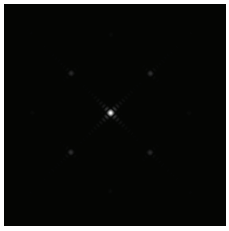


Fig. 3. OPC drum and laser



(a) Clustered Dot Halftones



(b) Frequency Domain representation of (a)

Fig. 4. Clustered Dot Halftones and Its Frequency Domain Representation

3.1. Dot Center Extraction

In this part, we consider estimation positions of the halftone-dot centers. We first perform a frequency domain analysis on the scanned image $I^s(a, b)$ and estimate the periodicity of halftone dots. Figure 4 (b) shows the frequency domain representation of the halftone image in Fig. 4 (a) where primary peaks can be utilized for periodicity estimation.

After finding the periodicity of the halftone dots, we sequentially estimate the dot positions from left to right. Since geometric distortion on the printed document varies smoothly, the distance between neighboring halftone dots deviates slightly from the estimated periodicity. For this reason, we also use the knowledge of previous dot positions to estimate the position of current halftone dot. In the $(i, j)^{th}$ halftone cell C we calculate the dot center (x_{ij}, y_{ij}) as the center of mass of the cell given by:

$$x_{ij} = \frac{\sum_{k,l \in C} I^s(k,l)k}{\sum_{k,l \in C} I^s(k,l)} \quad (1)$$

$$y_{ij} = \frac{\sum_{k,l \in C} I^s(k,l)l}{\sum_{k,l \in C} I^s(k,l)} \quad (2)$$

3.2. Rotation Compensation

The halftone dot positions are extracted from a scanned version of the printed document which is subject to small rotation in the scanning process. We need to compensate for this small rotation in order to correctly estimate the geometric distortion caused by the printer.

Firstly, we estimate the rotation angle from extracted dot positions. We fit two perpendicular lines as $y_{1j}^{est} = a_1 x_{1j} + b_1$ and $y_{i1}^{est} = a_2 x_{i1} + b_2$ to extracted dot positions in the a row and column, where x_{1j}, x_{i1} are abscissas of the dots, and y_{1j}^{est}, y_{i1}^{est} are estimated by line fitting[2]. We then find a_1, a_2, b_1 and b_2 by using least square estimates where we formulate the optimization problem as:

$$\min_{a_1,a_2} \sum_{i=1}^{m}\sum_{j=1}^{n} (y_{1j} - a_1 x_{1j} - b_1)^2 + (y_{i1} - a_2 x_{i1} - b_2)^2$$
$$\text{subject to } a_1 a_2 = -1$$

We then calculate the rotation angle $\theta = \arctan(a_1)$ and compensate for the rotation by performing coordinate transform as: $\hat{x}_{ij} = x_{ij}\cos\theta + y_{ij}\sin\theta$, $\hat{y}_{ij} = -x_{ij}\sin\theta + y_{ij}\cos\theta$.

3.3. Geometric Distortion Signature Extraction

In order to extract the geometric distortion signature from the output document, we need to estimate dot positions prior to printing. For this purpose, we use the fact that for AM clustered dot halftones, these positions lie on a period rectilinear grid. The periodicity of this grid is readily estimated from the peaks in the frequency spectrum of the scanned image as illustrated in Fig. 4(b). This periodicity provides the estimates (x_{ij}^r, y_{ij}^r) of the halftone dots on the regular grid at the input to the printer. A 2-D distortion displacement vector for each halftone dot is then obtained by calculating the displacement that printer causes:

$$\begin{pmatrix} \overline{x}_{ij} \\ \overline{y}_{ij} \end{pmatrix} = \begin{pmatrix} \hat{x}_{ij} \\ \hat{y}_{ij} \end{pmatrix} - \begin{pmatrix} x_{ij}^r \\ y_{ij}^r \end{pmatrix} \quad (3)$$

The collection of distortion displacement vectors over the halftone dots forms the distortion signature for the print.

3.4. Correlation-measure for Signature Similarity

We assess the similarity of a pair of geometric distortion signatures using the normalized correlation. For two signatures, $(\overline{x}_{ij}^1, \overline{y}_{ij}^1)$ and $(\overline{x}_{ij}^2, \overline{y}_{ij}^2)$ computed on a common support $\{(i,j)\}$, the normalized correlation is defined as

$$\rho_{12} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{\overline{x}_{ij}^1 \overline{x}_{ij}^2 + \overline{y}_{ij}^1 \overline{y}_{ij}^2}{\sqrt{(\overline{x}_{ij}^1)^2 + (\overline{y}_{ij}^1)^2}\sqrt{(\overline{x}_{ij}^2)^2 + (\overline{y}_{ij}^2)^2}} \quad (4)$$

[2]The rotation angle could alternately be estimated from the rotation of the peaks in the Fourier spectrum of the halftone image. For small rotations, the process described here provides a more accurate estimate in practice.

