Show-Through Cancellation in Scans of Duplex Printed Documents

Gaurav Sharma, *Senior Member, IEEE*

Abstract—In scanning pages with double-sided printing, often the printing on the back-side shows through in the scan of the front-side because the paper is not completely opaque. This *show-through* is an undesirable artifact that one would like to remove.

In this paper, the phenomenon of show-through is analyzed using first physical principles to obtain a simplified mathematical model. The model is linearized using suitable transformations and simplifying approximations. Based on the linearized model, an adaptive linear filtering scheme is developed for the electronic removal of show-through using scans of both sides of the document. Experimental results demonstrating the effectiveness of the method developed are presented.

Index Terms—Adaptive filtering, restoration, scanning, show-through.

I. INTRODUCTION

P RINTED documents are frequently captured as digital images for use in reproduction, communication, or automated processing. Examples of applications that involve image capture for these different purposes are copying, faxing, electronic document distribution and archival, optical-character-recognition (OCR), and electronic database storage for automating search and retrieval. A scanner is the most common device employed for the capture and digitization of hardcopy documents.

The image captured by a scanner is a two-dimensional (2-D) array of pixels, where each pixel value represents the reflectance of the document at the physical location corresponding to that pixel. The most common type of scanner is a flat-bed scanner. Fig. 1 is a schematic of the optical components of a flat-bed scanner. The document to be scanned is laid face down on a transparent glass platen and pressed flat against the platen by a backing. The scanner lamp illuminates the document through the platen glass and the light reflected off the spatial location corresponding to a given pixel is imaged by a lens onto a sensor. The resulting signal is digitized to obtain a representation of the image as a reflectance profile. Typically, the sensors are laid out in a linear CCD array, which allows for an entire row of pixels along one dimension of the document to be imaged in a single exposure step. The array is moved across the document and multiple exposures are performed to capture the complete 2-D image.

Images acquired using a scanner suffer from several degradations. Examples of these degradations include scanner noise
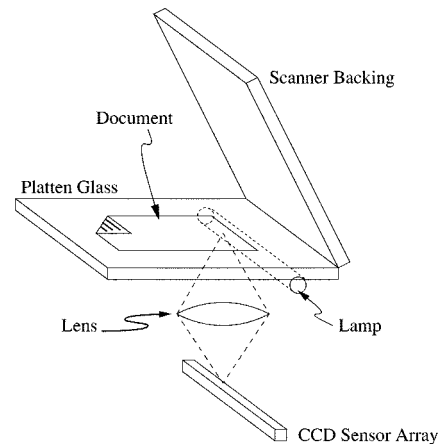
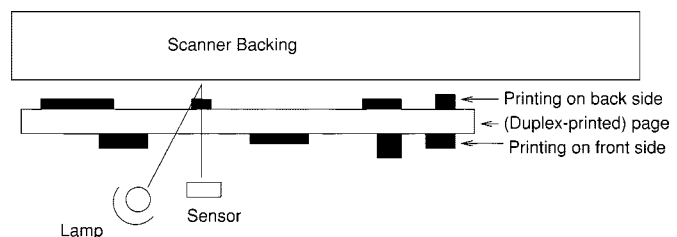Fig. 1. Schematic of optical components of a flatbed scanner.

Fig. 2. Scan of duplex printed page.

from various sources [1], optical-blur due to the limited bandwidth of the scanner modulation transfer function (MTF) [2], [3], jitter and other artifacts due to motion errors of the CCD array [2], [3], and optical integrating cavity effect [4]. A number of image processing techniques have been developed in order to enhance/restore the degraded images before they are displayed, archived or processed further [4], [5, Ch. 4], [6].

Hardcopy documents are often printed in duplex (double-sided) mode, with printed information on both sides of the page (common examples are most magazine and book pages). When a duplex printed page is scanned, information from the back-side printing can often be seen in the scan (of the front side of the page). We refer to this common artifact encountered while scanning duplex (double-sided) printed pages as *show-through* (since the text/image on the back side "shows-through" the paper). Fig. 2 graphically illustrates the process of scanning one side of a duplex printed page on a typical scanner, where the arrangement of the duplex printed page of paper in relation to scanner lamp, sensor, and backing is shown in a cross-sectional view. From the figure, it is clear that if the page is not completely opaque and the scanner uses a white backing behind the page, the sensor receives some

The Document Company Xerox

Xerox offers unrivaled expertise in the production and management of documents: color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.

Xerox and its partner, Fuji Xerox Co., Ltd. of Japan. offer the broadest array of document products and services in the industry: copiers, printers, fax machines, scanners, desktop software, digital printing and publishing systems, supplies, and comprehensive document-management services - from the running of in-house production centers to the creation of networks.

Fig. 3.   Original image printed on side one of a two-sided page.

Fig. 4.   Original image printed on side two of a two-sided page.

light that is transmitted through the paper, reflected from the backing and transmitted back through the paper. If there is no printing on the back side of the paper, this light produces no undesirable effect as the paper and the backing can be thought of as the effective substrate for the printing on the front side. If, however, the back side also has printing, then due to the transmission through the paper, the scan of the front side contains a residual (transposed) image of the back side. This contribution from the back side printing to the scan of the front side is the *show-through*. Since the transmittance of paper is low (in relation to its reflectance), in the scanned image, the dynamic range (i.e., contrast) of the show-through information is typically much lower than the dynamic range of the printing on the front side.

Figs. 3 and 4 represent the original "images" printed on two sides of a page and Figs. 5 and 6, respectively, are the images obtained by scanning the two sides of the duplex printed page with a white backing. In each scan, show-through from the (corresponding) back side can be clearly seen as a low contrast transposed image of the printing on the back side. The show-through is most visible in the regions in which there is no printing on the front side, though it can also be seen in light gray regions of the image side.

Show-through is clearly an unwanted artifact that one would like to eliminate. If the original information consists of simple black and white printed text, show-through can be removed by a simple process of thresholding. The thresholding converts pixels whose scanned reflectance is above a specified threshold value to white. This process converts the light gray regions resulting from the low-contrast back-side show-through to white and does not significantly influence the front-side text which is close to black.

The thresholding method, however, does not work for prints containing images or text that have additional gray levels beyond pure black and white. For regions with light gray printing on the front side, the thresholding either does not remove the show-through, or inadvertently eliminates the information in these regions by converting them to white.

