CROSS-TALK CANCELLATION IN SCANNED IMAGES

BOAZ OPHIR

CROSS-TALK CANCELLATION IN SCANNED IMAGES

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

IN ELECTRICAL ENGINEERING

BOAZ OPHIR

SUBMITTED TO THE SENATE OF THE TECHNION — ISRAEL INSTITUTE OF TECHNOLOGY

HESHVAN, 5767 HAIFA OCTOBER, 2006

THIS RESEARCH THESIS WAS SUPERVISED BY PROF. DAVID MALAH
UNDER THE AUSPICES OF THE ELECTRICAL ENGINEERING
DEPARTMENT

ACKNOWLEDGMENT

I would like to thank Prof. David Malah for his fruitful ideas and for the precious time he spent with me.

I would like to thank Dr. Ehud Karnin and Dr. Eugene Walach of the IBM Research Laboratory in Haifa, for presenting us with the problem and providing the data used in the experiments.

Finally, I owe a great debt to my family, for their patience, understanding, and support. This work is dedicated to them.

THE GENEROUS FINANCIAL HELP OF THE TECHNION IS GRATEFULLY
ACKNOWLEDGED

Contents

List of Figures

Abstract

Show-Through interference is a common occurrence when scanning duplex printed documents. The back-side printing shows through the paper thus contaminating the front side image. The same occurs when scanning the reverse side of the page. This is not a problem in low quality scans (as done in home/office scanners) where, up to a degree, image quality is not an issue. The matter becomes crucial when image quality is essential. Such a case is when creating a master copy in the digital printing industry.

Previous work focused on analyzing the process that causes the phenomenon, tracking the passage of light in the scanner mechanism as it passes through the document. Obviously, any such model includes a point spread function modelling the scattering effect caused to the light as it passes through the paper blurring the back-side Show-Through. It was shown that the process can be modelled as non-linear convolutive mixture of the desired images.

In this work we present two algorithms for removing the Show-Through.

The first algorithm, improving on earlier work, attempts to alleviate the problem via an adaptive decorrelation process. A cascaded multi-stage filtering scheme minimizes the image correlation. The algorithm adapts to local brightness variations, estimating local background brightness through a mean-shift process. The effects of the cross interference, in the reference signals, are minimized by an additional post

processing adaptive filtering stage.

The second algorithm treats the problem as a Blind Source Separation (BSS) problem, simultaneously estimating the images and mixing parameters. Image separation is achieved via an alternating minimization process, minimizing a cost functional combined of a Mean Squared Error fidelity term and Total-Variation (TV) regularization terms. The fidelity/regularization tradeoff is set by a location dependent scheme aiming to preserve image edges while removing unwanted show-through edges. Optimization is done via the Iterated Conditional Modes (ICM) method thus avoiding large scale optimization.

The decorrelation based algorithm achieves good results on a wide range of images at relatively low computational cost. The BSS algorithm also achieves good image separation, but at a much higher computational cost. Nevertheless, we believe that the BSS approach holds much promise for future development.

List of Symbols

BSS	Blind Source Separation
FIR	Finite Impulse Response
GNC	Graduated Non-Convexity
ICA	Independent Component Analisys
MAP	Maximum *a priori*
MI	Mutual Information
ML	Maximum Likelihood
MRF	Markov Random Field
MSE	Mean Squared Error
PDF	Probability Density Function
PSF	Point Spread Function
STD	Standard Deviation
TV	Total Variation

R	Reflectance
T	Transmittance
D	Optical Density
A	Optical Absorptance
subscript $f, 1$	Front side
subscript $b, 2$	Back side
subscript p	Paper
subscript bk	Scanner backing
subscript pw	White (unprinted) paper
$\mathbf{Y}$, y_i	Estimated signals
$\mathbf{S}$, s_i	Source signals
$\mathbf{X}$, x_i	Mixed signals
$\mathcal{H}$, $\mathbf{H}$, $h_{i,j}$	Non-linear/linear mixing operator, blurring filter of source j for mixture i.
f	Non-linear function
$\mathbf{w}$, $\mathbf{a}$, $\mathbf{b}$	Tap filters
μ	Filter update step
F, J	Cost functional
λ	Fidelity/Regularization tradeoff
I	Mutual Information
H	Entropy
I_n	Normalized mutual information
ϕ	MRF functional

Chapter 1

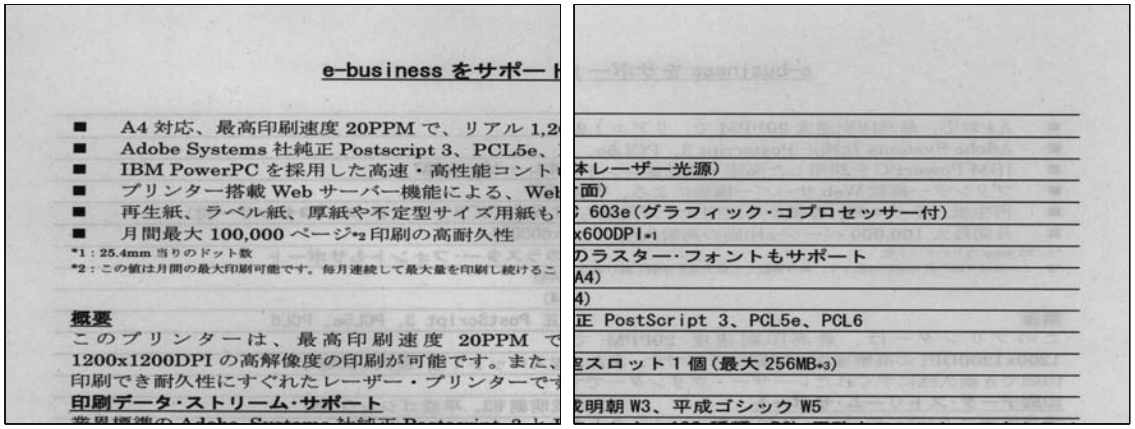
Introduction

1.1 The Show-Through Problem

Show-through interference is a common occurrence when scanning duplex printed documents. The back-side printing *shows* through the paper, thus contaminating the front side scan. This is not considered a problem in low quality scans (as done in home/office scanners) where, up to a degree, image quality is not an issue. The matter becomes crucial when image quality is essential. Such a case is when creating a master copy, in the digital printing industry. The problem is akin to the Cross Talk problem in multi-channel communication systems.

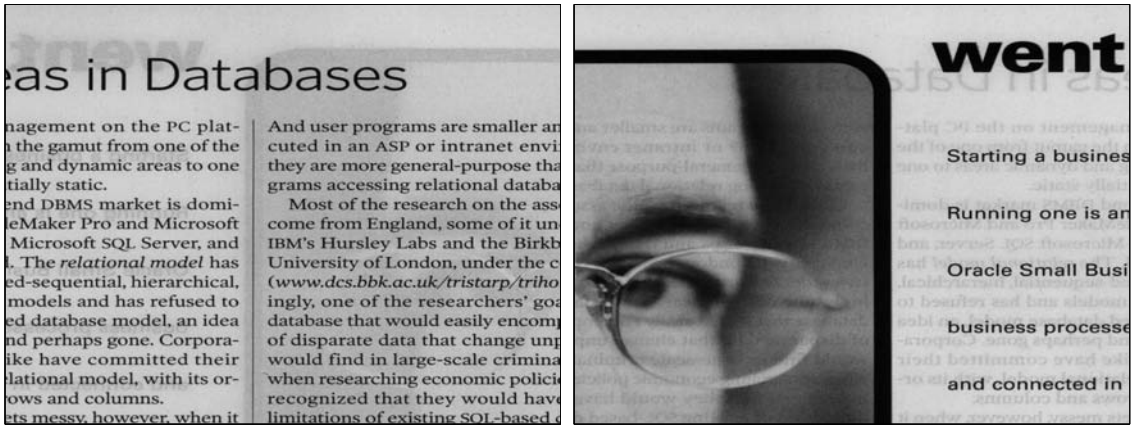
A naive approach to solving this problem is by thresholding, but this only works with scans of "simple" documents, such as scans of documents that contain only text (basically, binary images) as shown in Fig.1.1. When the document is more complex, such as a true grayscale image as shown in Fig.1.2, this approach fails.

Detailed analysis of the phenomenon (Chapter 2) reveals that show-through is the result of a complex process including non-linear effects and spatial dependency (convolution).



(a) Front (b) Back

Figure 1.1: Example scans of binary documents.



(a) Front (b) Back

Figure 1.2: Example scans of true grayscale documents.

1.2 Previous Work

While the Show-Through effects may in some cases be partially reduced through standard image de-noising techniques, recovery of a show-through corrected image cannot be accurately done using only a single side scan, because it is not possible to reliably distinguish between low-contrast printing on the front-side and the low-contrast show-through from the back-side.

Thus, Show-Through removal requires using scans of both sides of the paper. The first to suggest such an approach was Knox [22, 23]. However, the solution

proposed is overly simplistic. The mixing model described is linear point-wise, and more importantly, no practical method for estimating the mixing parameters is given.

The only comprehensive work regarding Show-Through was done by Sharma [37]. The main contribution of Sharma's work is in the physical modelling of the problem and the partial linearization of it. We describe this model in detail in Chapter 2.

Since very few works deal specifically with the problem of Show-Through, our review covers works on the more general subject of signal/image separation or cleaning.

Due to the complexity of the mixing model in the Show-Through scenario, most of these techniques are not directly applicable to our problem however we borrow ideas from them in our work.

1.2.1 Image Separation

The category of works, most closely related to the Show-Through problem, is that of separating mixtures of images. Many of these works are in the context of separating semi-reflections, a common occurrence in photography.

In [35] a solution is sought for the reflection separation problem that minimizes the structural correlations between the images. Minimization is done at varying scales using normalized gray-scale correlation as the measure of similarity. The mixing process is assumed to be pointwise linear. While promising at first, we found this method to be unapplicable and difficult to scale to convolutive mixtures.

Independent Component Analysis (ICA) techniques, using high-order statistics, have also been used for separation of reflections [16, 9]. These works use different polarizations to get linear pointwise image mixtures. The algorithm in [16] uses

graphical techniques, and [9] uses a sparsity prior, to achieve image separation. Neither of these methods can be readily scaled to non-linear convolutive mixtures.

In [36] focus cues are used. In this scenario, a convolutive mixture model is used. Each mixture contains one in focus image layer, and one out of focus layer. The separation criterion is Mutual Information, however minimization is done via exhaustive search, thus making it impractical.

In [24, 25] reflection separation is performed on a single image. The image mixture is assumed to be pointwise linear and fixed throughout the image. The algorithms assume statistic priors on the image and use *a-priori* information about the scene. In [24] human interaction is critical since the algorithm heavily depends on an "oracle". The algorithm in [25] suffers from high non-linearity of the cost function and convergence to the correct solution is not assured.

Other approaches include algorithms using motion cues [15] or stereo photography [44] to separate the images. These methods require images to be taken from different points of view (or different time), which is not relevant in our scenario.

In [40, 41, 17] a Maximum Likelihood (ML) approach is taken to separate linear pointwise image mixtures. Markov Random Field (MRF) models are used as image priors, requiring computational heavy non-convex optimization methods to be used. In [42] an attempt is made to generalize this approach to convolutive mixtures, however, the large number of variables makes this approach impractical for real world applications.

1.2.2 Related Problems

A more general category of works deal with multi-channel systems. Specifically our interest is in convolutive systems.

The prevalent approach in signal processing today falls under the heading of Independent Component Analysis (ICA) [1]. ICA signal separation techniques attempt to achieve statistical independence of the output signals. These methods typically use the Mutual Information (MI) measure or high order statistics as the separation criteria.

In order to calculate the MI, the signal entropies need to be calculated. A popular approach is to estimate entropies using simple parametric models of the PDF [6, 43]. The use of simple parametric models reduces the complexity of the problem but may lead to inaccurate solutions.

Many Blind Source Separation (BSS) algorithms approximate the MI criteria with high order statistics. Second order statistics are used in [46, 29] and forth order statistics in [48, 49, 39, 14]. We tested such an approach, cancelling 4th-order output cross cumulants, and found the results unimpressive.

A different approach, related to MI, uses maximum likelihood / maximum *a-priori* (ML/MAP) estimation [30, 32, 31, 27]. In this approach the separated sources are optimized to best fit the statistical model, and the quality of the results depends on the accuracy of the model.

Many of these works, dealing with one dimensional signals, make assumptions regarding the signals that are not relevant to images. Specifically, a popular assumption is that the signals are i.i.d.. Algorithms relying on this assumption many times

whiten the output signals thus corrupting the results severely.

A closely related problem to the convolutive mixture problem, is the blind deconvolution problem. We mention here just a handful of select works that influenced our work. In [10] a sparsity prior is used for deconvolution, while [11, 21] use Total-Variation and [4, 18] use MRFs.

1.3 Proposed Approach

We treat the show-through problem as an Image Separation problem, where neither desired images nor mixing parameters are known. Our inputs are two grayscale images created by scanning both the front and back sides of a duplex printed page. Each scan is obviously contaminated by the reverse side.

