Boise State University
ScholarWorks

Electrical and Computer Engineering Faculty Publications and Presentations

Department of Electrical and Computer Engineering

6-21-2004

Printer Modeling for Document Imaging

Margaret Norris
Boise State University

Elisa H. Barney Smith
Boise State University

Printer Modeling for Document Imaging

Margaret Norris[1] and Elisa H. Barney Smith[2]
Electrical and Computer Engineering Department
Boise State University
Boise, ID 83725-2075

ABSTRACT - The microscopic details of printing often are unnoticed by humans, but can make differences that affect machine recognition of printed text. Models of the defects introduced into images by printing can be used to improve machine recognition. A probabilistic model used to generate images showing toner placement bears similarities to actual printed images. An equation derived for the average coverage of paper by toner particles having probabilistic placement is developed using geometric probability. Simulations show that averages of 'printed images' do have the same average coverage as the derived average coverage equations.

Keywords: Printer modeling, image defects, geometric probability.

1. Introduction

Most printed documents are configured to be visually pleasing to humans. When printed text documents are to be viewed by machines for optical character recognition (OCR) or other document analysis purposes, details in the document images are more significant. Low OCR accuracy rates are most common in documents with image degradations caused by printing, scanning, photocopying and/or FAXing documents. The defects that occur through scanning have been studied in [1][2][3][6][9]. This paper analyzes the fine structure of the electrophotographic (laser) printing process and how through it, toner is placed on paper and on average what images result. This analysis of printing will enable the printer model to be combined with scanning defect model to produce a model of the complete document process.

Usually it is assumed that each image pixel is printed solidly black on a white paper background, Figure 1a. In reality, the toner adheres to the paper in amounts proportional to the amount that the laser discharges the photoconductor (PC) making a gradual transition from solid paper, to paper that is covered 100% by toner particles, Figure 1b. A detailed simulation model of the charge applied to the photoconductor in the electrophotographic printing process was developed by Yi [12][13] to be used by engineers at Hewlett-Packard to develop improved printer technologies. The simulation takes various parameters that define the print engine and calculates exposure energy and voltage on the PC surface for a given source image. The effects of electrophotographic printing and laser modulation have been incorporated in some halftoning algorithms [4][7]. This paper quantifies how the toner will be distributed on paper for a given laser trace representing a source image, and therefore, how black the page will be as a function of space. This is done through calculation of the average amount of toner per unit area, which can later be used to determine what intensity a scanner will see as it scans the page at high resolution.

High resolution pictures taken of toner are compared with our simulated toner placement in Section 2 for verification that the model is appropriate and for determination of model parameters. In Section 3 a mathematical expression is derived that describes the average amount of paper that is covered by toner. This starts with a uniform toner density, then expands under limiting conditions for varying toner density. Then this will be applied to toner densities that would correspond to specific shapes to be drawn on paper in Section 4. We show some simulations of toner being placed with some variable density patterns and this will be compared to the theoretical equation derived in Section 3. This is followed by the conclusion.

---

1. Presenting Author
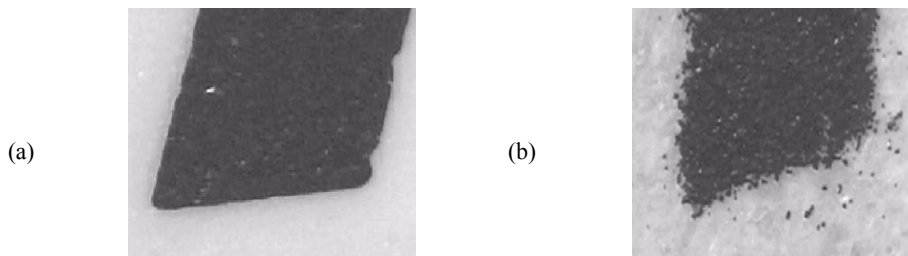2. Corresponding Author: EBarneySmith@BoiseState.edu

Figure 1:   Examples of characters printed with (a) phototypset printing and (b) laser printing. Pictures taken with high power microscope.