Prior to this paper, image enhancement and restoration techniques typically did not account for show-through. This paper presents image processing methods for electronic compensation of show-through [7]. The rest of this paper is organized as follows. Section II presents an analysis of show-through from first physical principles that yields a nonlinear mathematical model for the show-through phenomenon. In Section III, the model is linearized through suitable transformation and simplifying approximations. The notion of a show-through point spread function is introduced in Section IV to account for the spreading of light in the paper substrate. An algorithm for show-through correction based on adaptive linear filter theory is presented in
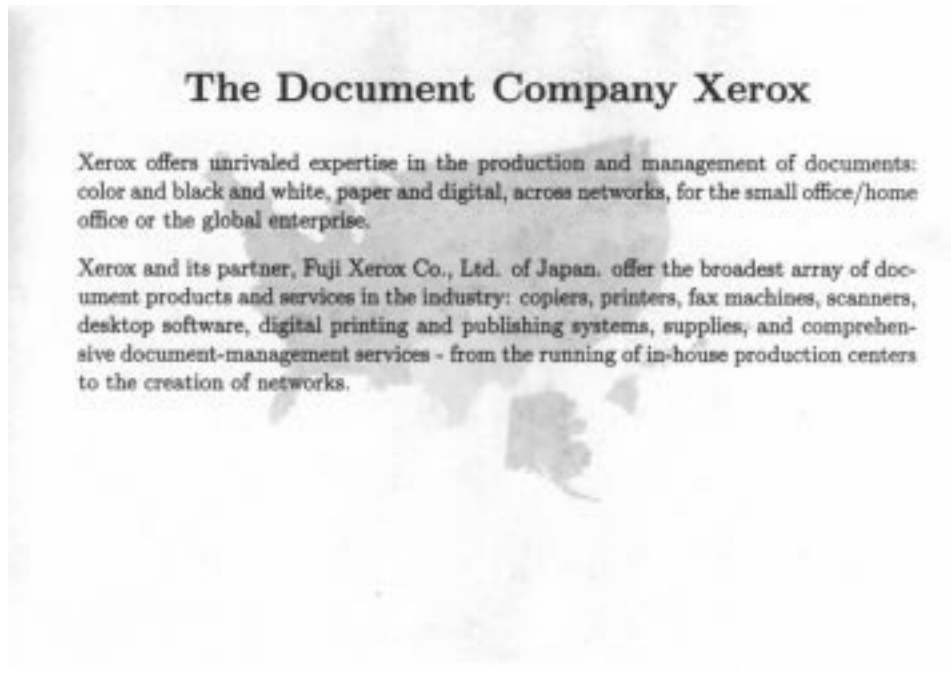
The Document Company Xerox

Xerox offers unrivaled expertise in the production and management of documents: color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.

Xerox and its partner, Fuji Xerox Co., Ltd. of Japan, offer the broadest array of document products and services in the industry: copiers, printers, fax machines, scanners, desktop software, digital printing and publishing systems, supplies, and comprehensive document-management services - from the running of in-house production centers to the creation of networks.

Fig. 5. Scan of side one of a two-sided page, demonstrating show-through from side two.

Fig. 6. Scan of side two of a two-sided page, demonstrating show-through from side two.

Section V. Experimental results demonstrating the effectiveness of the methods on actual scans are presented in Section VI and conclusions in Section VII.

II. SHOW-THROUGH MODEL AND ANALYSIS

Consider the simplified cross-sectional view of a scanner shown in Fig. 2, where a duplex printed document is being scanned. If there is no printing on either side, the light reaching the sensor has two main components, light that is scattered by the paper substrate and light that is transmitted through the paper, reflected by the backing and transmitted again through the paper. Thus the reflectance detected by the optical sensor for a white unprinted page is given by

$$R_p^w = S_p + T_p^2 R_{bk} \qquad (1)$$

where

S_p fraction of light scattered by the paper (in the forward direction);

T_p transmittance of paper;

R_{bk} reflectance of the backing.

Note that the superscript w has been used to indicate that R_p^w is the reflectance for "white" paper that has no printing on either side. From (1), one can see that the reflectance of unprinted paper on black backing ($R_{bk} = 0$) is given by S_p and that on white backing ($R_{bk} = 1$) is given by $S_p + T_p^2$.

When the scanned page has printed material on either side, the printing may be viewed as separate layers on either side of the page, with the printed information represented as the spatial transmittance profile of these layers. If $T_f(x, y)$ represents the transmittance of the "print layer" on the front side and $T_b(x, y)$ represents the transmittance of the "print layer" on the back side, the reflectance detected by the sensor when scanning the front side is given by

$$R_f^s(x, y) = T_f^2(x, y) \left(S_p + T_p^2 R_{bk} T_b^2(x, y) \right) \qquad (2)$$

where

x and y 2-D spatial coordinates on the paper being scanned;

subscript f front side;

superscript s denotes that this is the "scanned" reflectance.

Equation (2) indicates that the reflectance $R_f^s(x, y)$ detected by the scanner sensor depends on the front-side print layer transmittance, the paper scattering and transmittance parameters, the reflectance of the backing, and the back-side print layer transmittance. In particular, the dependence of the scanned reflectance $R_f^s(x, y)$ for the front side on the transmittance $T_b(x, y)$ of the back side print layer represents the undesired show-through in the front side scan.

The goal of show-through correction is to recover the "image" $T_f(x, y)$ printed on the front side with no dependence on the back side "image" $T_b(x, y)$. Recovery of the show-through corrected image cannot be accurately done using the front-side scan alone because it is not possible to reliably distinguish between light gray printing on the front-side and low-contrast show-through from the back side. Mathematically, this is manifested in the fact that the single equation (2) cannot be solved simultaneously for the two unknowns $T_f(x, y)$ and $T_b(x, y)$.

If the scan of the back-side is also available, then analogous to (2), the scanned reflectance for the back side can be written as

$$R_b^s(x, y) = T_b^2(x, y) \left(S_p + T_p^2 R_{bk} T_f^2(x, y) \right) \qquad (3)$$

where the terms on the right hand side are as defined earlier.

Observe that (2) and (3) represent two equations in the two unknowns $T_f(x, y)$ and $T_b(x, y)$. Therefore one can expect to remove the show-through if scans of both sides of the page are available and if these two equations can be solved for these unknowns. There are however, several obstacles that need to be overcome before this can be done. These will be addressed in the following sections.

III. LINEARIZED SHOW-THROUGH MODEL

The nonlinear nature of equations (2) and (3), and the lack of knowledge of the parameters S_p and T_p make an analytic or numerical solution impractical in their present form. In order to simplify these equations, it is advantageous to use the notion of a show-through corrected image. For the purposes of the discussion in this paper, it will be assumed that the "show-through corrected image" to be recovered from the scan, is the image that would have been obtained from the scanner if there was no printing on the back side of the scanned page (alternate definitions of the "show-through corrected image" are also possible, one benefit of this definition will be indicated in Section V). The assumption that there is no printing on the back-side is mathematically equivalent to setting the transmittance $T_b(x, y)$ of the back side print layer to unity in (2). Making this substitution the show-through corrected reflectance for the front side is given by

$$R_f(x, y) = T_f^2(x, y) \left(S_p + T_p^2 R_{bk} \right) \qquad (4)$$

and similarly, the show-through corrected reflectance for the back side is given by

$$R_b(x, y) = T_b^2(x, y) \left(S_p + T_p^2 R_{bk} \right). \qquad (5)$$

Dividing equation (2) by the reflectance of white (unprinted on either side) paper in (1), and taking the negative natural logarithm of the resulting equation we obtain

$$
\begin{aligned}
D_f^s&(x, y) \\
&\equiv -\ln \frac{R_f^s(x, y)}{R_p^w} \qquad\qquad\qquad\qquad\qquad (6) \\
&= -\ln \left(\frac{T_f^2(x, y) \left(S_p + T_p^2 R_{bk} T_b^2(x, y) \right)}{S_p + T_p^2 R_{bk}} \right) \\
&= -\ln(T_f^2(x, y)) - \ln \left(1 - \frac{T_p^2 R_{bk} \left(1 - T_b^2(x, y) \right)}{S_p + T_p^2 R_{bk}} \right) \\
&= D_f(x, y) - \ln \left(1 - \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} \left(1 - T_b^2(x, y) \right) \right) \quad (7)
\end{aligned}
$$

where

$$D_f(x, y) \equiv -\ln \frac{R_f(x, y)}{R_p^w} = -\ln \left(T_f^2(x, y) \right). \qquad (8)$$