We assume no prior knowledge about either the contents of the scanned document or of the physical attributes of the paper and ink. We assume the scans are made using a standard flatbed scanner and have no specific knowledge of the scanner mechanism or characteristics. The only assumption is that the images are aligned (registered) in advance. Image registration is itself not a trivial problem. In this work the images were registered manually.

Our goal is to generate two images, cleaned of show-through effects, as near as possible to the images that would be generated by scans of the same pages (front and back) if they were printed on a single side.

For this propose we developed two Show-Through removal algorithms. The first algorithm is based on the decorrelation criterion, improving on Sharma's algorithm. The second algorithm we developed uses a ML/MAP BSS approach. The algorithm

uses Total-Variation image priors and gradient based optimization techniques.

1.4 Thesis Outline

In Chapter 2 we formalize the show-through problem as an image mixture problem and present a physical model for the phenomenon.

In Chapter 3 we give a an outline of Decorrelation models for signal separation. In Chapter 4 we analyze Sharma's decorrelation-based algorithm, propose an improved decorrelation-based Show-Through removal algorithm, and present simulation results.

In Chapter 5 we formulate the Show-Through problem as a BSS minimization problem, minimizing energy functionals based on Maximum-a-Priori models. In Chapter 6 we present a minimization algorithm for Show-Through removal and present simulation results.

In Chapter 7 we summarize the work and provide suggestions for future work.

Chapter 2

Modelling Show-Through as an Image Mixture Problem

In this chapter we will show how the show-through phenomenon in scans of duplex printed images can be expressed as a non-linear convolutive mixture of images.

In order to facilitate understanding we first formalize the different types of image mixtures, give a brief overview of how print and scanners work, and then present a physical model of the show-through phenomena.

2.1 Image Mixtures

2.1.1 Pointwise Mixtures

The simplest image mixtures are pointwise mixtures, analogous to instantaneous mixtures in one dimensional signals. the mixtures $\mathbf{x}_f$ and $\mathbf{x}_b$ (front and back side images respectively), are mixed so that each pixel value at location (i, j) is a combination of the values at the same location of the original images $\mathbf{s}_f$ and $\mathbf{s}_b$.

A linear pointwise mixture can be expressed as:

$$\begin{aligned} \mathbf{x}_f(i,j) &= h_{11}\mathbf{s}_f(i,j) + h_{12}\mathbf{s}_b(i,j) \\ \mathbf{x}_b(i,j) &= h_{21}\mathbf{s}_f(i,j) + h_{22}\mathbf{s}_b(i,j) \end{aligned} , \tag{2.1}$$

where the h's are scalar mixing coefficients.

Defining the mixing matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} , \tag{2.2}$$

allows us to write (2.1) in compact form:

$$\mathbf{X} = \mathbf{HS} , \tag{2.3}$$

where $\mathbf{X}$ is a row stack matrix of the mixtures, and $\mathbf{S}$ of the sources. For $M \times N$ pixel images, these matrices have dimensions of $2 \times (MN)$.

2.1.2 Convolutive Mixtures

A more realistic scenario in the case of images is that of convolutive mixtures. In this case the sources are also blurred in the mixing process. Thus the mixing matrix's scalar coefficients need to be replaced by blur kernels, and the matrix multiplication by convolution. The mixture becomes:

$$\begin{aligned} \mathbf{x}_f(i,j) &= (\mathbf{h}_{11} * \mathbf{s}_f)(i,j) + (\mathbf{h}_{12} * \mathbf{s}_b)(i,j) \\ \mathbf{x}_b(i,j) &= (\mathbf{h}_{21} * \mathbf{s}_f)(i,j) + (\mathbf{h}_{22} * \mathbf{s}_b)(i,j) \end{aligned} . \tag{2.4}$$

In a typical scenario, each mixture contains two source images, only one of which is blurred, i.e.:

$$\begin{aligned} h_{11}(i,j) &= \mathbf{h}_{11}\delta(i,j) \\ h_{22}(i,j) &= \mathbf{h}_{22}\delta(i,j) \end{aligned} , \tag{2.5}$$

where δ is the Kronecker delta function.

The mixing operation can no longer be expressed simply as a matrix multiplication, however, for simplicity of notation, we define a mixing operator $\mathcal{H}$ such that:

$$\mathbf{X} = \mathcal{H}(\mathbf{S}) \ , \tag{2.6}$$

In the linear pointwise case, the mixing operator $\mathcal{H}$ is simply the matrix $\mathbf{H}$ multiplied by $\mathbf{S}$.

2.1.3 Non-Linear Mixtures

The most general scenario is one where the source signals are also non-linearly transformed during the mixing process. Obviously, this severely complicates the un-mixing process, especially when the non-linear functions are unknown. We do not presume to deal with the general non-linear mixture problem. However, we do consider the case where the non-linear function is *known*, as the show-through process, as will be shown in the following sections, can be modelled as:

$$\begin{aligned} \mathbf{x}_f(i,j) &= h_{11}\mathbf{s}_f(i,j) + (\mathbf{h}_{12} * f(\mathbf{s}_b))(i,j) \\ \mathbf{x}_b(i,j) &= (\mathbf{h}_{21} * f(\mathbf{s}_f))(i,j) + h_{22}\mathbf{s}_b(i,j) \end{aligned} \ , \tag{2.7}$$

where $f(\cdot)$ is a known non-linear function.

For convenience and simplicity we will use throughout this work the compact notation in (2.6) for all mixture models.

2.2 About Scanners and Print

The printing process involves the application of translucent [1] ink in a thin layer, on a reflecting surface, usually white paper. The transparency of the ink means that

[1] transmitting and diffusing light (Merriam-Webster)

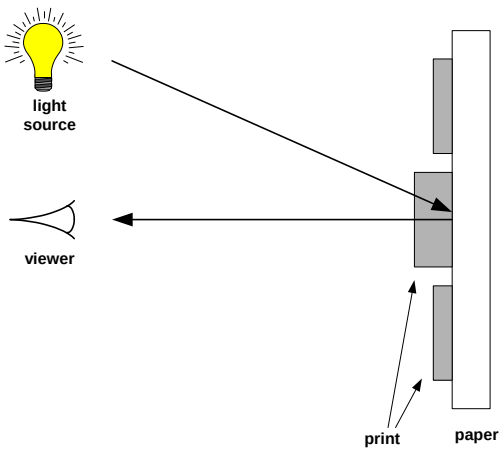


Figure 2.1: Viewing Printed Paper

when the printed paper is illuminated by white light, the light entering the viewer's eye (or the optical sensor) actually passes through the ink layer, to the paper, and is reflected by the paper back through the ink layer again, to the eye. The ink absorbs part of the light passing through it. The spatial profile of this absorbtion, or filtering, represents the printed information.

In a typical scanner, the document to be scanned is flattened on a transparent plate of glass by the scanner backing. The light originating from the scanner lamp, illuminates the document through the glass. The light is reflected from the document back to the sensor, thus creating a reflectance profile of the document.

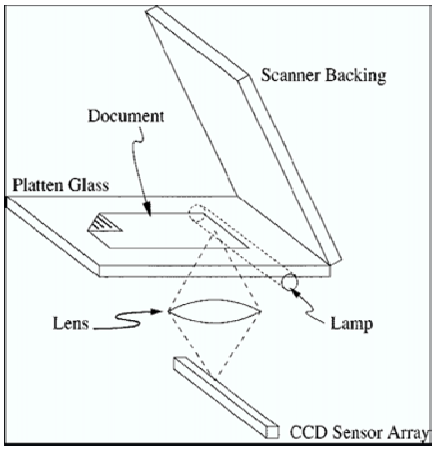


Figure 2.2: Schematic of Optical Components of Flatbed Scanner (from [37])

The show-through problem in duplex printed paper arises from the fact that the paper does not reflect 100% of the light that hits it. A fraction of the light actually passes through the paper, through the back-side print layer, and is reflected back toward the sensor by the scanner backing, as shown in Fig.2.3.

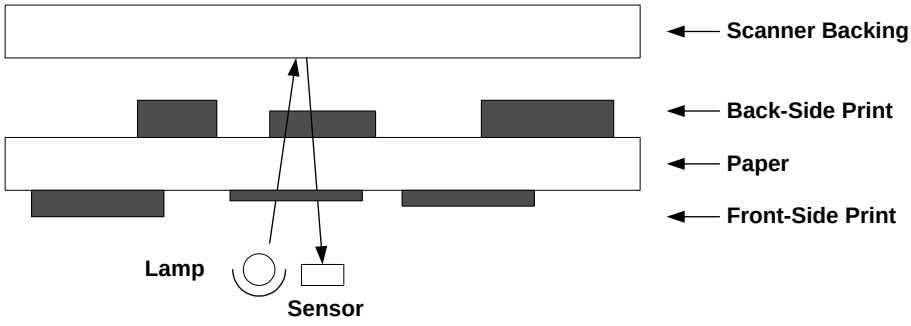


Figure 2.3: Passage of Light Through Duplex Printed Paper

2.3 Physical Model

The model developed by Sharma in [37] tracks the light passing through the printed page and the scanner mechanism. The results are given in terms of the reflectance and transmittance profiles of the print and paper.

The model described in the following sections is obviously a simplified one. Issues such as ink penetration into the paper substrate (when printing with normal inks on normal paper, there is no well-defined interface between the ink and paper) are ignored. Also, any Fresnel reflections at the print surface are neglected, as are secondary reflections between the paper and scanner backing. Nevertheless, the model captures enough of the complexity of the phenomenon to allow for useful correction algorithms.

2.3.1 Non-Linear Model

We define the reflectance of the paper R_p as the fraction of the light (optical energy/flux) reflected by the paper in the forward direction (toward the sensor). The transmittance of the paper T_p is defined as the fraction of light passed through it. A reasonable assumption is that most of the light that hits the paper is reflected by it $R_p \gg T_p$, and that these values are not location dependant, i.e. the paper is uniform.

The scanner backing also reflects light back toward the sensor. This value is denoted R_{bk}.

Thus, the reflectance of white paper (paper with no print on either side) is made up of two main components, light that is reflected by the paper and light that passes through the paper and is reflected by the scanner backing and passes back through the paper. We denote this value R_{pw}:

$$R_{pw} = R_p + T_p^2 R_{bk}. \tag{2.8}$$

The print layers on either side of the page exhibit transmittance properties $T_f(i,j)$ and $T_b(i,j)$, with the printed information represented by the spatial transmittance profile of these layers.

The reflectance detected when scanning the front side (neglecting second order effects) is given by:

$$\begin{aligned} R_f^s(i,j) &= T_f R_p T_f + T_f T_p T_b R_{bk} T_b T_p T_f \\ &= T_f(i,j)^2 (R_p + T_p^2 R_{bk} T_b^2(i,j)) \end{aligned}. \tag{2.9}$$

The expression $T_f R_p T_f$ represents light passing through the front-side print layer

T_f , reflected by the paper R_p and passing back again through the print to the sensor (see Fig.2.1). The expression $T_f T_p T_b R_{bk} T_b T_p T_f$ represents light passing through the front-side print T_f, the paper T_p and the back-side print T_b and reflected by the scanner backing R_{bk} back along the same path (Fig.2.3).

Similarly for the back side scan:

$$\begin{aligned} R_b^s(i,j) &= T_b R_p T_b + T_b T_p T_f R_{bk} T_f T_p T_b \\ &= T_b(i,j)^2 (R_p + T_p^2 R_{bk} T_f^2(i,j)) \end{aligned} . \tag{2.10}$$

The show-through component of the scanned image stems from the dependence of acquired reflectance on the reverse side transmittance. We define the desired result of a cleaning process as the images that would be recovered from the scan if no reverse side print existed:

$$\begin{aligned} R_f(i,j) &= T_f^2(i,j)(R_p + T_p^2 R_{bk}) \\ &= T_f^2(i,j) R_{pw} \end{aligned} , \tag{2.11}$$

and

$$\begin{aligned} R_b(i,j) &= T_b^2(i,j)(R_p + T_p^2 R_{bk}) \\ &= T_b^2(i,j) R_{pw} \end{aligned} . \tag{2.12}$$

Obviously, the scanned images are not a linear combination of the desired images R_f and R_b.