2. Simulations vs. Reality

All xerographic printers use similar basic steps in the process of creating a printed document. This starts with charging the photoconductor (PC) drum. The image that is to be printed by the computer is converted into a series of laser traces that will represent the image. The laser traces across the PC discharging the area it covers in a series of rows. Examples of pixel patterns and corresponding laser traces used in this paper are shown in Figure 2. The laser beam has a Gaussian intensity profile [11][12]. The profile is anisotropic and based on [12][13] where $\sigma_y=1.22\sigma_x$ which is used in this paper. The toner particles adhere to the discharged areas in quantities proportional to the charge. The toner is then transferred to the paper where rollers heat and compress the toner into the paper. Then the PC is cleaned and prepared for the next image.
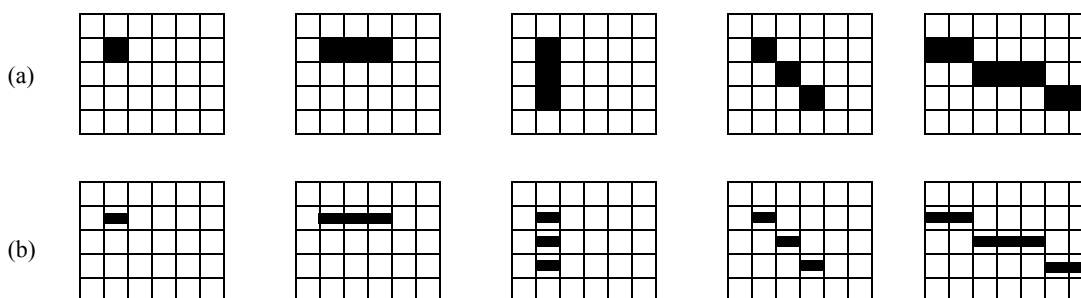

Figure 2:   (a) Conceptual pixel shapes and (b) corresponding laser traces for single pixel, three horizontally aligned pixels, three vertically aligned pixels, three diagonal pixels, and seven pixels in a 20 degree line.

With the assumption that toner particles are placed on paper with a density proportional to the energy of the laser trace, we first need to evaluate how well the model that toner is placed with a Gaussian density over the path of the laser trace matches with physical data. The unknown model parameters are the size of toner particles, r, the spread of the Gaussian laser energy in the horizontal direction, σ_x, and the number of toner particles likely to adhere to the paper per unit length of laser sweep (pixel) N. Figures 3 show the output of this model for a single pixel which corresponds to a laser tracing horizontally 1/600th of an inch. Figures 3a-i shows how the output varies for a range of σ_x and N parameter values. Figures 3j-l show the effect of varying the size of the toner particle, r. Nominal values for the model parameters were determined by comparing the simulated pixels to images of single pixels printed in isolation on paper, Figure 4. Several other image patterns were also evaluated. Based on several trial and error experiments, nominal values for the model parameters were determined. Their nominal values are shown in Table 1.

To confirm that the model would produce representative samples similar to actual samples, other laser trace patterns were considered. Images representing three pixels horizontally aligned, were produced by extending the length of the laser trace, Figure 5a and 5b. Finally we considered how adjacent laser traces would sum by considering vertically adjacent pixels. This is shown in Figure 5c and 5d. As can be seen

from the figures, even though the toner is placed randomly, there is a high degree of similarity between the simulated toner placement and the actual toner placement. This gives justification for using the proposed model and choice of model parameters.
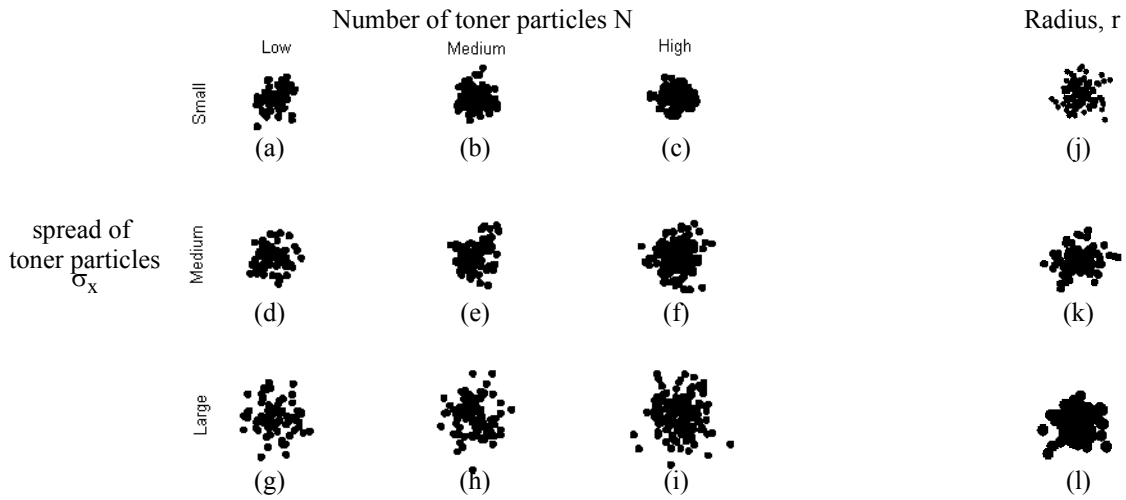