The normalized correlation takes values in $[-1, 1]$ where a high correlation indicates strong similarity between the two geometric distortion signatures and low similarity indicates that the signatures are quite distinct.

While the support of the estimated signatures is identical in cases where we have the same halftone frequencies and orientations, the support over which the signatures are computed is not identical when the method is applied across printers with different halftones (with different halftone frequencies and orientations). Since the geometric distortion is smoothly varying[3] in situations where the halftone frequencies corresponding to the two signatures under consideration are different, the coarser signatures can be interpolated to the finer sampled version, using the physical coordinates on the sheet of paper for suitably registering the signatures. We utilize this process in our comparison of signatures computed for halftones with disparate frequencies[4].

4. EXPERIMENTAL RESULTS

We evaluate the performance of the proposed method over six different printer models from various manufacturers as indicated in Table 1. Firstly, we consider clustering printed images that are originated from the same printer model and type. For this purpose, we use several printed images with different graylevels from each printer. For each printed image, we estimate a geometric distortion signature using the method of Section 3.3. We then compute the normalized correlation between the extracted signatures for each pair as described in Section 3.4. Figure 5 shows the histogram of the calculated correlations where the correlation between documents printed on the same printer are shown in magenta (light gray in monochrome printed versions) and correlations between documents printed on different printers are shown in black. For our test, four documents corresponding to different graylevels were used in each printer yielding $\binom{24}{2} = 276$ pairs of normalized cross correlations. For our limited sample set, the histograms shown in Fig. 5 clearly illustrate that by using a suitable threshold (0.87) documents printed by the same printer model can be reliably identified.

Table 1. Printer Used in Experiment

Brand	Model	DPI
Canon	IR5070	600
HP	4240n	1200
HP	4250dtn	1200
Xerox	Nuvera120	600
Xerox	DocuColor Igen3	600
Xerox	DocuColor8000	2400

We next evaluate the performance of the proposed method for identification of the printer model from printed images. We first generate a database of printer signatures by extracting geometric distortion signatures from several documents of each printer, and averaging these signatures (after registration to a common coordinate axis

[3]Note that smoothly varying distortions are largely imperceptible whereas non-smoothly varying geometric distortions give rise to objectionable artifacts. In fact, the so called random-bending geometric attack in StirMark [12] is modeled after the smoothly locally varying distortions inherently encountered in printing.

[4]Our preliminary experiments also indicate that, as might be expected from the underlying causes of geometric distortion, the estimated signatures are halftone invariant.

Table 2. Correlation between Printers

	Canon	HP4240n	HP4250dtn	Nuvera120	DocuColor Igen3	DocuColor8000
Canon	0.9814	0.1470	-0.3715	-0.1924	0.1770	0.0120
HP4240n	0.1470	0.9899	0.7558	-0.3927	-0.7086	-0.7963
HP4250dtn	-0.3715	0.7558	0.9763	-0.1862	-0.7528	-0.6742
Nuvera120	-0.1924	-0.3927	-0.1862	0.8416	0.4685	0.2610
DocuColor Igen3	0.1770	-0.7086	-0.7528	0.4685	0.9473	0.6360
DocuColor8000	0.0120	-0.7963	-0.6742	0.2610	0.6360	0.9731

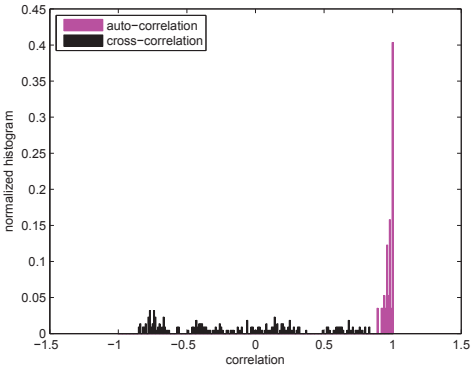


Fig. 5. Correlation of documents from six printers

on the physical sheet of paper) to get the printer signature. We then use a group of test images from each printer to identify the print origin by the proposed method.

