

CANCELLATION OF SHOW-THROUGH IN DUPLEX SCANNING

Gaurav Sharma

Dig. Imaging Tech. Ctr., Xerox Corp., MS0128-27E, 800 Phillips Rd, Webster, NY 14580.
Email: g.sharma@ieee.org

ABSTRACT

When scanning a page with printing on both sides, the printing on the back-side often shows through in the scan of the front-side because the page is not completely opaque. This phenomenon of show-through is analyzed and an image processing method for removal of this commonly encountered degradation is developed. A simplified mathematical model is obtained from first physical principles. 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.

1. INTRODUCTION

The phenomenon of *show-through* is a common artifact encountered while scanning duplex (double-sided) printed pages. Since the page is not completely opaque, the scanner receives light that is transmitted through the paper, reflected from the scanner backing and transmitted back through the paper. If there is no printing on the back side of the paper, this light produces no undesirable artifacts. If, however, the back side also has printing, then due to transmission through the paper, the scan of the front side contains a residual (transposed) image of the back side, which is referred to as *show-through* [1].

Figure 2 and 3, respectively, are the images obtained by scanning the two sides of a duplex printed page. 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 method, however, does not work for prints containing images or text that have gray levels other than 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.

2. SHOW-THROUGH MODEL AND ANALYSIS

Consider the simplified cross-sectional view of a scanner shown in Fig 1, where a duplex printed document is being scanned. The printing on the two sides 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. 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. Both these components encounter the transmittance of the front-side print layer but only the latter encounters the transmittance of the back-side print layer. 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

$$R_f^s(x, y) = T_f^2(x, y)(S_p + T_p^2 R_{bk} T_b^2(x, y)) \quad (1)$$

where x and y represent the two-dimensional spatial coordinates on the paper being scanned, S_p is the fraction of light scattered by the paper (in the forward direction), T_p is the transmittance of paper, R_{bk} is the reflectance of the backing, the subscript f denotes the front side, and the superscript s is used to denote that this is the “scanned” reflectance. Note that, the reflectance of unprinted paper is given by $R_p^w = S_p + T_p^2 R_{bk}$.

Equation (1) 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.

If the scan of the back-side is also available, then the scanned reflectance for the back side can be written in a manner analogous to (1) as

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

where the terms on the right hand side are as defined earlier. This yields a system of 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.

3. LINEARIZED SHOW-THROUGH MODEL

The nonlinear nature of equations (1) and (2), and the lack of knowledge of the parameters S_p and T_p make a direct analytic or numerical solution infeasible. In order to simplify these equations, it is advantageous to express them in terms of *optical density* by dividing by the paper reflectance and taking the (negative) natural logarithm

$$\begin{aligned} D_f^s(x, y) &\equiv -\ln \frac{R_f^s(x, y)}{R_p^w} \quad (3) \\ &= D_f(x, y) - \\ &\quad \ln \left(1 - \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} (1 - T_b^2(x, y)) \right) \\ &\approx D_f(x, y) + \frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}} A_b(x, y) \quad (4) \end{aligned}$$

where $D_f^s(x, y)$ is the paper-normalized density for the front-side scan, R_p^w denotes the reflectance of white (unprinted on either side) paper,

$$\begin{aligned} D_f(x, y) &\equiv -\ln(T_f^2(x, y)) \\ &= -\ln \frac{R_f(x, y)(S_p + T_p^2 R_{bk})}{R_p^w} \quad (5) \end{aligned}$$

represents the front-side density that would be obtained if there was no printing on the back-side, i.e., the show-through corrected density; $A_b(x, y) \equiv (1 - T_b^2(x, y)) = 1 - \frac{R_b(x, y)}{R_p^w}$ is the absorptance of the back side “print

layer”; and in (4) we have used the fact that for typical paper substrates, the fraction of light transmitted is much smaller than the fraction of light scattered, i.e., $T_p^2 \ll S_p$, which allows the approximation $\ln(1 - t) \approx -t$ for $|t| \ll 1$.

Equation (4) 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. A similar result can be obtained 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.*

4. SHOW-THROUGH CORRECTION

In practice, light undergoes spreading in the paper so the interaction between the front and the back sides is not purely point-wise as implied by (4). This spreading can be incorporated into the show-through model by replacing the scalar term $\frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}}$ in (4) with an empirical “show-through” point spread function (PSF) to obtain

$$\begin{aligned} D_f^s(x, y) &= D_f(x, y) + h(x, y) \otimes A_b(x, y) \\ &\approx D_f(x, y) + h(x, y) \otimes A_b^s(x, y) \quad (6) \end{aligned}$$

where $h(x, y)$ is the “show-through point spread function (PSF)”, \otimes represents the convolution operator, and the approximation in (6) replaces the absorptance for the back side “image layer”, i.e., $A_b(x, y)$, by the corresponding absorptances from the scan, given by

$$A_b^s(x, y) \equiv 1 - \frac{R_b^s(x, y)}{R_p^w} \quad (7)$$

Note that the show-through PSF represents the term $\frac{T_p^2 R_{bk}}{S_p + T_p^2 R_{bk}}$ in the physical model. The show-through PSF is therefore expected to be nonnegative and small in comparison to unity. Furthermore, it is determined by the scattering and transmission characteristics of the paper and the reflectance of the backing.