It may be noted here that the negative logarithm of the reflectance (or transmittance) is by definition the *optical density*.[1] Thus, the operation of dividing by the paper reflectance and taking the natural logarithm is a conversion from reflectance into (paper normalized) densities and the terms $D_f^s(x, y)$ and $D_f(x, y)$ in the above equations represent the (paper normalized) density of the front side seen by the scanner and the (paper normalized) density of the show-through corrected front side.

For typical paper substrates, the fraction of light transmitted is much smaller than the fraction of light scattered, i.e., $T_p^2 \ll S_p$. This assumption is directly supported by the observation that most paper substrates appear close to white even when placed on

---

[1]Normally, the logarithm to the base 10 is used in defining optical density but the natural logarithm is equally valid and is used here for notational simplicity.

a black backing [see (1)]. This assumption allows a significant simplification of (7) using the approximation

$$\ln(1-t) \approx -t \qquad \text{for } |t| \ll 1 \tag{9}$$

with

$$t = \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} \left(1 - T_b^2(x, y)\right).$$

With this approximation, (7) becomes

$$D_f^s(x, y) \approx D_f(x, y) + \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} \left(1 - T_b^2(x, y)\right)$$

$$= D_f(x, y) + \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} A_b(x, y) \tag{10}$$

where

$$A_b(x, y) \equiv \left(1 - T_b^2(x, y)\right) = 1 - \frac{R_b(x, y)}{R_p^w} \tag{11}$$

is the absorptance of the back side "print layer."

Using the same transformation and simplification for (3), the corresponding equation for the back side is obtained as

$$D_b^s(x, y) \equiv -\ln \frac{R_b^s(x, y)}{R_p^w}$$

$$\approx D_b(x, y) + \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} A_f(x, y) \tag{12}$$

where

$$D_b(x, y) \equiv -\ln \frac{R_b(x, y)}{R_p^w} = -\ln \left(T_b^2(x, y)\right) \tag{13}$$

is the paper-normalized density for the show-through corrected back-side and

$$A_f(x, y) \equiv \left(1 - T_f^2(x, y)\right) = 1 - \frac{R_f(x, y)}{R_p^w} \tag{14}$$

is the absorptance of the front side "print layer."

Equation (10) states that the paper-normalized density of the front side scan can be approximated by the sum of the paper-normalized density of the show-through corrected front side and the absorptance of the back-side print layer weighted by a small factor. It is clear that the second term represents the show-through. Equation (12) can be similarly interpreted for the back side. The major significance of these equations is that in the density domain the show-through separates into an additive distortion, which is further characterized as being a scalar multiple of the opposite side print layer's absorptance.

IV. SHOW-THROUGH POINT SPREAD FUNCTION

The show-through model developed so far in (10)–(12) is based on an extremely simplified physical view of the show-through phenomenon. One serious shortcoming of the simplification is its assumption of purely point-wise spatial interaction between the front and the back side images. In reality, spreading of light in the paper causes blurring of the show-through image. The spreading of light in paper has been studied extensively in the context of its impact on the reflectance of printed images [8]–[10]. For the show-through model, the blurring can be incorporated into the show-through model, by replacing the scalar

term $T_p^2 R_{bk}/(S_p + T_p^2 R_{bk})$, in (10) and (12) with an empirical "show-through" point spread function (PSF) to obtain

$$D_f^s(x, y) = D_f(x, y) + h(x, y) \otimes A_b(x, y) \tag{15}$$

$$D_b^s(x, y) = D_b(x, y) + h(x, y) \otimes A_f(x, y) \tag{16}$$

where $h(x, y)$ is the "show-through point spread function" and $\otimes$ represents the convolution operator. Since show-through PSF is a replacement for the term $T_p^2 R_{bk}/(S_p + T_p^2 R_{bk})$ in the physical model, it is clear that $h(x, y)$ is small in comparison to unity and physically accounts for the transmission and spreading of light in the paper and the reflectance of the backing.

In order to "solve" (15) and (16) for the show-through corrected densities $D_f(x, y)$ and $D_b(x, y)$, the absorptances for the back and front side "image layers," i.e., $A_b(x, y)$ and $A_f(x, y)$ may further be approximated by the corresponding absorptances from the scans, to obtain

$$D_f(x, y) = D_f^s(x, y) - h(x, y) \otimes A_b^s(x, y) \tag{17}$$

$$D_b(x, y) = D_b^s(x, y) - h(x, y) \otimes A_f^s(x, y) \tag{18}$$

where

$$A_b^s(x, y) \equiv 1 - \frac{R_b^s(x, y)}{R_p^w} \tag{19}$$

$$A_f^s(x, y) \equiv 1 - \frac{R_f^s(x, y)}{R_p^w} \tag{20}$$

$$\tag{21}$$

are the absorptances corresponding to the back and front side scans respectively. Note that this approximation can be avoided through an iterative use of (17) and (18), where the recovered show-through images from the previous iteration are used in computing $A_b(x, y)$ and $A_f(x, y)$ for the next iteration. The impact of the approximation above and in (9) is analyzed more completely in the Appendix.

In actual practice, the scanned images are in fact sampled over a discrete grid and (17) and (18), need to be approximated by their discrete versions

$$D_f(m, n) = D_f^s(m, n) - h(m, n) \otimes A_b^s(m, n) \tag{22}$$

$$D_b(m, n) = D_b^s(m, n) - h(m, n) \otimes A_f^s(m, n) \tag{23}$$

where m and n represent the sample indices for samples along the x and y spatial dimensions, respectively, and the other terms are as before.