Figure 3:   Simulations of single pixels with (a-i) varying $\sigma_X \in \{0.35, 0.50, 0.71\}$ and $N \in \{50, 100, 150\}$ (j-l) varying toner particle radii, $r \in \{0.14, 0.20, 0.28\}$.
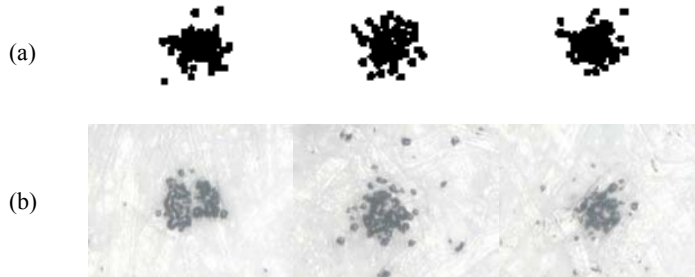

Figure 4:   (a) Sample images of isolated pixels shown together with (b) pictures of isolated pixels taken with a microscope.
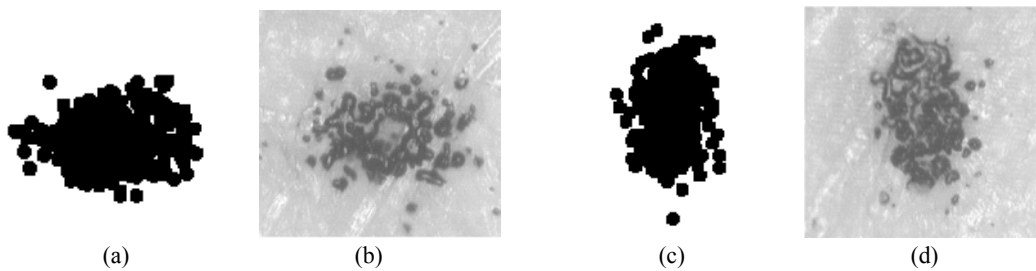

Figure 5:   (a) Sample images of horizontally adjacent pixels shown together with (b) pictures taken with a microscope. (c) Sample images of vertical adjacent pixels shown together with (d) pictures taken with a microscope.

Table 1: Nominal printer model parameters

r	σ	N_a
0.00033 inches =0.2 pixels	0.00083 inches =0.5 pixels	100/pixel

3

3. Average Coverage

Given that the toner will be placed in a probabilistic fashion on the paper, an equation is needed that relates the density of toner particles to the average or expected paper coverage. The derivation of average coverage will begin with toner being placed on a piece of paper with uniform distribution as was described by Kendall and Moran [8]. This is then expanded to allow for an arbitrary density of toner particles.

3.1   Uniform toner density

The toner is to be placed onto a region of paper that is square with unit length sides. The toner is restricted such that the center of the toner particle must lie within this unit square region. The probability of any point of unit square being covered by at least one toner particle is

$$cvg = P(\text{area filled by toner}) \tag{1}$$
$$= 1 - P(\text{area not filled by toner})$$
$$= 1 - P(\text{no circle covers each point on the unit square}).$$

Let C_i represent the ith toner particle, and SNC represent 'surface not covered'. Then the equation describing the amount of coverage in the unit square will be

$$cvg = 1 - P(\text{SNC by C}_1 \text{ AND SNC by C}_2 \text{ AND ... AND SNC by C}_N) \tag{2}$$
$$= 1 - \prod_{i=1}^{N} P(\text{SNC by } C_i)$$

Since each toner particle is the same size and shape, and is not connected to each other, each toner particle is placed independently of each other toner particle. So

$$P(\text{SNC by C}_1) = P(\text{SNC by C}_2) = \ldots = P(\text{SNC by C}_N) . \tag{3}$$

Then

$$cvg = 1 - (P(\text{SNC by C})^N) \tag{4}$$
$$= 1 - (1 - P(C))^N$$
$$= 1 - \left(1 - \frac{\text{area of toner particle}}{\text{area of W}}\right)^N.$$