If the show-through PSF and the relative alignment of the front and back side images is known precisely, the show-through corrected image can be obtained from the scans of the two sides using (6). In practice, however, both the above assumptions are unrealistic. Using methods from adaptive signal processing, however, the show-through correction can be performed without these unrealistic constraints. Since equation (6) is

linear, the problem of show-through correction can be viewed as the two-dimensional equivalent of the one-dimensional echo-cancellation problem in speech telephony [2, pp. 53-56]. As in speech echo-cancellation, adaptive linear filters [2] can be used for automatically estimating and tracking the show-through PSF and implicitly the alignment between the front and the back side images.

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.

Using the estimate of R_p^w , scan data from the front side is converted to density relative to paper white as per (3) and data for the back side is converted to absorbance using (7). The image pixels are processed one at a time. At each pixel location (m, n) , the show-through corrected density is computed as

$$D_f(m, n) = D_f^s(m, n) - \sum_{k=-M}^M \sum_{l=-N}^N w^{m,n}(k, l) \times A_b^s(k - m, l - n) \quad (8)$$

where $w^{m,n}(k, l)$ is a two-dimensional adaptive FIR filter with support $[-M \dots M] \times [-N \times N]$ 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 several known algorithms in adaptive filter theory [2]. For the following description, the simplest and perhaps best-known least-mean square (LMS) algorithm [2, pp. 302] will be used. For the LMS algorithm, the filter coefficients for the next pixel location (m', n') are computed as

$$w^{m',n'}(k, l) = w^{m,n}(k, l) + \mu D_f(m, n) \times A_b^s(k - m, l - n) \quad (9)$$

where μ is the LMS adaptation step-size parameter. 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 (8) represents the error term used in the more general specification of the LMS algorithm [2, pp. 302].

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 (5) to obtain the output show-through corrected image. Note that the show-through correction of (8) 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.

5. EXPERIMENTAL RESULTS

The show-through model and the cancellation algorithm developed here were tested using a UMAX powerlook desktop scanner. The scanner was operated in a mode for which the scanner output is linearly related to scanned reflectance. The two sides of a duplex-printed page were scanned at 600 dpi resolution as 8-bit gray-scale images. The resulting scanned images are shown in Fig 2 and 3. 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-degree stroke lines at the bottom of the upper case “X” on the first line (the first “X” in “The Document Company Xerox”). 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.

Using the alignment information and the estimated white reflectance, the images were processed using the LMS adaptive-filtering algorithm presented earlier. Successive pixels to be processed were chosen along a serpentine raster so as to preserve spatial contiguity and thereby allow the adaptive filter to track the inevitable changes in alignment between the front and back side images over the page. 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.

The show-through corrected images for the text-side and the image-side obtained from the show-through cancellation algorithm developed are shown in Fig. 4 and 5, respectively. The results clearly indicate that the algorithm is successful in cancelling show-through. In comparison to the original scans, both the corrected images have almost no show-through and very minor processing artifacts (these minor artifacts can be seen better in the electronic versions of these images when displayed on a monitor).

6. REFERENCES

- [1] K. T. Knox, "Show-through correction for two-sided documents," United States Patent No. 5832137, 3 Nov. 1998.
- [2] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, NJ, second edition, 1991.

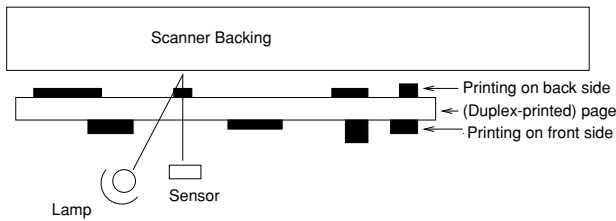


Figure 1: Scan of Duplex Printed Page.

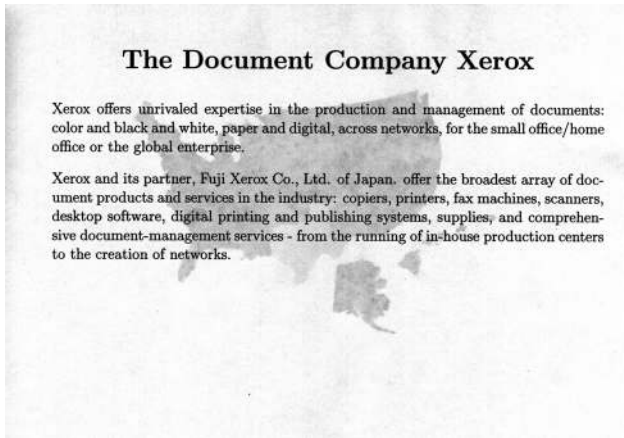


Figure 2: Scan of Side 1 of a two sided page, demonstrating show-through from Side 2.



Figure 3: Scan of Side 2 of a two sided page, demonstrating show-through from Side 1.

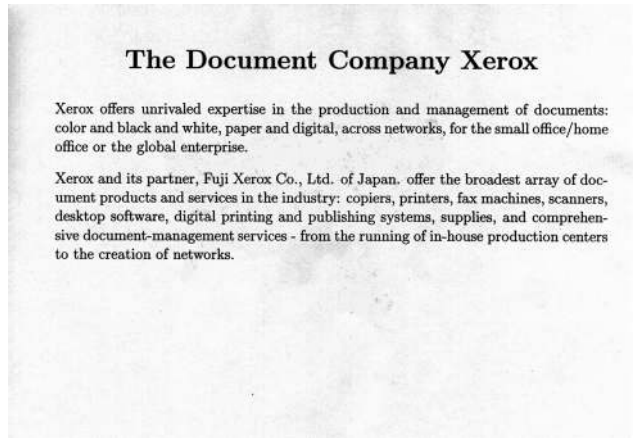


Figure 4: Side 1 scanned data after show-through correction.



Figure 5: Side 2 scanned data after show-through correction.