V. SHOW-THROUGH CORRECTION ALGORITHM

In order to use the "solution" of (22) and (23), the show-through point spread function should be known *a priori* and the relative alignment of the images of the front and the back-side should be known precisely. Since neither of these requirements are satisfied in practice, these equations cannot be directly used for the cancellation of show-through. Note however, that the linearity of these equations implies that the problem of show-through cancellation can be viewed as the 2-D equivalent of the one-dimensional (1-D) echo-cancellation problem in speech telephony. Methods from linear filter theory applied to the problem of speech telephony can therefore be adapted to the problem of show-through cancellation. In particular, adaptive
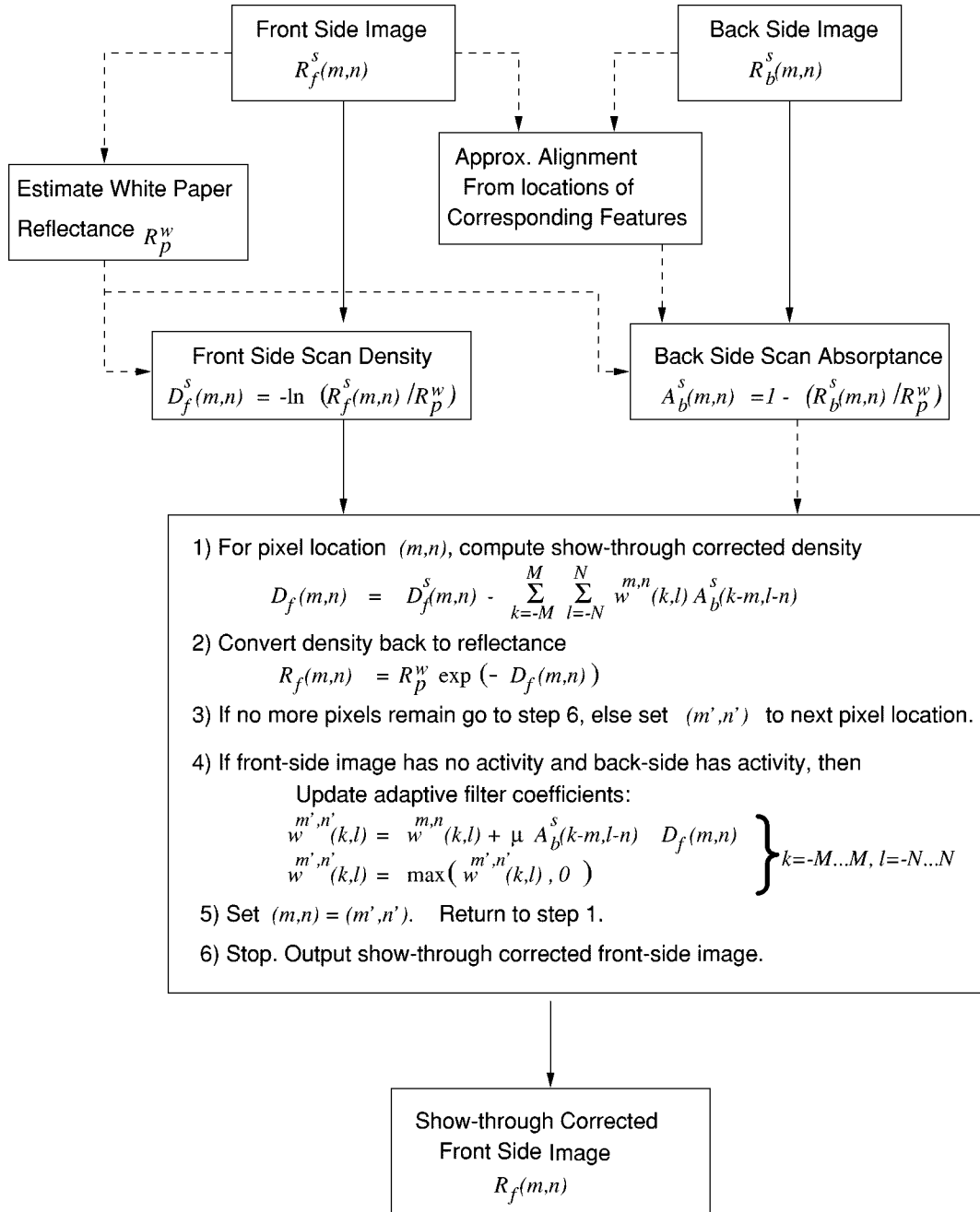
Fig. 7.   Show-through cancellation algorithm.

linear filters can be used for automatically estimating and tracking the show-through PSF. For the following description, it is assumed that the show-through is to be cancelled from the "front-side" scan. The processing for the back side can be similarly described.

The complete process of show-through correction is described as follows. First, approximate alignment for the front and back side scans is determined by identifying corresponding image features in the front side show-through and the back side scan. The reflectance R_p^w of white paper unprinted on both sides (with the scanner backing) is estimated by averaging the reflectance values over a region of the scan that has no printing on either side. Note that the fact that no additional parameters other than R_p^w (such as the paper scattering fraction S_p or the

paper transmittance T_p or the reflectance of the backing R_{bk}) are required for the show-through correction algorithm is a direct result of the definition of the show-through corrected image that has been adopted here. The definition was chosen for this very reason.

Using the estimate of R_p^w, scan data from the front side is converted to density relative to paper white as per (6) and data for the back side is converted to absorptance using (19). The image pixels are processed one at a time in an order that preserves spatial contiguity, for example, by processing along a serpentine raster. Maintaining spatial contiguity in the processing ensures that the inevitable changes in alignment between the front and back side images over the page can be tracked more readily by the adaptive filter.

The Document Company Xerox

Xerox offers unrivaled expertise in the production and management of documents: color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.

Xerox and its partner, Fuji Xerox Co., Ltd. of Japan, offer the broadest array of document products and services in the industry: copiers, printers, fax machines, scanners, desktop software, digital printing and publishing systems, supplies, and comprehensive document-management services - from the running of in-house production centers to the creation of networks.

Fig. 8. Side one scanned data after show-through correction.

Fig. 9. Side two scanned data after show-through correction.

At each pixel location (m, n), the show-through corrected density is computed as

$$D_f(m, n) = D_f^s(m, n) - \sum_{k=-M}^{M} \sum_{l=-N}^{N} w^{m, n}(k, l) A_b^s(k - m, l - n) \quad (24)$$

where $w^{m, n}(k, l)$ is a 2-D adaptive finite-impulse-response (FIR) filter with support $[-M \cdots M] \times [-N \times N]$ that represents the show-through point spread function.