2.3.2 Linearized Model

In order to linearize the mixture model two concepts are introduced. Optical Density, defined as the logarithm of the ratio of incident to reflected light, or in our case, the negative logarithm of the reflectance:

$$D = -\ln R, \tag{2.13}$$

and Optical Absorptance, defined as the ratio of absorbed to incident light, or in our case the complement of the reflectance:

$$A = 1 - R, \tag{2.14}$$

Normalizing reflectance values by the white paper reflectance, the Normalized Density of the front side print is given by:

$$\begin{aligned} D_f^s(i,j) &\equiv -\ln\left(\frac{R_f^s(i,j)}{R_{pw}}\right) \\ &= -\ln\left(\frac{T_f^2(i,j)(R_p + T_p^2 R_{bk} T_b^2(i,j))}{R_p + T_p^2 R_{bk}}\right) \\ &= -\ln(T_f^2(i,j)) - \ln\left(1 - \frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}}(1 - T_b^2(i,j))\right) \end{aligned} \quad , \tag{2.15}$$

similarly, the desired density image (without show-through) is given by:

$$D_f(i,j) \equiv -\ln\left(\frac{R_f(i,j)}{R_{pw}}\right) = -\ln(T_f^2(i,j)). \tag{2.16}$$

Substituting eq.2.16 into eq.2.15 gives:

$$D_f^s(i,j) = D_f(i,j) - \ln\left(1 - \frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}}(1 - T_b^2(i,j))\right) \tag{2.17}$$

Since $R_p \gg T_p$ we can approximate (2.17) using the relation $\ln(1-t) \approx -t$ for $|t| \ll 1$.

$$D_f^s(i,j) \approx D_f(i,j) + \frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}}(1 - T_b^2(i,j)) \tag{2.18}$$

Recalling that the normalized back-side absorptance is given by:

$$A_b(i,j) = 1 - \frac{R_b(i,j)}{R_{pw}} = (1 - T_b^2(i,j)), \tag{2.19}$$

we can express the observed front side density by the following linear combination of the desired front side density and back side absorptance:

$$D_f^s(i,j) \approx D_f(i,j) + \frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}} A_b(i,j). \tag{2.20}$$

Notice that the scalar term $\frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}}$ depends only on paper and machine characteristics and not on the print layers.

The model, so far, ignores the spatial interaction between the front and back sides, i.e., the blurring caused by the passage of light through the paper. This spreading of light through paper has been studied at length particularly in the context of half toning. The Yule-Nielsen effect (1951) reflects the fact that a photon which enters the ink dot can exit from the non-inked paper or vice versa due to light diffusion in the substrate (see [33] and references within). This effect, also called optical dot gain, depends on the optical properties of the materials (paper, ink) and geometrical distribution of ink dots (resolution, size and shape). This behavior can be characterized by a point spread function.

This spatial interaction can be modelled by replacing the scalar term $\frac{T_p^2 R_{bk}}{R_p + T_p^2 R_{bk}}$ in (2.20) by the point spread function h, resulting in the following relation:

$$D_f^s(i,j) = D_f(i,j) + (h * A_b)(i,j). \tag{2.21}$$

The unknown parameters are incorporated in the PSF as an attenuation factor.

Similarly for the back side scan:

$$D_b^s(i,j) = D_b(i,j) + (h * A_f)(i,j). \tag{2.22}$$

Obviously, the relations depicted in (2.21) and (2.22) between the scanned and

desired images are not strictly linear. However, they do express the image mixtures as a linear mixture of aspects of the desired front and back side images.

In summary, we can now express the show-through problem as a non-linear convolutive mixture of images. In the *optical density* domain the mixing operator becomes:

$$\mathcal{H} = \begin{bmatrix} 1 & \mathbf{h}_{12} * f(\cdot) \\ \mathbf{h}_{21} * f(\cdot) & 1 \end{bmatrix} \tag{2.23}$$

where $f(d) = 1 - e^{-d}$ is the non-linear function that translates density values to absorptance.

Working in the reflectance domain (standard gray-scale images) requires transforming to and from density and absorptance values, when estimating the mixture parameters. This requires an estimation of R_{pw}. An automatic method for this is proposed in section 4.2.1.

2.4 Similarity Measures

Measuring the amount of show-through goes to the heart of the show-through removal problem. Finding a reliable measure of the amount of show-through consistent with our human perception, is in itself a nontrivial problem. In theory, once such a measure is found, algorithms can be devised minimizing the measure and removing the show-through.

One way of assessing the amount of show-through can be done by measuring the similarity of the front and back side images. Theoretically, the greater the similarity, the more show-through is present in the images. Several such similarity measures can be defined between the images.

In practice, while the show-through removal process usually caused a reduction in the similarity measures we rarely saw a direct correlation between image quality and the degree of improvement in the measures.

2.4.1 Cross Correlation

The correlation of two images s_1 and s_2 is defined by:

$$Cor(s_1, s_2) = \sum_{m,n} (s_1(m,n) - E\{s_1\}) * (s_2(m,n) - E\{s_2\}) \tag{2.24}$$

Normalizing by the auto-correlation of the images gives the following cross-correlation measure:

$$XC(s_1, s_2) = \frac{Cor(s_1, s_2)}{\sqrt{Cor(s_1, s_1) * Cor(s_2, s_2)}} \tag{2.25}$$

2.4.2 Mutual Information

A measure of Mutual Information is defined in [36]. This measure is a normalized version of the Kullback-Leibler distance which measures how far the images are from statistical independence.

Let the probabilities for certain brightness values $\widetilde{s}_1$ and $\widetilde{s}_2$ be $P(\widetilde{s}_1)$ and $P(\widetilde{s}_2)$, respectively. These probabilities are estimated by the histograms of the images. The joint probability is $P(\widetilde{s}_1; \widetilde{s}_2)$ (which is in practice estimated by the joint histogram of the images). The mutual information is then:

$$I(s_1, s_2) = \sum_{\widetilde{s}_1, \widetilde{s}_2} P(\widetilde{s}_1, \widetilde{s}_2) \log \frac{P(\widetilde{s}_1, \widetilde{s}_2)}{P(\widetilde{s}_1) P(\widetilde{s}_2)} \tag{2.26}$$

The entropy, or self-information, of each image is defined as:

$$H(s) = -\sum_{\widetilde{s}} P(\widetilde{s}) \log(P(\widetilde{s})) \tag{2.27}$$

Normalizing the mutual information, by the average of the image entropies, gives the ratio of mutual information to the self information:

$$I_n(s_1, s_2) = \frac{I(s_1, s_2)}{[H(s_1) + H(s_2)]/2} \tag{2.28}$$

In this chapter we showed how the show-through phenomenon in scans of duplex printed images can be modelled as a non-linear convolutive mixture of images. Based on this model, we now offer two different approaches to removing the show-through, described in subsequent chapters.

Chapter 3

Decorrelation Models

3.1 Single Channel Interference

The most widely used approach to two-channel signal separation was proposed by Widrow et al. [47]. An interfering signal s_2 is coupled into the primary signal s_1 through an unknown system H_{12}. The signal mixing system is depicted in Fig.3.1.

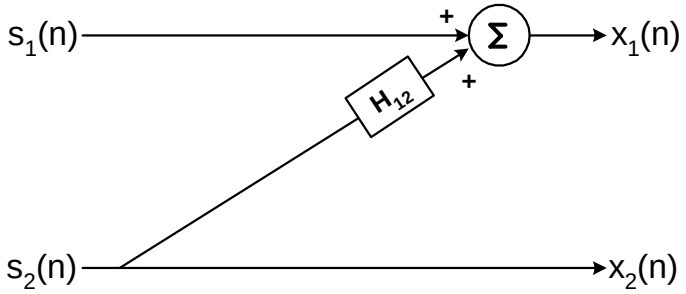


Figure 3.1: Single Channel Interference

The underlying assumption is that the signals s_1 and s_2 are uncorrelated. The objective is to reconstruct an estimate y_1 of s_1 using the interfering signal x_2 as a reference.

It is suggested in [47] that the unknown system H_{12} can be identified by minimizing the average power of the reconstructed signal. This system can be used for cancellation of the interfering signal as illustrated in Fig.3.2.

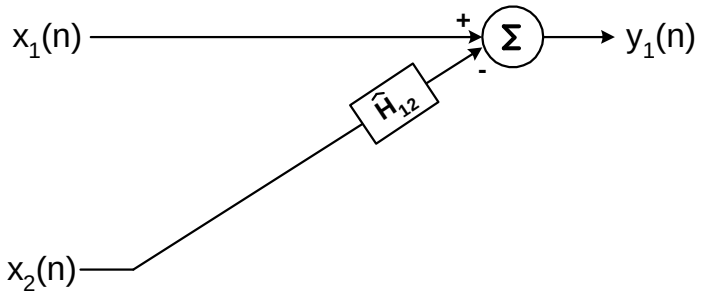


Figure 3.2: Least Squares Reconstruction

Adjusting the filter to minimize the average power of the reconstructed signal corresponds to estimating H_{12} by a least-squares fit of the interfering (reference) signal x_2, to the primary signal x_1. Minimizing the output power, minimizes the output noise power and maximizes the output signal-to-noise ratio.

Recursive (RLS) and sequential (LMS) algorithms are widely used in these contexts [20].

3.2 Adaptive Filtering Methods

3.2.1 LMS Method

The Least-Mean-Square (LMS) algorithm is a stochastic gradient based algorithm. The estimated recovery filter $\widehat{H}_{12}$ is assumed to be a FIR filter with tap weight vector $\mathbf{w}$ of length M.

The algorithm consists of two basic processes:

- An adaptive process for the adjustment of the filter tap weights.
- A filtering process, generating an estimation of the desired response and an estimation error used in the adaptation process.

The algorithm attempts to minimize the mean-squared error, which at time n is given by:

$$J(n) = E[y_1^2(n)] \tag{3.1}$$

A steepest descent tap-weight updating scheme can be described as a recursive relationship:

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \frac{1}{2}\mu[-\nabla(J(n))] \tag{3.2}$$

where μ is a positive real-valued constant controlling the size of the incremental correction to the tap-weight vector.

The gradient-decent algorithm requires the knowledge of the gradient vector at time n:

$$\nabla(J(n)) = -2\mathbf{p} + 2\mathbf{R}\mathbf{w}(n) \tag{3.3}$$

where, $\mathbf{R} = E[\mathbf{x}_2(n)\mathbf{x}_2^H(n)]$ is the M-by-M correlation matrix of the M-by-1 tap-input vector $\mathbf{x}_2(n)$ (composed of the tap input series $x_2(n)$, $x_2(n-1)$, ..., $x_2(n-M+1)$). Likewise, $\mathbf{p} = E[\mathbf{x}_2(n)x_1^*(n)]$, is the M-by-1 cross correlation vector between the tap inputs and the desired response $x_1(n)$. Superscript H indicates Hermitian transposition.

However, exact measurements of the gradient vector are impossible since this would require prior knowledge of both the correlation matrix $\mathbf{R}$ and cross correlation vector $\mathbf{p}$. Thus, the gradient vector must be estimated from the available data. An instantaneous estimate of the gradient vector is given by substituting instantaneous estimates of $\mathbf{R}$ and $\mathbf{p}$ into equation 3.3:

$$\widehat{\mathbf{R}}(n) = \mathbf{x}_2(n)\mathbf{x}_2^H(n) \tag{3.4}$$

$$\widehat{\mathbf{p}}(n) = \mathbf{x}_2(n)x_1^*(n) \tag{3.5}$$

$$\widehat{\nabla}(J(n)) = -2\mathbf{x}_2(n)x_1^*(n) + 2\mathbf{x}_2(n)\mathbf{x}_2^H(n)\widehat{\mathbf{w}}(n) \tag{3.6}$$

Substituting (3.6) into (3.2) gives the following tap-weight adaptation:

$$\widehat{\mathbf{w}}(n+1) = \widehat{\mathbf{w}}(n) + \mu \mathbf{x}_2(n)[x_1^*(n) - \mathbf{x}_2^H(n)\widehat{\mathbf{w}}(n)] \tag{3.7}$$

The output signal produced is given by:

$$y_1(n) = x_1(n) - \widehat{\mathbf{w}}^H(n)\mathbf{x}_2(n) \tag{3.8}$$

As can be seen the LMS algorithm is a very simple optimization process. Among it's advantages are that it does not require measurements of correlation functions or matrix inversions. The algorithm does not converge to the Wiener solution due to gradient noise, however it does display two convergence behaviors when μ is properly chosen:

1. Convergence in the mean to the Wiener solution

$$E[\widehat{\mathbf{w}}(n)] \xrightarrow[n\to\infty]{} \mathbf{w}_0 \tag{3.9}$$

   where $\mathbf{w}_0$ is the Wiener solution defined by $\mathbf{R}\mathbf{w}_0 = \mathbf{p}$.

2. Convergence in the mean square:

$$J(n) \xrightarrow[n\to\infty]{} J(\infty) \tag{3.10}$$

   where $J_{min} < J(\infty) < \infty$ is a finite value but obviously greater than the minimum mean squared error that corresponds to the Wiener solution.

The convergence of the algorithm depends on the choice of the step-size parameter μ. Small μ gives slow adaptation, which is equivalent to long system "memory". The choice of large μ gives faster adaptation at the cost of larger excess mean-squared error (smaller "memory" degrades the estimation performance and gives larger steady-state error).

Detailed analysis of the convergence properties of the algorithm and the choice of μ are given in [20] and references within.