Toner particles are assumed to be circular with a radius r. The potential area (W) that could be covered by any piece of toner is $1 + 4r + r^2\pi$. The area of each piece of toner particle is $r^2\pi$, so

$$cvg = 1 - \left(1 - \frac{\pi r^2}{1 + 4r + \pi r^2}\right)^N. \tag{5}$$

Next consider an area of dimension L by L. The region that could contain part of a circular toner particle will now be W_L, with area

$$\text{area of } W_L = L^2 + 4Lr + \pi r^2. \tag{6}$$

Then through a similar derivation, the average coverage will be

$$cvg = 1 - \left(1 - \frac{\text{area of toner particle}}{\text{area of W}_L}\right)^N \tag{7}$$
$$= 1 - \left(1 - \frac{\pi r^2}{L^2 + 4Lr + \pi r^2}\right)^N$$

4

3.2   Non-uniform toner density

When considering arbitrary toner densities, the paper boundary effects must be removed. To do this, the limit is taken as the length of the paper approaches infinity:

$$cvg = \lim_{L \to \infty} \left[1 - \left(1 - \frac{\pi r^2}{L^2 + 4Lr + \pi r^2} \right)^N \right] \tag{8}$$

and the particle density is defined

$$\rho = N/L^2 , \tag{9}$$

so

$$cvg = \lim_{L \to \infty} \left[1 - \left(1 - \frac{\pi r^2}{L^2 + 4Lr + \pi r^2} \right)^{L^2 \rho} \right] \tag{10}$$

$$= 1 - \lim_{L \to \infty} \left[\left(1 - \frac{1}{\frac{L^2 + 4Lr}{\pi r^2} + 1} \right)^{\frac{L^2}{\pi r^2}} \right]^{\rho \pi r^2}$$

$$cvg = 1 - e^{-\pi r^2 \rho} . \tag{11}$$

While the density was initially restricted to a constant or uniform value, here it can vary as a function of position on the paper $\rho(x,y)$. This can be used to predict the amount of paper covered by toner particles for an arbitrary density of toner.

4. Simulation vs. Theory

The density of the toner is assumed to be dictated by the charge on the PC. This will be determined by the path the laser traces and the rate at which the PC is discharged by the laser. Based on [11][12][13] it is known that the laser discharges the PC with a Gaussian distribution from the center of the laser trace. First the number of images that need to be averaged is evaluated, then several different laser traces are considered.

4.1   Averaging

The equations formulated in the previous section can be used to generate a simulation of a single printing of an image on paper. The coverage equation represents the average coverage per unit area. So this section will investigate how many simulated images will need to be created and averaged so that the average of the simulated images will compare favorably to the coverage equation, Equation 11, formulated in Section 3.

The image used in this experiment is a single pixel. One simulation consists of starting with a clean sheet of "paper" and randomly placing toner particles on the "paper." The paper with the image is set aside so that another simulation can be made. After the second image on paper was made, it was added to the previous simulated paper image. This continued until the specified number of simulated images were made. The average image is calculated from the sum of simulated images by dividing by the number of simulations, N_{sim}. The absolute difference between the coverage equation and the average image was calculated and the difference is noted in Table 2. The values for N_{sim} used for this experiment were 10, 100, 1000, 10,000, and 100,000.
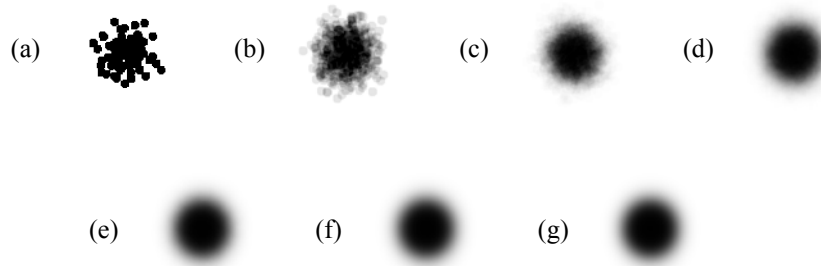
Figure 6:   The grayscale images with increasing number of simulations contributing to the average (a) 1 (b) 10 (c) 100 (d) 1000 (e) 10,000 (f) 100,000. The theoretical coverage is shown in (g).