The scanned image values for the front and back side images are then examined and compared to the paper white reflectance to determine if the local neighborhood about the current pixel location (and including it) contains any printing on either side. If the local neighborhood has printing on the back-side but no printing on the front-side (this corresponds to "far-end single-talk" in the speech-telephony echo-cancellation analogy), the filter coefficients $w^{m, n}(k, l)$ are adapted. The adaptation of the filter coefficients can be done in accordance with any of the
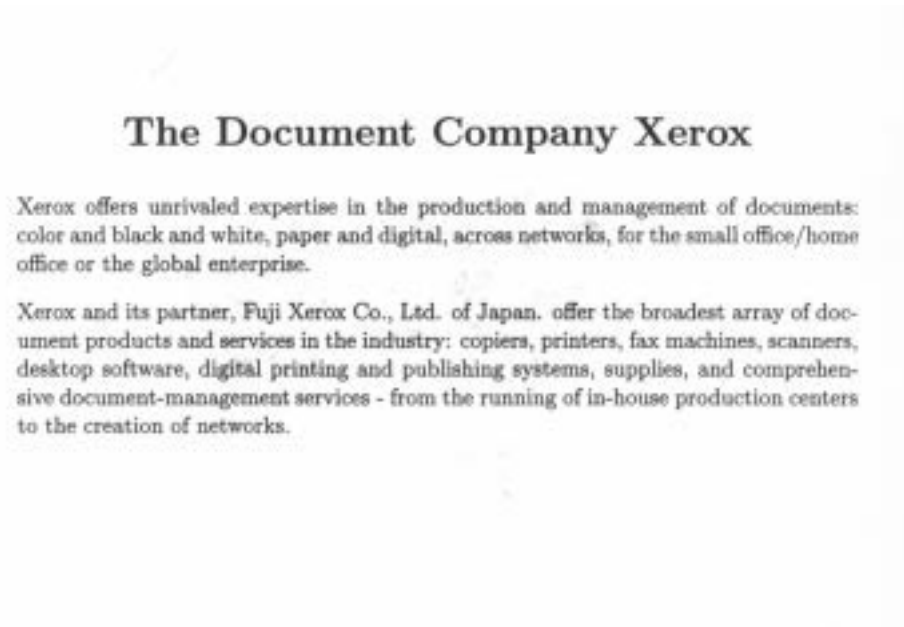
Fig. 10. Side one scanned data after thresholding.

Fig. 11. Side two scanned data after thresholding.

several known algorithms in adaptive filter theory [11], [12]. For the following description, the simplest and perhaps best-known least-mean square (LMS) algorithm [11, p. 302] will be used. For the LMS algorithm, the filter coefficients for the next pixel location (m', n') are computed as

$$w^{m',n'}(k, l) = w^{m,n}(k, l) + \mu D_f(m, n) A_b^s(k - m, l - n)$$
$$k = -M \cdots M, l = -N \cdots N \qquad (25)$$

where μ is the LMS adaptation step-size parameter. Note that at pixels where the filter coefficients are adapted, there is no printing on the front side, therefore, the desired show-through corrected density at the pixel (m, n) is zero and the value $D_f(m, n)$ computed in (24) represents the error term used in the more general specification of the LMS algorithm [11, p. 302]. Since the adaptive filter represents the show-through PSF which is expected to be nonnegative [because it represents the physical quantity $T_p^2 R_{bk}/(S_p +$

The Document Company Xerox

Xerox offers unrivaled expertise in the production and management of documents: color and black and white, paper and digital, across networks, for the small office/home office or the global enterprise.

Xerox and its partner, Fuji Xerox Co., Ltd. of Japan. offer the broadest array of document products and services in the industry: copiers, printers, fax machines, scanners, desktop software, digital printing and publishing systems, supplies, and comprehensive document-management services - from the running of in-house production centers to the creation of networks.

Fig. 12.   Side one scanned data after show-through correction and thresholding.

Fig. 13.   Side two scanned data after show-through correction and thresholding.

$T_p^2 R_{bk}$)], any coefficients that turn negative as a result of the adaptation are truncated to zero. Note that since the white reflectance R_p^w is an estimate based on an average of a pixel region, the actual scanned reflectance values at the different pixels will vary above and below this average. The computed values for D_f and A_b^s therefore take on both negative and positive values. This allows the adaptation in (25) to occur in either the positive or negative direction.

The processing is then repeated for the next pixel location, continuing until the complete image has been processed. The show-through corrected density for the front-side $D_f(m, n)$ is

converted to reflectance by inverting the relation of (8) to obtain the output show-through corrected image. Note that the show-through correction of (24) is applied at each pixel, though the adaptation is performed only in regions where the back-side has printing and the front-side does not. Since the desired output is not known for regions with printing on the front side, the filter coefficients cannot be adapted in those regions. No adaptation is performed for regions that have no printing on either side to keep the filter coefficients from drifting due to the noise (and no signal) in these regions. The complete show-through cancellation algorithm is summarized in Fig. 7.

VI. EXPERIMENTAL RESULTS

The show-through model and the cancellation algorithm developed here were tested using a UMAX powerlook desktop scanner. The scanner has an optical resolution of 600 dots-per-inch (dpi) and is capable of scanning in 24-bit color or 8-bit gray-scale. For the purposes of the experiment, the scanner was operated in a linear mode (i.e., gamma of 1.0), in which the scanner output is linearly related to scanned reflectance.[2] A duplex printed page with the printing on the two sides as shown in Figs. 3 and 4 was used as the input. The two printed sides of the page were scanned at 600 dpi resolution as 8-bit gray-scale images. In scanning the images, care was taken to minimize skew by keeping the upper edge of the page closely aligned with the edge of the scanner platen (the effect of a relative skew between the front and back side image on the algorithm will be discussed later in this section). The resulting scanned images have been presented earlier in Figs. 5 and 6, and will be referred to as the "text-side" and "image-side" scans, respectively. In both images, show-through from the other side, can clearly be seen as mentioned earlier.

The relative alignment between the scanned images for the two sides was obtained by locating the position of the intersection of the two lower 45° stroke lines at the bottom of the upper case "X" on the first line (the first "X" in "The Document Company Xerox"). In actual applications, if an automatic document feeder is utilized for scanning the page in duplex mode, the relative alignment between the images on the two sides can be obtained from the feeder's geometry and the detected paper edges. Alternately, techniques from image registration [13], [14] may be adapted for aligning the images.

For the images in Figs. 5 and 6, the (scaled) reflectance R_p^w of white paper unprinted on either side was computed by averaging the scanned values over a 400 pixel by 400 pixel square in the top corner of the image-side scan. This average value was 250.56 (for the 8-bit scan with pixel digital values ranging between 0 and 255). In order to estimate the level of show-through, the average value over a region with black printing on the back side and no printing on the front sides was also computed and found to be 245.68. Note this latter value is not required in the show-through correction algorithm and is included here only to give the reader a quantitative estimate for the magnitude of show-through.

Using the alignment information and the estimated white reflectance, the images were processed using the algorithm of Section V. The sequence for processing pixels was chosen along a serpentine raster, i.e., even scanlines were traversed left-to-right and odd scanlines were traversed right-to-left. For the processing, the filter size was chosen to be 31×31 (i.e., $M = N = 15$), the filter was initialized to all zeros at the start, and the LMS adaptation parameter μ was set to 0.001. For checking for the presence of printing on either side (detecting "far-end single talk"), a 15×15 neighborhood was used. Printing was deemed present on a given side if the minimum value over the neighborhood was below 75% of the estimated white reflectance R_p^w (a digital value of 190 for the example presented here).

Fig. 14. Adaptive filter coefficients $w^{m,n}(k, l)$ at two different image locations.