A useful variant of the LMS algorithm is the Normalized LMS (NLMS)[20]. The standard LMS algorithm suffers from gradient noise amplification when $\mathbf{x}_2(n)$ is large (since the tap-weight correction is proportional to the tap-input $\mathbf{x}_2(n)$). To overcome this, the tap-weight correction is normalized with respect to the squared Euclidian norm of the filter input at time n. Thus the tap-weight adaptation becomes:

$$\widehat{\mathbf{w}}(n+1) = \widehat{\mathbf{w}}(n) + \frac{\mu}{\|\mathbf{x}_2(n)\|^2}\mathbf{x}_2(n)[x_1^*(n) - \mathbf{x}_2^H(n)\widehat{\mathbf{w}}(n)] \tag{3.11}$$

3.2.2 RLS Method

The Recursive Least Squares (RLS) method is an algorithm for the design of adaptive transversal filters. Given the least squares estimate of the tap-weight filter at time $n-1$, the algorithm computes an updated estimate given the new data at time n. The algorithm utilizes at each update all the input data back to the point the algorithm was initiated.

The exponentially weighted least squares cost function to be minimized is:

$$J(n) = \sum_{i=1}^{n} \lambda^{n-1}|y_1(i)|^2 \tag{3.12}$$

The weighting factor λ ensures that data from the distant past is "forgotten" thus allowing for filter adaptation in non-stationary environments. The optimum tap weight vector that minimizes the cost function is defined by:

$$\mathbf{R}(n)\widehat{\mathbf{w}}(n) = \mathbf{p}(n) \tag{3.13}$$

where $\mathbf{R}(n)$ is an estimate of the correlation matrix with exponential weighting, defined as:

$$\mathbf{R}(n) = \sum_{i=1}^{n} \lambda^{n-1}\mathbf{x_2}(i)\mathbf{x_2}^H(i) \quad n = 1, 2, \ldots \tag{3.14}$$

and $\mathbf{p}(n)$ an exponentially weighted cross-correlation vector:

$$\mathbf{p}(n) = \sum_{i=1}^{n} \lambda^{n-1}\mathbf{x_2}(i)x_1^*(i) \tag{3.15}$$

The recursion equations for updating these expressions are given by:

$$\mathbf{R}(n) = \lambda \mathbf{R}(n-1) + \mathbf{x_2}(n)\mathbf{x_2}^H(n) \quad n = 1, 2, \ldots \quad (\mathbf{R}(0) = 0) \tag{3.16}$$

and:

$$\mathbf{p}(n) = \lambda \mathbf{p}(n-1) + \mathbf{x_2}(n)x_1^*(n) \tag{3.17}$$

Through matrix algebra relations it is possible to compute recursively the inverse of the correlation matrix. Thus the tap-weight vector $\widehat{\mathbf{w}}$ can be estimated without the need to invert $\mathbf{R}$ directly [20].

The rate of convergence of the RLS algorithm is typically of an order of magnitude faster than the LMS, however this is achieved at the cost of a large increase in computational complexity. Different versions of the algorithm, as well as in-depth analysis of the algorithm properties are given in [20] and references within.

3.3 Dual Channel Interference

Inherent in the previously described approach is the assumption that there is no leakage of the primary input signal s_1 into the interfering signal s_2. The cross channel interference scenario is depicted in Fig.3.3.

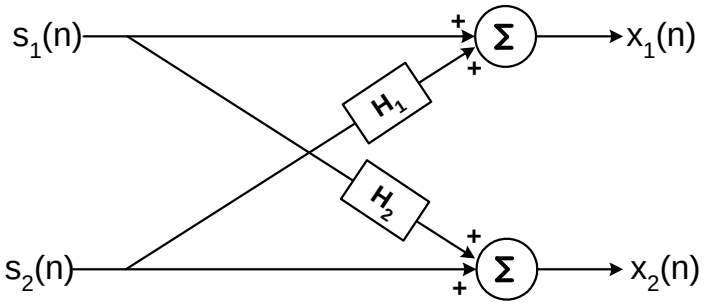


Figure 3.3: Cross Channel Interference

The naive reconstruction approach is to use two LS reconstruction systems (Fig.3.2), to create output signals y_1 and y_2. However, when both signals are coupled into each sensor, the performance of the LS systems may deteriorate. The LS approach will cause a portion of the primary signal to be cancelled out while removing the interfering signal, thus distorting the recovered signal.

Widrow *et al.* [47] analyze the amount of distortion caused to the primary signal, concluding that if the signal-to-noise ratio is high enough in the primary signal and the energy seepage into the reference signal is low enough, then the distortion is not enough to render the cleaning process useless. However, they do not offer solutions to counter the distortion.

An in-depth analysis of the cross channel scenario is offered by Weinstein *et al.*

in [46]. The coupling system analyzed is of the form:

$$\mathcal{H} = \begin{bmatrix} 1 & H_{12} \\ H_{21} & 1 \end{bmatrix} \tag{3.18}$$

The system is assumed to be invertible, i.e., $\det(\mathcal{H}) = 1 - H_{12}H_{21} \neq 0 \quad \forall\omega$, and $s_1(n)$ and $s_2(n)$ are assumed to result from uncorrelated wide-sense stationary random processes, with zero mean. This is equivalent to:

$$E\{s_1(n)s_2^*(n-\tau)\} = 0 \qquad \forall\tau \tag{3.19}$$

The aim is to find estimates $\widehat{H}_{12}$ and $\widehat{H}_{21}$ of H_{12} and H_{21}, respectively, and an appropriate reconstruction system, so that the reconstructed signals $y_1(n)$ and $y_2(n)$ be uncorrelated:

$$E\{y_1(n)y_2^*(n-\tau)\} = 0 \qquad \forall\tau \tag{3.20}$$

This implies the Cross-Spectra $P_{y_1y_2}(\omega)$ is zero for all ω.

Using the relationships between the Power-Spectra of the inputs and outputs of an LTI system, the following de-correlation criterion is obtained:

$$\begin{aligned} P_{y_1y_2}(\omega) = {} & P_{x_1x_2}(\omega) - \widehat{H}_{12}(\omega)P_{x_2x_2}(\omega) \\ & - \widehat{H}_{21}^*(\omega)P_{x_1x_1}(\omega) + \widehat{H}_{12}(\omega)\widehat{H}_{21}^*(\omega)P_{x_2x_1}(\omega) = 0 \end{aligned} \tag{3.21}$$

Based on this criterion several possible algorithms were developed for signal separation.

Two implementations of reconstruction systems are offered in [46]. These implementations are depicted in Fig.3.4 and Fig.3.5.

3.3.1 Recovery System 1

In this system a post-processing filtering stage, $\frac{1}{1-\widehat{H}_1\widehat{H}_2}$, is appended to the decorrelation system, as depicted in 3.4. The post filters counter the distortion caused to the decorrelated signals.

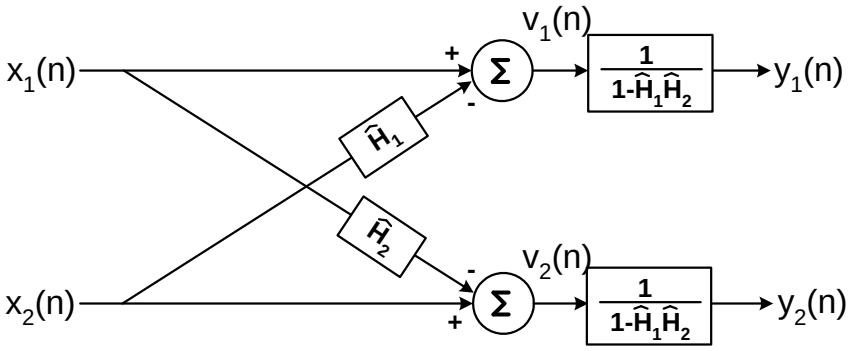


Figure 3.4: First Reconstruction System

The decoupling filters $\widehat{H}_{12}$ and $\widehat{H}_{21}$ are assumed to be discrete time casual FIR filters:

$$\widehat{H}_{12}(\omega) = \sum_{k=0}^{q_1} a_k e^{-j\omega k} \tag{3.22}$$

$$\widehat{H}_{21}(\omega) = \sum_{k=0}^{q_2} b_k e^{-j\omega k} \tag{3.23}$$

Generating the estimated signal under the reconstruction system presented in Fig.3.4, is given by:

$$v_1(n) = x_1(n) - \sum_{k=0}^{q_1} a_k x_2(n-k) \tag{3.24}$$

$$v_2(n) = x_2(n) - \sum_{k=0}^{q_2} b_k x_1(n-k) \tag{3.25}$$

and y_1 and y_2 are generated from v_1 and v_2 by the relation:

$$\sum_{k=0}^{q_1+q_2} d_k y_i(n-k) = v_i(n) \qquad i = 1, 2 \tag{3.26}$$

$$d_k = \delta_k - \sum_{l=0}^{k} a_l b_{k-l} \qquad k = 0, 1, ..., (q_1 + q_2) \tag{3.27}$$

Developing the cross and auto power spectrum relations between s_i and v_i, defined by the recovery system, and requiring that the de-correlation criterion (3.21) be met, give the following relations:

$$P_{x_2v_2}(\omega)\widehat{H}_{12}(\omega) = P_{x_1v_2}(\omega) \tag{3.28}$$

$$P_{x_1v_1}(\omega)\widehat{H}_{21}(\omega) = P_{x_2v_1}(\omega) \tag{3.29}$$

In the time domain the equations become:

$$\sum_{k=0}^{q_1} a_k c_{x_2v_2}(\tau - k) = c_{x_1v_2}(\tau) \tag{3.30}$$

$$\sum_{k=0}^{q_2} b_k c_{x_1v_1}(\tau - k) = c_{x_2v_1}(\tau) \tag{3.31}$$

where $c_{x_iv_j}(\tau)$ is the cross-correlation:

$$c_{x_iv_j}(\tau) = E\{x_i(n)v_j^*(n-\tau)\} \tag{3.32}$$

Defining $\mathbf{c}_{x_i\mathbf{v}_j} = E\{\mathbf{v}_j^* x_i(n)\}$ and $\mathbf{C}_{\mathbf{x}_i\mathbf{v}_j} = E\{\mathbf{v}_j^*(n)\mathbf{x}_i^T(n)\}$ allows concatenating equations (3.30) and (3.31) into matrix form:

$$\mathbf{C}_{\mathbf{x}_2\mathbf{v}_2}\mathbf{a} = \mathbf{c}_{x_1\mathbf{v}_2} \tag{3.33}$$

$$\mathbf{C}_{\mathbf{x}_1\mathbf{v}_1}\mathbf{b} = \mathbf{c}_{x_2\mathbf{v}_1} \tag{3.34}$$

Substituting the correlation functions by sample based approximations, recursive and sequential algorithms are developed for adjusting $\mathbf{a}$ and $\mathbf{b}$.

The recursive (RLS like) algorithm [46]:

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mathbf{Q}(n)\mathbf{v}_2^*(n)v_1(n;\mathbf{a}(n-1)) \tag{3.35}$$

where $v_1(n;\mathbf{a}(n-1))$ is the estimate of signal v_1 at time n based on the previous estimate of $\mathbf{a}$, and,

$$\begin{aligned}\mathbf{Q}(n) &= [\sum_{k=1}^{n} \beta_1^{n-k}\mathbf{v}_2^*(k)\mathbf{x}_2^T(k)]^{-1} = \\ &= \frac{1}{\beta_1}[\mathbf{Q}(n-1) - \frac{\mathbf{Q}(n-1)\mathbf{v}_2^*(n)\mathbf{x}_2^T(n)\mathbf{Q}(n-1)}{\beta_1 + \mathbf{x}_2^T(n)\mathbf{Q}(n-1)\mathbf{v}_2^*(n)}]\end{aligned} \tag{3.36}$$

Replacing **a** and **b** by their current estimates gives a sequential (LMS like) algorithm:

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mu \mathbf{v}_2^*(n; \mathbf{b}(n-1)) v_1(n; \mathbf{a}(n-1)) \tag{3.37}$$

$$\mathbf{b}(n) = \mathbf{b}(n-1) + \mu \mathbf{v}_1^*(n; \mathbf{a}(n-1)) v_2(n; \mathbf{b}(n-1)) \tag{3.38}$$

An open issue in the implementation of this system has to do with the implementation of the post filter $G = \frac{1}{1-\widehat{H}_1\widehat{H}_2}$. Direct computation of the IIR filter G from the estimated coefficients of filters H_1 and H_2 is problematic. While stability of this filter is assured when the exact coefficients of H_1 and H_2 are known ($\det(\mathcal{H}) = 1 - H_{12}H_{21} \neq 0 \quad \forall\omega$) this is not the case when using the estimates, and stabilizing procedures need be taken.

A FIR approximation of G is proposed in [36]. In this method an approximation $\widehat{G}(m)$ is created by:

$$\widehat{G}(m) = \sum_{k=1}^{m} (H_{12}H_{21})^{k-1} \tag{3.39}$$

which in the spatial domain becomes:

$$\begin{aligned} \widehat{g}_m(x,y) = \delta(x,y) + \overbrace{\mathbf{a} * \mathbf{b}}^{once} + \overbrace{\overbrace{\mathbf{a} * \mathbf{b}}^{} * \overbrace{\mathbf{a} * \mathbf{b}}^{}}^{twice} + \cdots \\ + \overbrace{\overbrace{\mathbf{a} * \mathbf{b}}^{} * \cdots * \overbrace{\mathbf{a} * \mathbf{b}}^{}}^{m-1 \quad times} \end{aligned} \tag{3.40}$$

The filter support is proportional to m. Thus, for large m the border effects on finite size images are significant.