Table 2: The maximum error for each number of simulations

N_{sim}	10	100	1,000	10,000	100,000
Max Difference	0.527	0.186	0.058	0.020	0.015

Figure 6a shows one simulated image of placement of toner particles on paper. The simulated image of toner particles on paper appears similar to the magnified images. The gray scale image in Figure 6b illustrates the average of ten simulated images. The average image appears to be a crude form of the theoretical average coverage shown in Figure 6g. It can clearly be seen that the image is randomly changing each time. Figure 6c is closer to the gray scale image of the theoretical average coverage. The gray scale images in Figure 6 illustrate that as the number of simulations increases, the gray scale image more closely resembles the theoretical average coverage. Between Figure 6e, 6f, and 6g, the naked eye cannot distinguish the difference between the images.

4.2   Comparison of other source image patterns

Experiments were conducted to generate a charge pattern for several different source images. These are a single pixel, pixels aligned horizontally, pixels aligned vertically, pixels at a 45 degree angle, and a line at 20 degrees. For each source pattern, images are created representing both the theoretical coverage amount and average images calculated from averaging 100,000 instances of images created from randomly placed toner particles drawn from the specified toner distribution probabilities. The single pixel was shown in Figure 6. Results for four other cases are shown in Figure 7. Table 3 displays the maximum absolute error. In all cases the error was small and the average image closely matches the derived coverage equations.

Table 3: Error between coverage and averaged toner placement.

	single pixel	horizontally aligned pixels	vertically aligned pixels	45 degree line	20 degree line
Absolute Error	0.0145	0.0154	0.0197	0.0137	0.0181

5. Conclusion

Using the probabilistic placement of toner particles, a deterministic model of the amount of paper expected to be covered by toner has been developed. The output of this model has a visible appearance that

Figure 7:   (a) Average images and (b) theoretical coverage for a three horizontal pixel, three pixels with vertical alignment, three pixels in 45 degree diagonal and a 20 degree line.

resembles toner images when viewed with a microscope, and matches simulations of images created with these statistical properties.

This model can be used as the input to an imaging system to see the optical response expected from text generated by the electrophotographic (laser) printing process. The combination of this model with prior modeling and analysis completed by the second author will enable document image defects to be better understood and for OCR systems to compensate for imaging defects in printed and photocopied documents.

6. References

1. Henry S. Baird, "Calibration of document image defect models," *Proc. of Second Annual Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, April 1993, pp. 1-16.
2. Elisa H. Barney Smith, "Characterization of Image Degradation Caused by Scanning," *Pattern Recognition Letters*, Volume 19, Number 13, 1998, pp. 1191-1197.
3. E. H. Barney Smith, "Scanner Parameter Estimation Using Bilevel Scans of Star Charts," Proc. International Conference on Document Analysis and Recognition 2001, Seattle, WA, 10-13 September 2001, pp. 1164-1168.
4. Farhan A. Baqai and Jan P. Allebach, "Printer Models and the Direct Binary Search Algorithm," Proc. ICASSP, Seattle, WA, May 1998, pp. 2949-2952.
5. J.D. Foley, A. vanDam, S.K. Feiner and J.F. Hughes, "Computer Graphics: Principles and Practice," Addison-Wesley 1996.
6. Tin Kam Ho and Henry S. Baird, "Large-Scale Simulation Studies in Image Pattern Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 10, October 1997, pp. 1067-1079.
7. D. Kacker, T. Camis, J.P. Allebach, "Electrophotographic Process Embedded in Direct Binary Search," Proc. SPIE Color Imaging: Device-Independent Color, Color Hardcopy and Graphic Arts V, San Jose, CA, January 2000, pp 468-482.
8. M.G. Kendall and P.A.P. Moran, *Geometrical Probability,* Hafner Publishing Company, 1963.
9. Theo Pavlidis, Minghua Chen, and Eugene Joseph, "Sampling and Quantization of Bilevel Signals," Pattern Recognition Letters, Vol. 14, July 1993, pp. 559-562.
10. M.J. Stanich, "Print-quality enhancement in electrophotographic printers," IBM Journal of Research & Development, Vol. 41, No. 6, 1997, pp. 66-6789.
11. E.M. Williams, *The Physics and Technology of Xerographic Processes*, Krieger Publishing Company, 1993.
12. J. Yi, *A Xerographic Simulation Model*, MS Thesis, University of Idaho, May 1999.
13. A. Vongkunghae, J. Yi, R. Wells, "A Printer Model Using Signal Processing Techniques," IEEE Trans. Image Processing, Vol 12, No. 7, July 2003, pp 776-783.