The show-through corrected images for the text-side and the image-side obtained from the show-through cancellation algorithm developed are shown in Figs. 8 and 9, respectively. The results clearly indicate that the algorithm is successful in cancelling show-through. In comparison to the original scans, both the show-through corrected images have no apparent show-though and very minor processing artifacts (these minor artifacts can be seen better in the electronic versions of these images when displayed on a cathode-ray-tube (CRT) monitor).

As mentioned in the introduction, simple thresholding provides an efficient way for elimination of show-through from black and white text images. The thresholding method was tested on the images used in the above example. The images were thresholded, so as to convert values above 200 to white i.e., 255. This threshold was determined by progressively reducing the threshold value starting at the average for a region with printing on the back side alone (which as noted earlier, was 245.68), till the show-through was significantly reduced and text was not significantly degraded. The results of thresholding the scanned images for the text and the image-side in this

---

[2]Obviously, the algorithm can be used for any known scanner gamma setting with appropriate pre and post conversions. The linear mode was chosen to keep the presentation simpler.

Fig. 15. Scan of side two of a two-sided page, demonstrating show-through.

manner are shown in Figs. 10 and 11, respectively. Note that the thresholding method eliminates most of the show-through over the text-side. Over the image-side, however, the thresholding significantly lightens the light gray regions (the state of Texas in the map) and does not fully eliminate the show-through from these regions. Compared to the printed versions of the images, these artifacts can be seen more clearly when the images are viewed on a CRT display.

The thresholding operation does have the advantage that it reduces the noise present in the white regions since it eliminates most of the variations caused by scanner noise and by changes in the paper reflectance. The white background in Figs. 10 and 11 therefore appears a lot cleaner than in Figs. 8 and 9. The same benefit can be realized for the show-through correction technique presented in this paper by adding thresholding as a final post-processing step. Since the purpose of thresholding in this case is elimination of background variation and not show-through removal, the threshold value used can be significantly higher. For the 400 pixel by 400 pixel square over which the average value of white was computed, the observed standard deviation was $\sigma_w = 5.94$. In order to eliminate the background variation, a threshold value of $226 \approx R_p^w - 4 * \sigma_w$ was used, and all pixels above this threshold were converted to white. The resulting images are shown in Figs. 12 and 13 for the text and image sides, respectively. The background noise in the white regions of

Fig. 16.   Scan of side two of a two-sided page, demonstrating show-through.

these images is significantly reduced. In addition, unlike the directly thresholded images in Figs. 10 and 11, there is no residual show-through in the light gray region of the image (corresponding to the state of Texas).

In order to visualize the show-through PSF, the adaptive filter coefficients $w^{m,n}(k,l)$ were recorded at several positions in the processing of the image-side scan. The filter coefficients at two different image locations are shown in Fig. 14. The filter coefficients have a primarily unimodal distribution, which is consistent with what is expected physically for the show-through PSF. The spatial extent of the significant filter coefficients is considerably smaller than the 31×31 filter used, but the position of the filter coefficient's peak is shifted in the

second plot in relation to the first. Change in relative alignment of the front and back side images for the two image locations is the probable cause of this shift. Note that a skew between the front and back side scans is one example of mis-alignment that would produce a variation in the relative alignment of the front and back side images over the page. Even a small skew angle results in a large pixel shift over the length/width of the page and therefore the adaptive filter needs a large support size in order to track any "shifts" due to skew (for example, at 600 dpi scan resolution, a skew of just $0.2°$ corresponds to a pixel shift of almost eight pixels over a page length of 11.5 in). If, however, the images on the front and the back side can be aligned well with minimal skew (for instance, when an

Fig. 17.   Show-through corrected version of the scan in Fig. 15.

automated document feeder is used for scanning), the support of the adaptive FIR filter can be significantly reduced, thereby obtaining a significant reduction in computation. Since optical distortions and mechanical positioning errors in the scanner also contribute to relative mis-alignment between the front and back side images a perfect global alignment is usually impractical at useful scan resolutions.

The influence of the different parameters used in the show-through correction algorithm was studied experimentally by varying these parameters. The effect of errors in the estimated white reflectance R_p^w was studied by varying over the range from 245 to 255 the value actually used for R_p^w in the conversion to density and absorptance (instead of the actual estimate

value of 250.56). Only minor degradation was observed in the output of the show-through cancellation algorithm due to this variation, indicating that the algorithm is fairly robust to small errors in the estimated white reflectance. Next, the impact of using different size support for the adaptive filter (changing M and N) was investigated. For filter sizes, smaller than 21×21 residual show-through artifacts could be seen in the images after correction. The artifacts observed are even more severe if the adaptive filter coefficients at a given image location are recorded and used as a fixed filter over the entire image. As noted in the last paragraph these artifacts are probably because the smaller areas of support for the adaptive filter do not allow the filters to track any location dependent changes in the align-

Fig. 18.   Show-through corrected version of the scan in Fig. 16.

ment of the front and back side images resulting from relative skew. The impact of changing the adaptation parameter μ was similar to that observed in other applications of the LMS algorithm [11]: larger values of μ provided better tracking but also increased noise and smaller values took longer for the filter to converge and also produced some artifacts due to slower tracking of any changes caused by change in relative alignment.

At the end of Section IV, it was mentioned that the error introduced due to the approximation of the other side's "print-layer absorptance" with the scan absorptance [$A_b(x, y) \approx A_b^s(x, y)$

for the front-side correction] can be reduced by iterating the show-through correction algorithm by using the output of the last show-through correction step as the input images for the next iteration. This iterative scheme was also tested on the images presented in this section and the second application of show-through correction algorithm was seen to reduce any remaining show-through artifacts in the images after the first correction step, when the images in electronic form were displayed on the screen. However, the remaining show-through artifacts were too small for the improvement to be distinguishable in the printed images and has therefore not been included here.

Fig. 19.   Thresholded version of the scan in Fig. 15.

Finally, it is worth noting that the excellent results obtained from the show-through correction algorithm are due to the combination of a good physical model with the adaptive linear filtering and not exclusively due to the use of adaptive filters. If the model is not explicitly used, i.e., the adaptive filtering is performed with both images in the same reflectance/absorptance/density space, the resulting cancellation is typically good for regions where there is no printing in the front side, but results in residual show-through or over correction in other regions. For the same reason, the algorithm presented in this paper makes significant improvements over the method described in [15], where the idea of compensating for show-through by using scans of both sides of the page was first proposed.