Computing G directly from the filter estimates also means that transients and errors in the LS estimation process will also be directly reflected in the post processing stage. We therefore prefer using the filtered signals to estimate the post filters. The filters are estimated using an LMS process between signals v_1 and x_1 and between signals v_2 and x_2. For further details see 4.

3.3.2 Recovery System 2

An alternative reconstruction system is depicted in Fig.3.5.

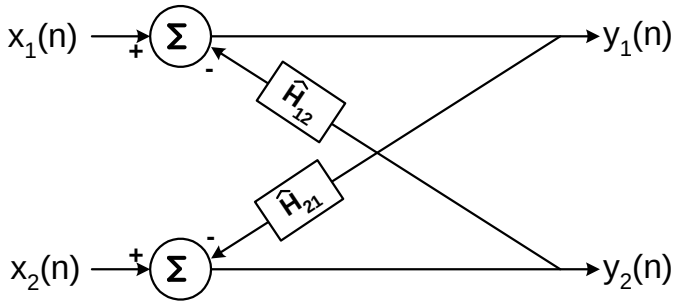


Figure 3.5: Second Reconstruction System

In this scheme the estimated signal is generated by:

$$y_1(n) = x_1(n) - \sum_{k=0}^{q_1} a_k y_2(n-k) \tag{3.41}$$

$$y_2(n) = x_2(n) - \sum_{k=0}^{q_2} b_k y_1(n-k) \tag{3.42}$$

Developing an LMS-like filter coefficient update equation, for each of the two summation operations, gives the following sequential algorithm [46]:

$$\mathbf{a}(n) = \mathbf{a}(n-1) + \mu \mathbf{y}_2^*(n; \mathbf{b}(n-1)) y_1(n; \mathbf{a}(n-1)) \tag{3.43}$$

$$\mathbf{b}(n) = \mathbf{b}(n-1) + \mu \mathbf{y}_1^*(n; \mathbf{a}(n-1)) y_2(n; \mathbf{b}(n-1)) \tag{3.44}$$

An RLS-like algorithm can be developed in a similar manner [46].

While seemingly, this approach is superior to the previous one (no post filtering), there is an inherent problem in expanding the algorithm to two dimensions. No matter in what order the images are scanned, the 2-D filters are no longer casual. This is not a problem when the reference image is known (as in recovery system 1), but this is not the case when both signals are reconstructed in unison, each serving as a reference for the other.

The recovery algorithm needs to be adjusted to solve this problem. We considered the following alterations:

- Adding a delay to the coefficient updating mechanism. The length of the delay is proportional to the size of the filter and the image size (row length).
- Recovering the image in two passes - one forward, one backward. During the forward pass, only the casual part of the filter is updated. During the backward pass the whole filter can be updated.

In our simulations we found the second method preferable.

Chapter 4

Solving the Show-Through Problem by Decorrelation

In this chapter we present the basic show-through removal algorithm proposed by Sharma [37], point out it's weaknesses and propose an improved algorithm.

4.1 Basic Algorithm

The show-through removal algorithm proposed in [37] is a 2-D adaptation of 1-D echo-cancellation techniques used in telephony. Adaptive linear filters estimate and track the show-through point spread function through an LMS process.

The Algorithm:

1. Manually estimate the white paper reflectance R_{pw} (averaging reflectance values of areas with no print on either side).

2. Convert front-side reflectance values to density (equation 2.13).

3. Convert back-side reflectance values to absorbance (equation 2.14).

4. For each pixel (progressing in a spatial contiguous order):

   (a) Compute show-through corrected density $\widehat{D}_f$ from the scanned front-side density D_f^s and back-side absorbance A_b^s:

   $$\begin{aligned}\widehat{D}_f(m,n) &= D_f^s(m,n) \\ &- \sum_{k=-N}^{k=N} \sum_{l=-N}^{l=N} w(k,l) A_b^s(m-k, n-l).\end{aligned} \tag{4.1}$$

   where w are the filter tap weights.

   (b) If back side has activity but not front side, update filter coefficients by LMS method

   $$w'(k,l) = w(k,l) + \mu \widehat{D}_f(m,n) A_b^s(m-k, n-l). \tag{4.2}$$

   $$w(k,l) = w'(k,l).$$

   A pixel at location (m,n) is deemed active if the minimum value over neighboring pixels is below a certain percentage of the estimated white paper reflectance.

   (c) Convert density to reflectance.

5. Repeat steps 1-4 for the back-side image.

4.1.1 Algorithm Analysis

The algorithm proposed by Sharma suffers from several faults and weaknesses which we rectified.

The first issue we addressed was that of the background estimation. Other than the obvious need to automate the process we also found that using only a global reflectance value (a single value for the whole image) may lead to unwanted "whitening" artifacts in the recovered images. We propose using "local" (location dependent) background reflectance values.

Another problem is related to inherent limitations in the LMS process, using a single filtering stage as proposed by Sharma may not be the best strategy. We propose a cascaded filtering structure to improve the results.

Thirdly, a structural fault with Sharma's algorithm is that it uses the standard LMS recovery system (Fig.3.2). Obviously the show-through problem is a case of dual channel interference. Thus, an appropriate recovery system is required.

The following sections provide the details of our modifications.

4.2 Proposed Algorithm Development

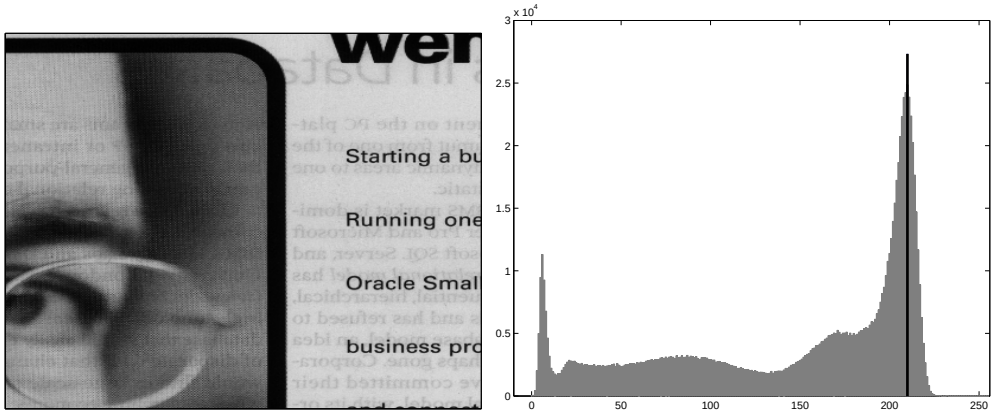
4.2.1 Background Estimation

4.2.1.1 Global Background

The original algorithm described in [37] computes the white paper reflectance R_{pw} by taking the average reflectance value of a manually selected area of the images, not containing print on either the front or back side. We propose an automatic method

based on the *mean shift* algorithm proposed in [12] and [13].

Typically, the brightness histogram of a scanned document is multi-modal (Fig. 4.1). We define the white paper reflectance as the peak of the brightest (rightmost) mode in the image histogram. While other choices are also possible (such as taking the brightest pixel value in the image), our choice is both intuitive and robust, as well as giving good results.



(a) Image (b) Image histogram

Figure 4.1: Multi-modal PDF of image (R_{pw} marked by dark line).

The mean shift algorithm is a simple non-parametric technique for estimation of the probability density gradient. An iterative steepest ascent algorithm is implemented, operating on the marginal probability density (the brightness histogram). Beginning at an initial position - a certain brightness level, the algorithm computes the mean value of neighboring brightness levels. The algorithm then "shifts" to this mean position and repeats the process until convergence.

The algorithm, (with properly selected parameters) converges to a local probability density maximum (mode peak). Details on setting the parameters (the size of neighborhood and the initial location) are discussed following the algorithm steps.

The Mean Shift Algorithm:

1. Choose the search window radius (in our case the number of brightness levels).
2. Choose initial window center location (in our case the brightness level).
3. Compute mean value of pixels contained in the window and shift window center to it .
4. Repeat step 3 until convergence.

The point of convergence obviously depends on the starting point. Since we are seeking the peak of the brightest mode, the brightest pixel value in the image is chosen as the initial window position.

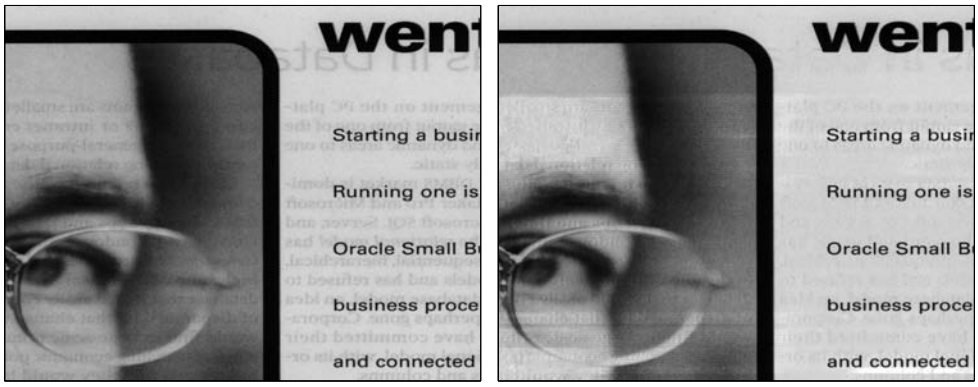
As a rule of thumb - the smaller the window radius - the better the separation resolution of the modes. However, if the radius is too small, the algorithm may get stuck on local maxima caused by discontinuities in the PDF estimate. Thus we want to find the maximal radius that gives us sufficient resolution, depending on the pixel distribution statistics of the image. In [12] the window radius chosen is proportional to the STD of the pixel brightness values. Values between 0.2σ and 0.4σ were chosen according to the application. We found that a radius of 0.1σ produced good results on the images tested.

We denote the *global white paper reflectance* values obtained by the above algorithm, R^f_{pg} and R^b_{pg} for the front and back-side images, respectively.

4.2.1.2 Local Background

The cleaning processes described in [37] attempts to cancel or minimize in some sense (specifically least mean square) a difference, or error, value. According to the normalization proposed in [37], the zero value corresponds to the global white paper

reflectance. This means that when removing show-through, the algorithm attempts to bring the brightness value closer to R_{pg}, whether or not this is the "desired" value. The effect will be one of over "whitening" of areas in the image that contained show-through, as can be seen in Fig.4.2 (notice especially the man's forehead). The desired value we are seeking is that of the background reflectance in the vicinity of the pixel being cleaned.



(a) Before (b) After

Figure 4.2: Whitening Artifacts

Typically, least-squares algorithms operate on zero mean signals. This is due to the fact that separation of the DC component in signal mixtures is an ill-conditioned problem [36]. The normalization by R_{pw} serves to set the background, or "DC", value to zero. However, as can be seen in Fig.4.2, the non-stationarity of the scanned images pertains (among other things) to non-constant DC (or background) values. Therefore, we conclude that the normalization should be done on a local level according to local background levels.

We define R_{pl}, the local background value, as the gray level value at the peak of the brightest mode, in a square area ($L \times L$ pixels) surrounding each pixel location. This value can generally be found using a mean-shift process, with a few alterations, as explained below. The collection of background values for all pixel locations define

a background image.

As before, the brightest pixel value in the area, is chosen as the initial window position. However, since the sample size is limited to a small area surrounding each pixel, the histogram may not give a good approximation of the density. Too small a mean-shift window radius may cause the mean-shift process to get stuck on an isolated value or local maxima that is not a true mode. In other cases the image segment analyzed contains only one narrow mode. This typically happens in image segments that are more or less uniform (such as segments with no print). In these cases, setting the window size as a small fraction of the STD, gives too small a value that may result in the algorithm getting stuck. A reasonable choice for the window radius (taking into account that the STD may be very low) may be as high as 0.5σ.

The solution to these problems is an *adaptive mean-shift* process [13] in which the window radius is allowed to vary throughout the iterations. The minimal radius is set in proportion to the STD, however if the window covers too few pixels (less than a certain percentage of the $L \times L$ area pixels), the radius is increased until the window covers enough pixels. Increasing window size guarantees that the algorithm does not get stuck on isolated values, instead of an actual mode. We used 5% of the pixels as the minimal window coverage for a window of size $L \times L = 31 \times 31$.

In some cases, where back-side activity is high, it is impossible to estimate the correct local background value based on the immediate vicinity alone, since too many of the pixels are contaminated by show-through. In these cases the best we can do is use the global value.

Back-side activity is measured by comparing the local average back-side reflectance (measured on neighboring pixels) value to back-side R_{pg} and by comparing average

back-side and front-side reflectance values. The darker the back-side the more likely it is to be active. By comparing the back-side and front-side values we can identify areas where the back-side activity results from show-through. If the average back-side value is lower than a fraction (say $60\% - 75\%$) of the back-side R_{pg} and if the local mean back-side value is lower than the local mean front-side value, the back-side is deemed active. In this case the global front-side value is chosen.