Table 2 lists the averages of the correlation values between the estimated signatures for the test documents and precomputed printer signatures in the "database". The rows and columns in Table 2 are both indexed by the printer model, with rows corresponding to the printer used to generate the test documents and the columns corresponding to the printer whose signature is under consideration. The correlation values are the averages of correlations among documents on each printer. From the tabulated numbers we see that signatures of test images exhibit a strong correlation with the true source printer signature (diagonal entries in Table 2) and low correlation with the rest(off diagonal entries in Table 2). Thus, the source printer can be identified as the one that shows the highest correlation with the output document.

5. CONCLUSION AND DISCUSSION

Estimates of geometric distortion introduced in electrophotographic printers can be obtained from printed images. These estimates serve as a useful forensic signature for identifying the originating printer and/or for grouping together images produced by the same printer. The experimental results presented here demonstrate the promise of this forensic technique: geometric distortion signatures for documents printed on the same printer exhibit a strong correlation across varying graylevels whereas the signatures for documents printed on different printers demonstrate significantly lower correlation.

The results presented here illustrate the promise of the proposed geometric distortion signature as a forensic tool. Since the performance is evaluated over a random selection of printers available in our office environment, it is anticipated that the method generalizes to larger sets of printers. This needs to, however, be experimentally validated and is something that is being undertaken in ongoing work.

In order to fully enable printed document forensics through signal/image processing, a toolkit comprising of multiple techniques must be assembled. The method presented here represents one component in such a toolkit. Finally, we note that extensions of the geometric distortion signatures to text documents and to color prints also represent problems that are worthy of further exploration.

6. REFERENCES

[1] G. Sharma and S.-G. Wang, "Show-through watermarking of duplex printed documents," in *Proc. SPIE: Security, Steganography, and Watermarking of Multimedia Contents VI*, E. J. Delp and P. W. Wong, Eds., vol. 5306, Jan. 2004.

[2] B. Oztan and G. Sharma, "Continuous phase modulated halftones and their application to halftone data embedding," in *Proc. IEEE Intl. Conf. Acoustics Speech and Sig. Proc.*, vol. II, May 2006, pp. 333–336. [Online]. Available: http://www.ece.rochester.edu/~basak/Papers/icassp06paper.pdf

[3] K. Solanki, U. Madhow, B. S. Manjunath, S. Chandrasekaran, and I. El-Khalil, "Print and scan resilient data hiding in images," *IEEE Trans. Info. Forensics and Security*, vol. 1, no. 4, pp. 464–478, Dec. 2006.

[4] P. Chiang, A. Mikkilineni, E. Delp, J. Allebach, and G. Chiu, "Extrinsic Signatures Embedding and Detection in Electrophotographic Halftone Images through Laser Intensity Modulation," *International Conference on Digital Printing Technologies*, vol. 22, p. 432, 2006.

[5] O. Bulan, V. Monga, G. Sharma, and B. Oztan, "Data embedding in hardcopy images via halftone-dot orientation modulation," in *Proc. SPIE: Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, Eds., vol. 6819, Jan. 2008, pp. 68 190C–1–12.

[6] V. Talbot, P. Perrot, and C. Murie, "Ink Jet Printing Discrimination Based on Invariant Moment," in *International Conference on Digital Printing Technologies*, vol. 22, 2006, p. 427.

[7] J. Oliver and J. Chen, "Use of Signature Analysis to Discriminate Digital Printing Technologies," *International Conference on Digital Printing Technologies*, pp. 218–222, 2002.

[8] O. Arslan, R. Kumontoy, P. Chiang, A. Mikkillineni, J. Allebach, G. Chiu, and E. Delp, "Identification of inkjet printers for forensic applications," in *Proceedings of the IS&Ts NIP21: International Conference on Digital Printing Technologies*, vol. 21, 2005, pp. 235–238.

[9] A. Mikkilineni, G. Ali, P. Chiang, G. Chiu, J. Allebach, and E. Delp, "Signature-embedding in printed documents for security and forensic applications," in *Proc. SPIE: Security, Steganography, and Watermarking of Multimedia Contents VI*, vol. 5306, Jan. 2004, pp. 455–466.

[10] T. Kazama and Y. Matsuzaki, "Image forming apparatus, image forming method, and storage medium storing program for forming image," United States Patent Application Publication No. US2007/0139715 A1, 21 Jun. 2007.

[11] D. N. Curry, "Two dimensional linearity and registration error correction in a hyperacurity printer," United States Patent No. 5732162, 1998.

[12] F. Petitcolas and M. Kuhn, "Stirmark software," Available from http://www.petitcolas.net/fabien/watermarking/stirmark/.