To demonstrate the performance of the show-through correction algorithm on additional documents, a duplex printed page from the May 2000 issue of IEEE SPECTRUM [16] was used as the input document. The page was scanned in using the UMAX scanner and the settings described for the above experiment. The scans of the two sides of paper corresponding to this page are shown in Figs. 15 and 16. Corresponding show-through corrected versions of these scans are shown in Figs. 17 and 18, respectively, and the results obtained from thresholding are shown in Figs. 19 and 20, respectively. The images in these figures reinforce the conclusions obtained from the scans presented in the earlier figures. The show-through correction algorithm successfully eliminates the show-through in all regions of the image, whereas the simple thresholding method leaves residual show-through artifacts in the light gray regions and also causes degradation of these light gray regions due to the thresholding process. In order to better demonstrate the advan-

Fig. 20.   Thresholded version of the scan in Fig. 16.

tage of the show-through cancellation algorithm over the simple thresholding scheme, cropped regions of the show-through corrected images in Figs. 17 and 18 are shown in Figs. 21 and 23, respectively. Corresponding cropped versions of the thresholded images are shown in Figs. 22 and 24, respectively. In the thresholded images, one can clearly see the abrupt contouring in the light gray regions caused by the thresholding and also the residual show-through in several image regions. The images obtained from the show-through correction algorithms are free from both these artifacts.

Fig. 25 demonstrates the adaptive filter coefficients for two different image locations for the above case. Note that in comparison to Fig. 14 the coefficients in Fig. 15 demonstrate a smaller physical spread. This difference is to be expected because the paper sheets used in the two cases are different and therefore have different show-through PSF's corresponding to their differing light transmission and spreading characteristics. If perfect alignment of the two sides were to be achieved by independent means, and a fixed filter was utilized for the show-through correction, it would be necessary to account for
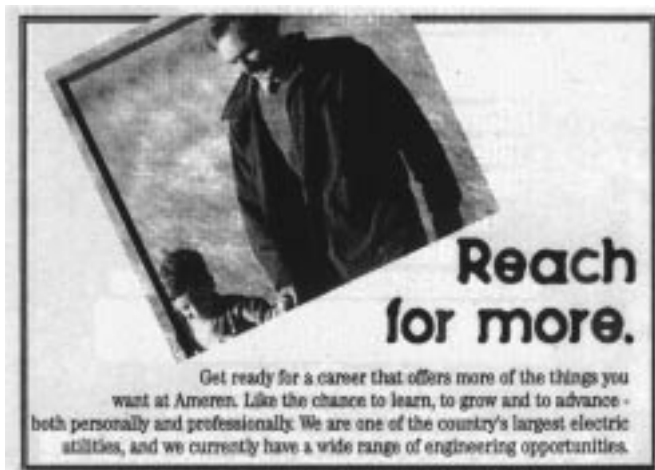
Fig. 21. Expanded region from show-through corrected version of the scan in Fig. 15.

Fig. 22. Expanded region from thresholded version of the scan in Fig. 15.

Fig. 23. Expanded region from show-through corrected version of the scan in Fig. 16.

Fig. 24. Expanded region from thresholded version of the scan in Fig. 16.

these differing characteristics of different papers by having separate fixed filters for each type of paper and incorporating a mechanism for identifying the different papers. The use of an adaptive filter allows these differences in the PSF to be automatically accounted for in the show-through correction.

VII. CONCLUSIONS

In scanning duplex printed pages, show-through, i.e., a residual low-contrast image of the back side in the scan of the front side, is a commonly encountered image degradation. This paper presented a physical analysis of the phenomenon of show-through, that resulted in a nonintuitive but elegant linear model for the phenomenon. Based on the analysis, an image processing method was developed for the correction of show-through using scans of the two sides. The method cancels show-through using adaptive linear filters in much the same way as echo cancellation is achieved in speech telephony.

The algorithm is found to be very effective in correction of show-through in tests performed on actual scans of duplex

Fig. 25. Adaptive filter coefficients $w^{m,n}(k,l)$ at two different image locations.

printed documents. In particular, the algorithm offers great improvement over the commonly used method of thresholding in that it preserves low-contrast information in the front-side while successfully eliminating (low-contrast) show-through from the back side.

APPENDIX
MODEL APPROXIMATION ERROR

Equations (17) and (18) were obtained through from the model of (7) using the approximation in (9) and assuming that $A_b(x, y) \approx A_b^s(x, y)$ and $A_f(x, y) \approx A_f^s(x, y)$. In this Appendix, the impact of these approximations is analyzed by obtaining corresponding exact expressions using the complete Taylor series expansions instead of the approximations. For notational simplicity, in the following analysis, the notion of a show-through point spread function will not be introduced. Denote

$$h = \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}}. \qquad (26)$$

Using the series expansion

$$\ln(1 - t) = -\sum_{l=1}^{\infty} \frac{t^k}{k} \qquad \text{for } |t| < 1 \qquad (27)$$

we have from (7)

$$D_f^s(x, y) = D_f(x, y) + \sum_{l=1}^{\infty} \frac{1}{l} (hA_b(x, y))^l \qquad (28)$$

$$D_b^s(x, y) = D_b(x, y) + \sum_{l=1}^{\infty} \frac{1}{l} (hA_f(x, y))^l \qquad (29)$$

$$\qquad (30)$$

and

$$\begin{aligned}
A_b(x, y) &= \left(1 - \frac{R_b(x,y)}{R_p^w}\right) \\
&= 1 - \exp(-D_b(x, y)) \\
&= 1 - \exp(-D_b^s(x, y) - \ln(1 - hA_f(x, y))) \\
&= 1 - \exp(-D_b^s(x, y)) \frac{1}{1 - hA_f(x, y)} \\
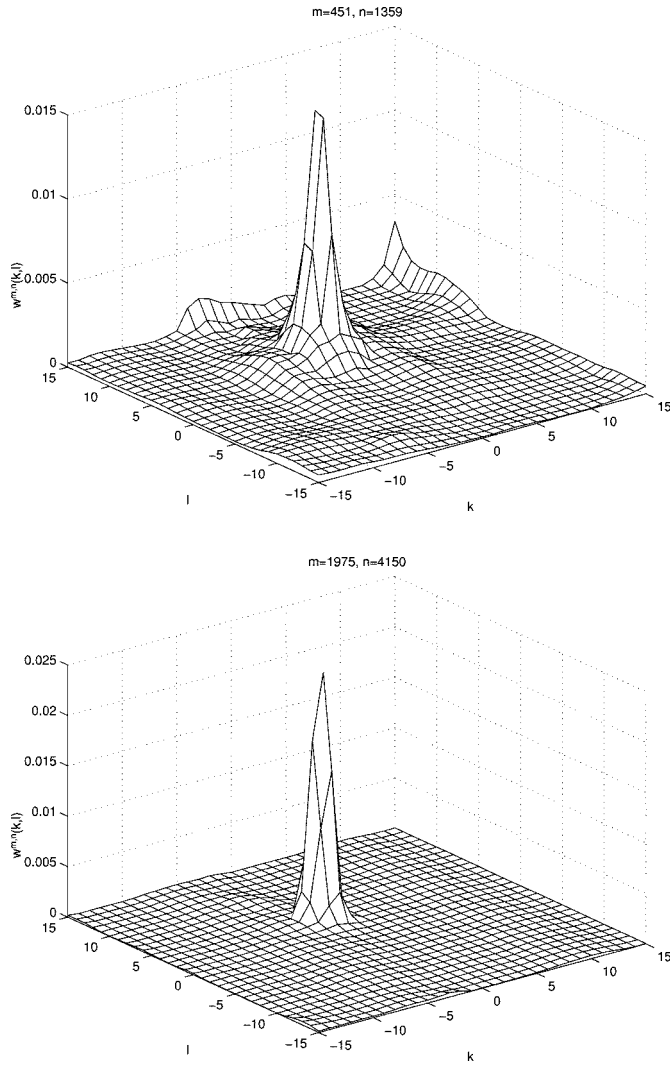&= 1 - \exp(-D_b^s(x, y)) - \exp(-D_b^s(x, y)) \\
&\quad \cdot \sum_{k=1}^{\infty} (hA_f(x, y))^k \\
&= A_b^s(x, y) - \frac{R_b^s(x, y)}{R_p^w} \sum_{k=1}^{\infty} (hA_f(x, y))^k \qquad (31)
\end{aligned}$$