The process described above is computationally heavy - a mean-shift process is run for most pixels. Several measures may be taken to reduce the computational load. Local background values need not be calculated for every pixel. Instead, we found it sufficient to re-calculate at horizontal and vertical intervals equal to half the dimensions of the window. In order to avoid unwanted discontinuities, the local background images are smoothed by low-pass filters (Gaussian, 15×15 pixels, std = 2). Additionally, when the STD is very low (for example less than 20% of the STD of the total image) ,the data is typically single modal, and it is sufficient to use the mean pixel value in the neighborhood, instead of completing a mean-shift process.

A similar process is run for the back-side. We denote the local background reflectance values $R^f_{pl}(m,n)$ and $R^b_{pl}(m,n)$, for front-side and back-side, respectively. These values replace the R_{pw} when calculating density and absorbance thus providing adaptation to local background levels.

To summarize, the procedure for calculating the local background reflectance is given by:

1. Compute front-side global background reflectance R^f_{pg} by a mean-shift process run on the whole image.

2. For selected pixel locations (m, n) if back-side activity is high:

   Set local value equal to global value $R^f_{pl}(m, n) = R^f_{pg}$

   else:

   Compute local value by an adaptive mean-shift process.

3. Interpolate background image for locations not set in previous step (nearest neighbor interpolation).

4. Smooth local background image R^f_{pl} with Gaussian filter.

5. Repeat steps 1-4 for the back-side image.

Using this approach greatly reduces the "whitening" artifacts. A demonstration of the results obtained with this algorithm is deferred to section 4.3.

4.2.2 Filter Cascade

One of the drawbacks of a MSE process, is that it has a tendency to concentrate it's effort on the larger filter coefficients, where most of the energy is concentrated. In doing so the process neglects the smaller coefficients which may not converge correctly.

The proposed solution to this problem is a cascaded filter structure (Fig. 4.3), replacing the single filtering stage (Fig. 3.2), with a large filter support, in the original algorithm. The filter support is increased with each stage of the cascade (for example a 3-stage cascade with filter supports of 5×5, 9×9 and 15×15).

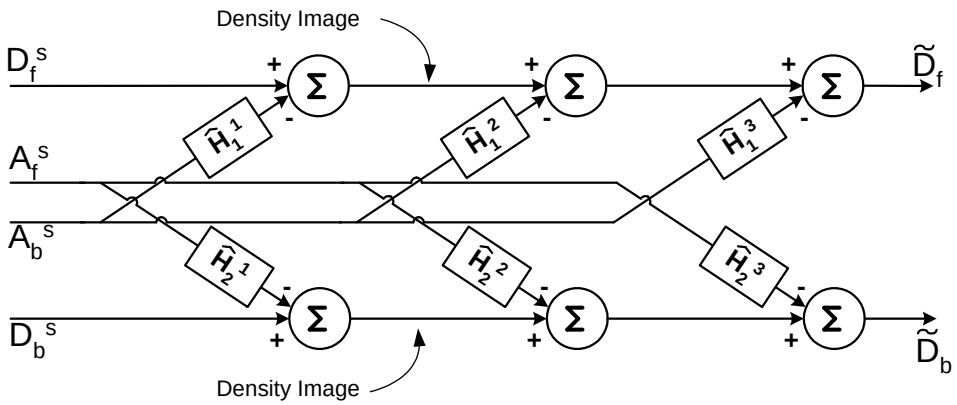


Figure 4.3: Cascaded filter structure.

Each stage uses the previous stage output as it's main input, but uses the same original reference (interference) signal, as shown in Fig.4.3. Another option is to compute new reference signals at each stage, from the previous stage output (converted to absorbance). We found this option less preferable since recomputing the reference signal, from partially cleaned signals, hurts the stability of the cleaning process.

Predictably, the later filtering stages contribute less with each stage. The first filtering stage produces the biggest improvement both visually and in similarity measures. Both mutual information and cross correlation were calculated between front and back side images. We expected these values to decrease in magnitude with the reduction of show-through. The latter filtering stages did provide visual improvement. However, improvement in the similarity measures was found to be negligible. Is some cases even a certain degradation occurs. We found that two or three filtering stages are appropriate.

4.2.3 Post Processing

When using the reconstruction system depicted in Fig.3.4 a post processing stage is required. Estimating the post-processing filter is in itself not a trivial process. For linear mixtures this filter is $G = \frac{1}{1-\widehat{H}_{12}\widehat{H}_{21}}$ but in our case the distortion caused to

the primary signal by the LS process is more complex due to the non-linearity of the system. In light of this, estimating the post-processing filter directly from the estimates of H_{12} and H_{21} is unfeasible. Instead, we propose a non-parametric approach to estimate the filter G from the data itself.

We propose using the fact that the scanned (contaminated) image, D_i^s, contains an undistorted version of the clean image and additive interference. We assume that the interference is uncorrelated with the clean signal. The show-through cleaned image, $\widetilde{D}_i$, is a distorted version of the clean image with much reduced interference. We estimate the post-processing filter and cancel the distortion using an LMS process (Fig.4.4). In theory the error signal, e, should contain, after convergence, only the uncorrelated show-through, and the filter output, $\widehat{D}_i$, the undistorted image. This process is run separately for the front and back side images.

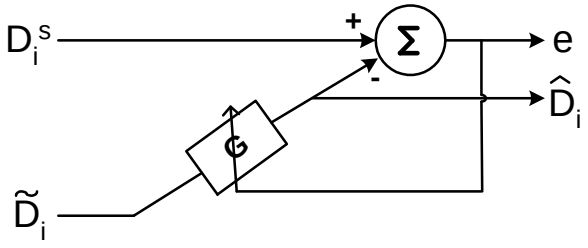


Figure 4.4: Post processing filter estimation.

This approach was found to give good results and although the distortion it corrects isn't usually significant, the resulting images are more aesthetically pleasing.

4.2.4 Summary of Proposed Algorithm

To summarize, the proposed algorithm is:

1. Estimate the global white paper reflectance, R_{pg}, using the mean-shift process on the whole image for both front and back sides.

2. Estimate the local white paper reflectance images, R_{pl}, for both front and back: local mean-shift process coupled with activity estimate.

3. Convert front-side reflectance values to density using front-side R^f_{pl} (2.13) [1].

4. Convert back-side reflectance values to absorbance around pixel location using front-side R^f_{pl} value at pixel location (2.14) [2].

5. For each pixel (progressing in a spatial contiguous order):

   (a) Compute show-through corrected density.

   (b) If back side has activity but not front side, update filter coefficients by LMS method.

6. Clip density values brighter than Density of R_{pg}.

7. Repeat stages 5 and 6 with larger filter supports (Fig.4.3).

8. Post-processing stage - cancel distortion on front and back-side recovered images using their respective scanned versions in an LMS process (Fig.4.4).

9. Convert density back to reflectance using R^f_{pl}.

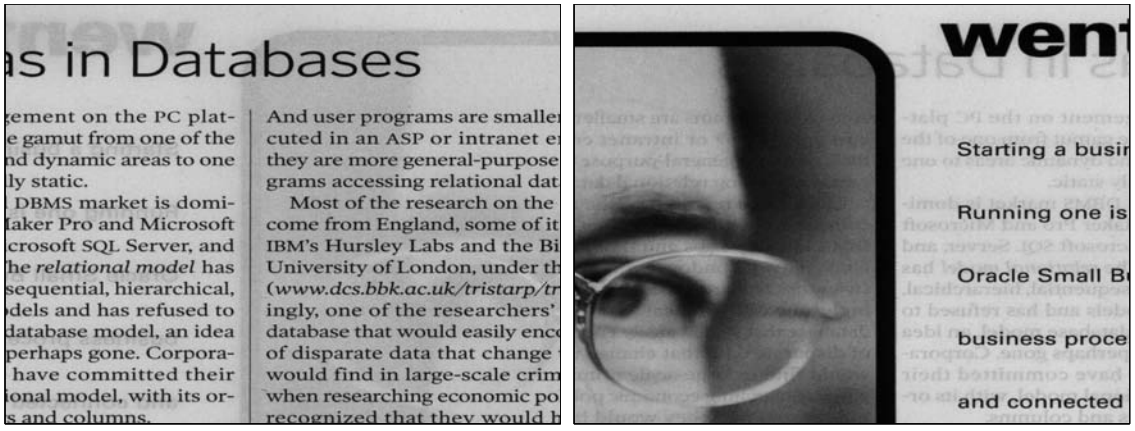
[1] Stages 3-9 are run simultaneously for the back-side.

[2] The area around the pixel location (m, n) is converted to absorbance according to $R^f_{pl}(m, n)$. This operation has heavy memory costs since for each location we must save an area of size equal to the support of the largest filter used. Alternatively, this operation can be done for each location in turn as part of stage 5.

4.3 Simulation Results

The following example is for a pair of 256 gray level images of 1000×1000 pixels. For convenience The images are shown in their correct orientation but obviously during registration and throughout the cleaning process the mirror image (flip left-right) of one of the images is used.

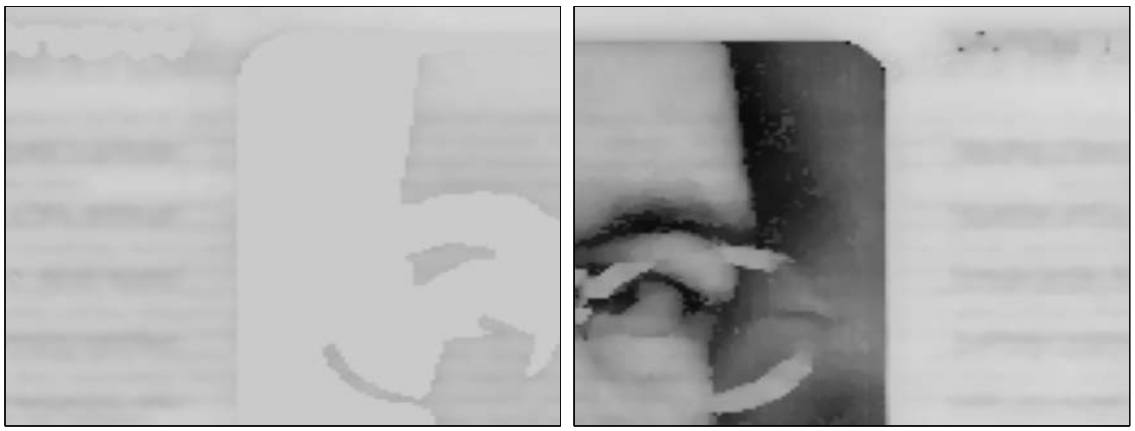
Fig.4.5 depicts the original scans. Notice that this example contains not only text (binary image) but also contains various shades of gray, thus making naive thresholding cleaning impossible.



(a) Front (b) Back

Figure 4.5: Original Images : MI = 0.105, Corr = 0.2478

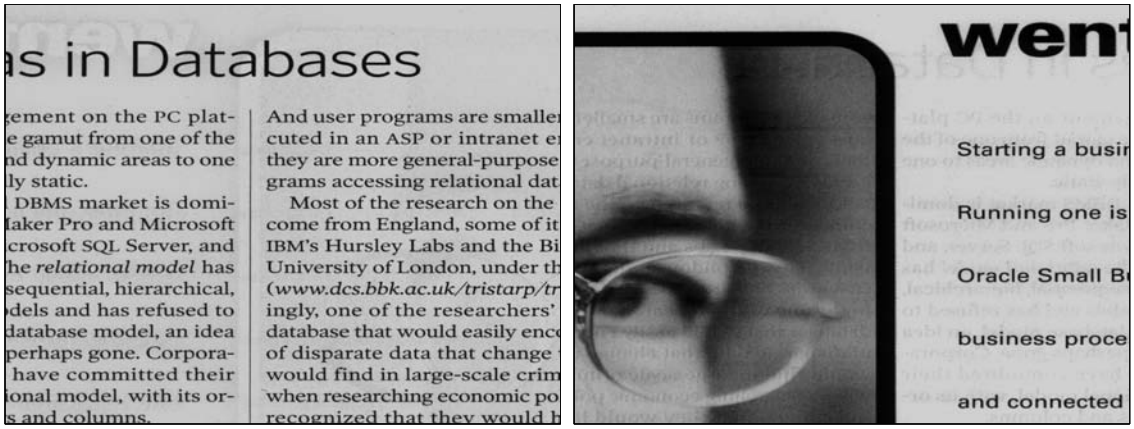
Fig.4.6 depicts the local background images. Notice the large area in the center of the front-side image where the global background value was chosen due to the large area affected by show-through caused by the dark area in the back-side. Of particular importance is the correct background estimation in the left of the back-side image (the man's forehead for instance).



(a) Front (b) Back

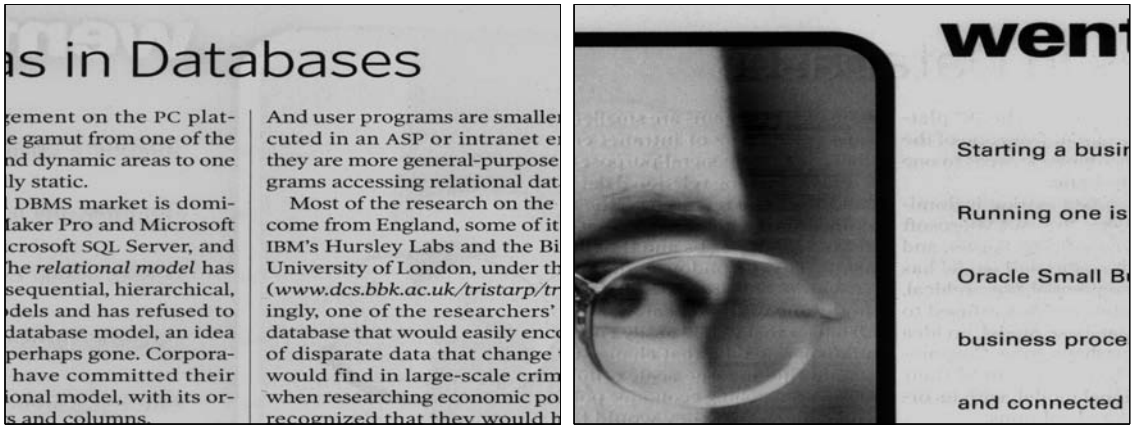
Figure 4.6: R_{pl} Images. The global background values for these images : $R^f_{pg} = 203$, $R^b_{pg} = 209$

Figures 4.7-4.9 show the output of each of three cascaded filtering stages. Visual improvement is evident, with each stage improving on the previous stage output (notice for example the text show-through in the right side of the back-side image) even though the objective measures worsen somewhat (although all still improve on the original measures).



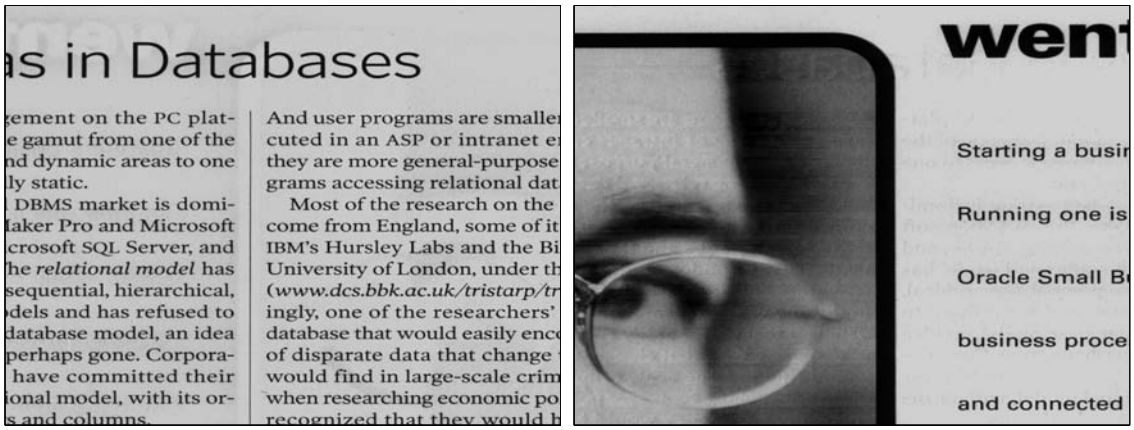
(a) Front (b) Back

Figure 4.7: First Stage Output (5×5 filter): MI = 0.025, Corr = 0.002



(a) Front (b) Back

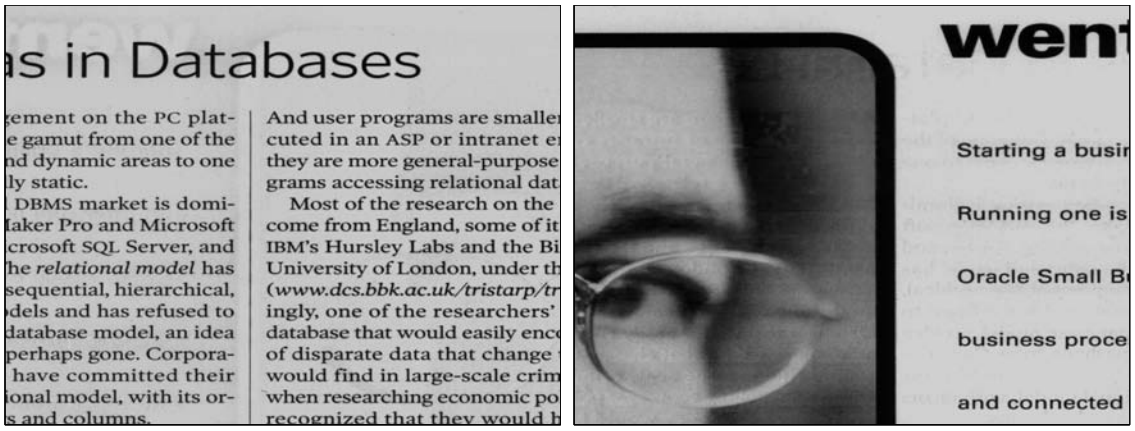
Figure 4.8: Second Stage Output (9×9 filter): MI = 0.027, Corr = Corr = 0.022



(a) Front (b) Back

Figure 4.9: Third Stage Output (15×15 filter): MI = 0.027, Corr = Corr = 0.035

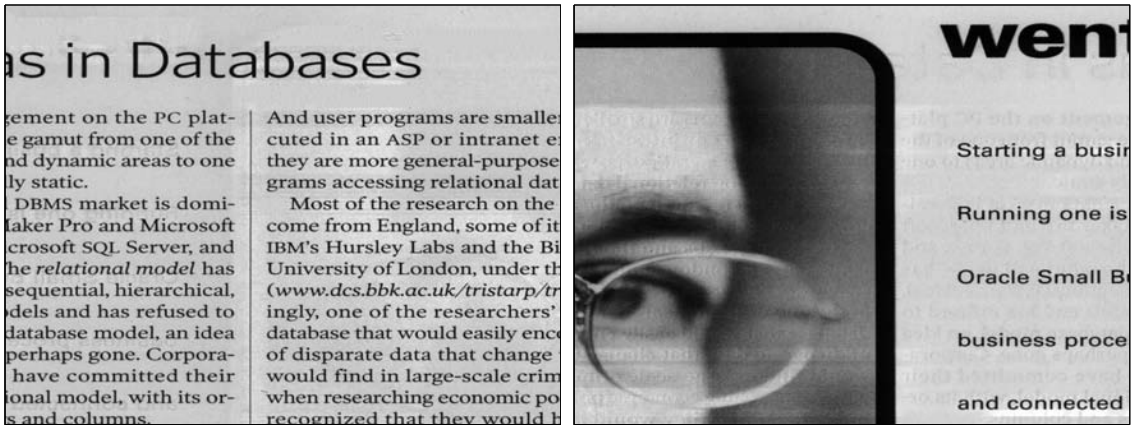
Fig.4.10 depicts the final output (after the post-processing stage) of our improved cleaning algorithm. While the effects of the post-processing stage are subtle, these images are more aesthetically pleasing than the output of the filter cascade (Fig.4.9) and bear closer likeness to the original images (for example in textures) but without the show-through.



(a) Front (b) Back

Figure 4.10: Post-Processing Output : MI = 0.029, Corr = 0.013

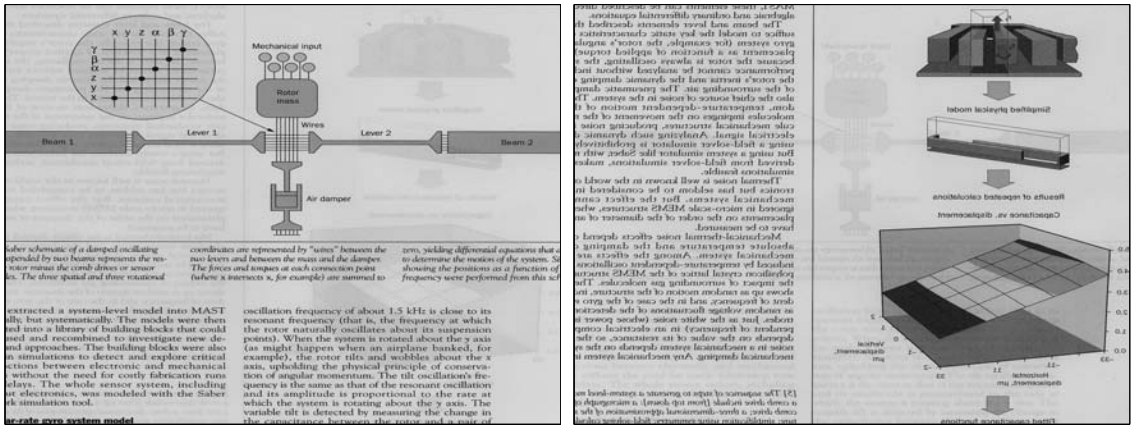
In comparison to our results, Figure 4.11 shows the results of Sharma's algorithm for the same images.



(a) Front (b) Back

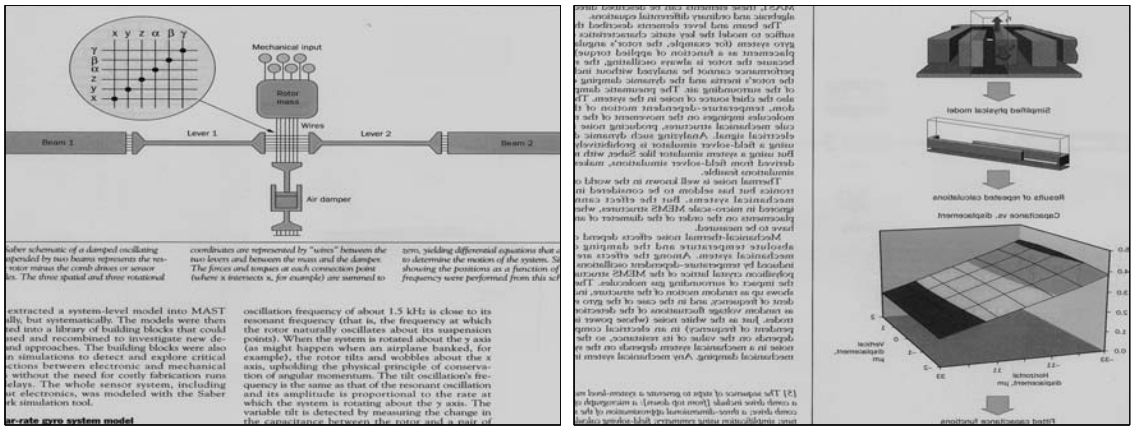
Figure 4.11: Output of Sharma's Algorithm : MI = 0.027, Corr = 0.052

Fig.4.12 depicts another example for a pair of 2000 × 2000 pixel images. The cleaned images are shown in Fig.4.13 and zooms of interesting parts in Figures 4.14 and 4.15.



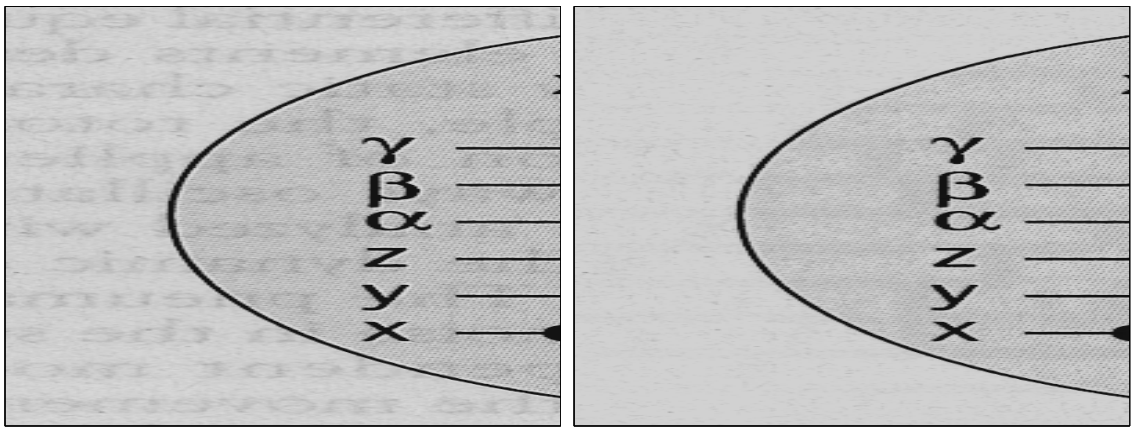
(a) Front (b) Back

Figure 4.12: Original Images : MI = 0.050, Corr = 0.129



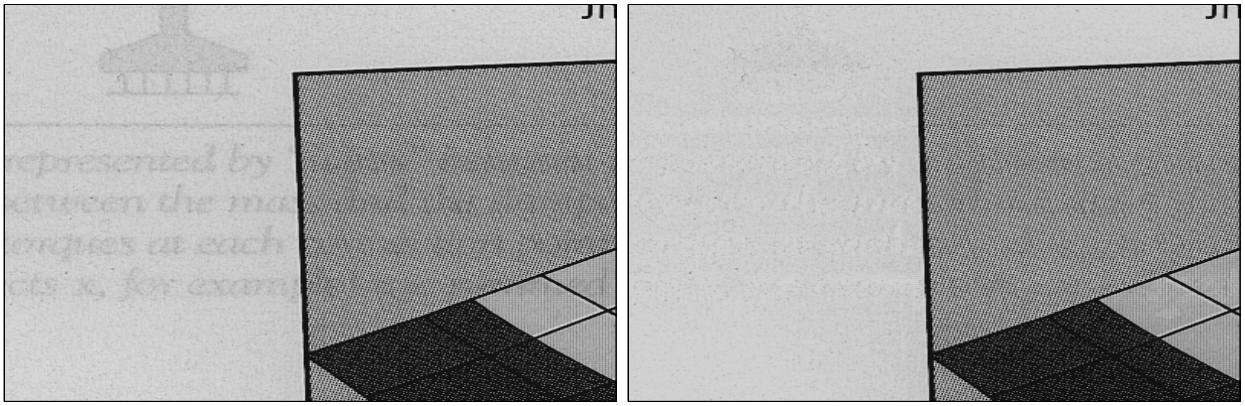
(a) Front (b) Back

Figure 4.13: Full algorithm's output : MI = 0.006, Corr = 0.012



(a) Before (b) After

Figure 4.14: Zoom of front side



(a) Before (b) After

Figure 4.15: Zoom of back side

In conclusion, in this chapter we presented a decorrelation-based Show-Through removal algorithm that improves on previous results. The algorithm adapts to local image behavior while using a more suitable filtering scheme to achieve better show-through removal with significantly less artifacts.