where

$$A_b^s(x, y) \equiv 1 - \exp(-D_b^s(x, y)) = 1 - \frac{R_b^s(x, y)}{R_p^w} \qquad (32)$$

is the absorptance corresponding to the back side scan. Likewise

$$A_f(x, y) = A_f^s(x, y) - \exp\left(-D_f^s(x, y)\right) \sum_{k=1}^{\infty} (hA_f(x, y))^k \qquad (33)$$

where

$$A_f^s(x, y) \equiv 1 - \exp\left(-D_f^s(x, y)\right) = 1 - \frac{R_f^s(x, y)}{R_p^w} \qquad (34)$$

is the absorptance corresponding to the front-side scan.

Making the appropriate substitutions the exact versions of (17) and (18) can be obtained as

$$\begin{aligned}
D_f(x, y) &= D_f^s(x, y) - \sum_{l=1}^{\infty} \frac{1}{l} (hA_b(x, y))^l \\
&= D_f^s(x, y) - hA_b^s(x, y) \\
&\quad + h \frac{R_b^s(x, y)}{R_p^w} \sum_{k=1}^{\infty} (hA_f(x, y))^k - \sum_{l=2}^{\infty} \frac{1}{l} h^l \\
&\quad \cdot \left(A_b^s(x, y) - \frac{R_b^s(x, y)}{R_p^w} \sum_{k=1}^{\infty} (hA_f(x, y))^k\right)^l \\
\end{aligned}$$
$$\qquad (35)$$

$$
\begin{aligned}
D_b(x, y) &= D_b^s(x, y) - \sum_{l=1}^{\infty} \frac{1}{l} \left(h A_f(x, y)\right)^l \\
&= D_b^s(x, y) - h A_f^s(x, y) \\
&\quad + h \frac{R_f^s(x, y)}{R_p^w} \sum_{k=1}^{\infty} \left(h A_b(x, y)\right)^k - \sum_{l=2}^{\infty} \frac{1}{l} h^l \\
&\quad \cdot \left(A_f^s(x, y) - \frac{R_f^s(x, y)}{R_p^w} \sum_{k=1}^{\infty} \left(h A_b(x, y)\right)^k \right)^l .
\end{aligned}
\tag{36}
$$

From these expressions, it is clear that the approximations in (17) and (18) are obtained from the above equations by dropping the terms in h^2 and higher powers of h. As has been already noted, $h \ll 1$ for typical paper substrates and therefore the approximation error is quite small.

REFERENCES

[1] J. E. Farrell and B. A. Wandell, "Scanner linearity," *J. Electron. Imag.*, vol. 2, no. 3, pp. 225–230, July 1993.

[2] D. R. Lehmbeck and J. C. Urbach, "Scanned image quality," in *Optical Scanning*, F. Marshall Gerald, Ed. New York: Marcel Dekker, 1991, pp. 83–157.

[3] H. S. Baird, "Document image defect models," in *Structured Document Image Analysis*, H. S. Baird, H. Bunke, and K. Yamamoto, Eds. New York: Springer-Verlag, 1992, pp. 546–556.

[4] K. T. Knox, "Integrating cavity effect in scanners," in *Proc. IS&T/OSA Optics Imaging Information Age*, Rochester, NY, Oct. 20–24, 1996, pp. 83–86.

[5] R. C. Gonzalez and P. Wintz, *Digital Image Processing*, 2nd ed. Reading, MA: Addison-Wesley, 1987.

[6] R. P. Loce and E. R. Dougherty, *Enhancement and Restoration of Digital Documents: Statistical Design of Nonlinear Algorithms*. Bellingham, WA: SPIE Press, 1997.

[7] G. Sharma, "Show-through compensation apparatus and method," U.S. Patent Application 09/200984, Nov. 30, 1998.

[8] J. A. C. Yule and W. J. Neilsen[sic], "The penetration of light into paper and its effect on halftone reproduction," in *TAGA Proc.*, May 7–9, 1951, pp. 65–76.

[9] F. R. Ruckdeschel and O. G. Hauser, "Yule–Nielsen effect in printing: A physical analysis," *Appl. Opt.*, vol. 17, no. 21, pp. 3376–3383, Nov. 1978.

[10] M. Maltz, "Light scattering in xerographic images," *J. Appl. Photogr. Eng.*, vol. 9, no. 3, pp. 83–89, June 1983.

[11] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[12] G. O. Glentis, K. Berberidis, and S. Theodoridis, "Efficient least squares adaptive filter algorithms for FIR transversal filtering," *IEEE Signal Processing Mag.*, vol. 16, pp. 13–41, July 1999.

[13] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surveys*, vol. 24, no. 4, pp. 325–376, Dec. 1992.

[14] Q. Tian and M. N. Huhns, "Algorithms for subpixel registration," *CVGIP: Graph. Models Image Process.*, vol. 35, no. 2, pp. 220–233, Mar. 1986.

[15] K. T. Knox, "Show-through correction for two-sided documents," U.S. Patent 5 832 137, Nov. 3, 1998.

[16] *IEEE Spectrum*, vol. 37, May 2000.

Gaurav Sharma (S'88–M'97–SM'00) received the B.E. degree in electronics and communication engineering from University of Roorkee, India, in 1990, the M.E. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, in 1992, and the M.S. degree in applied mathematics and Ph.D. degree in electrical engineering from North Carolina State University (NCSU), Raleigh, in 1995 and 1996, respectively.

From August 1992 to August 1996, he was a Research Assistant with the Center for Advanced Computing and Communications, Electrical and Computer Engineering Department, NCSU. Since August 1996, he has been a Member of Research and Technical Staff with the Digital Imaging Technology Center, Xerox Corporation, Webster, NY. He is also involved in teaching in an adjunct capacity with the Electrical Engineering Department, Rochester Institute of Technology, Rochester, NY. His current research interests include color science and imaging, signal restoration, image security and halftoning, and error correction coding.

Dr. Sharma is a member of Sigma Xi, Phi Kappa Phi, Pi Mu Epsilon, and is the current Treasurer for the Rochester Chapter of the IEEE Signal Processing Society.