Chapter 5

Maximum-a-Priori Models

In recent years Bayesian estimation has gained popularity as a framework for image restoration and much work has been done in the field. The BSS problem can also be formulated as a Bayesian estimation problem. This approach allows easy incorporation of prior knowledge, thus regularizing and stabilizing an otherwise ill-conditioned problem.

Under the Bayesian approach a likelihood function $P(\mathbf{Y}, \mathcal{H}, \mathbf{X})$ is formulated. Where $\mathbf{Y}$ are the estimated images, $\mathcal{H}$ the estimated mixing operator and $\mathbf{X}$ the measured images. The image reconstruction thus becomes:

$$< \mathbf{Y}^*, \mathcal{H}^* >= \arg\max_{\mathbf{Y},\mathcal{H}} P(\mathbf{Y}, \mathcal{H}, \mathbf{X}) \tag{5.1}$$

Decomposing the probability expression:

$$P(\mathbf{Y}, \mathcal{H}, \mathbf{X}) = P(\mathbf{X}|\mathbf{Y}, \mathcal{H})P(\mathbf{Y})P(\mathcal{H}) \tag{5.2}$$

allows us to formulate the problem as a Maximum-a-Priori problem:

$$< \mathbf{Y}^*, \mathcal{H}^* >= \arg\max_{\mathbf{Y},\mathcal{H}} P(\mathbf{X}|\mathbf{Y}, \mathcal{H})P(\mathbf{Y})P(\mathcal{H}) \tag{5.3}$$

An equivalent formulation of the problem is that of minimizing an energy functional. This formulation can be achieved directly from the Bayesian approach (by taking the $-log$ of the likelihood function) or independently of it.

5.1 Energy Functional

The image separation process is thus based on the unconstrained minimization of an energy functional, typically composed of two main terms:

A. A Fidelity term describing the likelihood of our estimation, i.e. the closeness of the estimate to the measured images.

B. A Regularization term describing prior knowledge of the signals, typically controlling the irregularity of the images.

$$J = Fidelity + \lambda * Regulariztion \tag{5.4}$$

where λ is a parameter determining the trade-off between the goodness of the fit to the measured data and the amount of regularization done.

In certain situations additional penalty terms may be added to the functional.

An alternative formulation of this problem is that of a constrained minimization problem. Typically in this formulation, the regularization term is minimized subject to a known distortion level. This formulation is less convenient in our scenario where not only the images $\mathbf{Y}$ need be estimated but also the mixing operator $\mathcal{H}$.

5.1.1 Fidelity

The most popularly used fidelity term used is the mean squared error. In compact form we write:

$$Fidelity = \| \mathcal{H}(\mathbf{Y}) - \mathbf{X} \|_2^2 \tag{5.5}$$

In our case, where each mixture contains a pair of images, one of which is blurred

and non-linearly transformed, this term becomes:

$$Fidelity = \sum_{(n,m)\in S} \left[h_{11}y_1(n,m) + \sum_{(i,j)\in K} h_{12}(i,j) f(y_2(n-i,m-j)) - x_1(m,n) \right]^2 + \left[\sum_{(i,j)\in K} h_{21}(i,j) f(y_1(n-i,m-j)) + h_{22}y_2(n,m) - x_2(m,n) \right]^2 \tag{5.6}$$

where $f(\cdot)$ is the non-linear transformation.

Minimization of the fidelity term alone in terms of the joint space $\{\mathbf{Y}, \mathcal{H}\}$, is insufficient. The global minimum of this term is the trivial solution ($\{\mathbf{X}, \mathbf{I}\}$). Also, of practical concern is the fact that this term is non-convex in the joint space.

5.1.2 Regularization

The regularization term allows us to introduce prior knowledge on the signals into the process. An exact model for modelling images does not currently exist, however some reasonable priors are accepted as suitable for images. These priors will typically attempt to control the irregularity of the images since images are typically piecewise smooth functions, while at the same time attempting to preserve edges.

In the BSS framework the underlying assumption of the *independence* of the mixed images allows us to decompose the regularization term into a sum of two independent terms, one for each image.

$$Reg(\mathbf{Y}) = Reg(\mathbf{y}_1) + Reg(\mathbf{y}_2) \tag{5.7}$$

Two types of regularization terms, suitable for images, are considered in our work. The first type of terms is based on the *Markov Random Field* (MRF) model for images [26]. The second type of regularization is the *Total Variation* term.

5.1.2.1 Markov Random Fields

Under the MRF model, an image s has a distribution according to the Gibbs form:

$$P(s) = \frac{1}{Z} exp^{-U(s)} \tag{5.8}$$

where U is the prior energy, composed of a sum of potential functions over a set of cliques of interacting pixel locations. Z is a normalization factor.

The regularization term is taken to be the prior energy U. For example, considering the set of cliques composed of adjacent pixel locations $\{l, r\}$:

$$U(s) = \sum_{\{l,r\}} \phi_i(s(l) - s(r)) \tag{5.9}$$

The functional ϕ is chosen according to the degree of correlation assigned to adjacent samples. This functional describes the regularity of the image by penalizing high gradient values. However, in order to preserve steep edges in the image, edge preserving functionals are preferred.

MRF functionals may be convex, such as:

$$\phi(\Delta s) = \sqrt{(\Delta s)^2 + \varepsilon} \tag{5.10}$$

or non-convex, such as [40]:

$$\phi(\Delta s) = \frac{|\Delta s|/T}{1 + |\Delta s|/T} \tag{5.11}$$

where T is a threshold parameter. MRF functionals of varying complexity were used in the context of image separation [41, 42].

The use of convex functions allows for standard optimization methods based on gradient decent to be used. Non-convex functions require the use of specially tailored

optimization methods, such as *Graduated Non-Convexity* (GNC). Under this scheme the optimization process starts with convex functions, which are gradually morphed into the non-convex MRF functional. For details see 6.

5.1.2.2 Total Variation

The second type of regularization term considered is based on the non-linear *Total Variation* norm proposed by Rudin, Osher and Fatemi [34].

In this approach the regularization term is of the form:

$$TV(s) \equiv \int_{\Omega} |\nabla s| dxdy \equiv \int_{\Omega} \sqrt{s_x^2 + s_y^2} dxdy \tag{5.12}$$

where Ω denotes the image domain and s_x and s_y are the image derivatives in the x and y axes, respectively. The functional $TV(s)$ measures the total variation in image s. This measure has the advantage of being both convex and edge preserving. This approach has been found to be especially good in preserving sharp edges while not penalizing smooth images, in image denoising [34, 45] and restoration [50, 3].

An important aspect of the Total Variation approach is the discretization of (5.12) and the precision of the gradient estimation. Simply put, better estimation leads to a more stable algorithm and better results.

An example of a simple discretization using central derivatives is:

$$\begin{aligned} s_x(i,j) &= \frac{1}{2}(s_{i+1,j} - s_{i-1,j}) \\ s_y(i,j) &= \frac{1}{2}(s_{i,j+1} - s_{i,j-1}) \end{aligned} \tag{5.13}$$

The following, more intricate discretization, is proposed in [38], where the gradient field is computed on points in between pixel locations (Fig.5.1):

$$TV(u) = \frac{1}{2}\left[\sum_{i=1}^{n-1}\sum_{j=1}^{n}\sqrt{[s_x(i+\frac{1}{2},j)]^2 + [s_y(i+\frac{1}{2},j)]^2} + \sum_{i=1}^{n}\sum_{j=1}^{n-1}\sqrt{[s_x(i,j+\frac{1}{2})]^2 + [s_y(i,j+\frac{1}{2})]^2}\right] \tag{5.14}$$

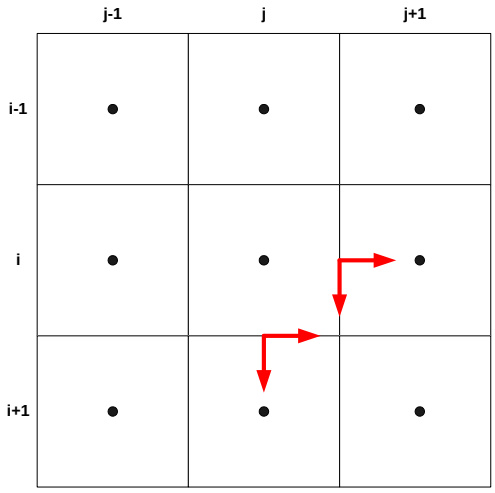


Figure 5.1: Gradient field estimation - gradient is estimated between pixel locations at half pixel horizontal/vertical shifts. At each location both vertical and horizontal derivatives are calculated.

where:

$$s_x(i+\frac{1}{2},j) = s_{i+1,j} - s_{i,j}$$
$$s_y(i,j+\frac{1}{2}) = s_{i,j+1} - s_{i,j}$$

and,

$$s_y(i+\frac{1}{2},j) = \frac{1}{2}minmod[(s_{i,j+1} + s_{i+1,j+1} - s_{i,j} - s_{i+1,j}), (s_{i,j} + s_{i+1,j} - s_{i,j-1} - s_{i+1,j-1})]$$
$$s_x(i,j+\frac{1}{2}) = \frac{1}{2}minmod[(s_{i+1,j} + s_{i+1,j+1} - s_{i,j} - s_{i,j+1}), (s_{i,j} + s_{i,j+1} - s_{i-1,j} - s_{i-1,j+1})]$$

The function *minmod* is defined as follows:

$$minmod(a,b) = \frac{sign(a) + sign(b)}{2} min(|a|,|b|)$$

5.1.3 Fidelity/Regularization Tradeoff

An issue that is often neglected is that of the tradeoff between the relative "strengths" of the fidelity term and the regularization term, determined by the parameter λ. This

is particularly true in image processing applications. In [38] Strong *et al* propose using a varying, location dependent, weight term, for the purpose of image denoising. Under this scheme, a high weight is assigned to the regularization term in smooth image areas, and a low one at image edges. Incorporating λ into the TV term, (5.14) becomes:

$$\begin{aligned} TV(s) = \frac{1}{2}\Bigg[& \lambda_{i+\frac{1}{2},j} \sum_{i=1}^{n-1}\sum_{j=1}^{n} \sqrt{[s_x(i+\frac{1}{2},j)]^2 + [s_y(i+\frac{1}{2},j)]^2} + \\ & \lambda_{i,j+\frac{1}{2}} \sum_{i=1}^{n}\sum_{j=1}^{n-1} \sqrt{[s_x(i,j+\frac{1}{2})]^2 + [s_y(i,j+\frac{1}{2})]^2} \Bigg] \end{aligned} \tag{5.15}$$

where $\lambda_{i+\frac{1}{2},j}$ and $\lambda_{i,j+\frac{1}{2}}$ are given by:

$$\begin{aligned} \lambda_{i+\frac{1}{2},j} &= \frac{1}{|s(i+1,j) - s(i,j)| + \epsilon} \\ \lambda_{i,j+\frac{1}{2}} &= \frac{1}{|s(i,j+1) - s(i,j)| + \epsilon} \end{aligned} \tag{5.16}$$

Thus, strong denoising can be achieved while simultaneously preserving image edges.

5.1.3.1 Location Dependent Weighting for Image Separation

Since we are dealing with an image separation scenario, the tradeoff term λ can be set separately for each image. Furthermore, in this scenario additional considerations need to be taken in determining the location dependent tradeoff, since the interference component in each image is itself derived from the second image. Thus the interference itself contains edges that we want to overcome in the total variation minimization process. We therefore want to recognize these interference edges (as opposed to true edges), and assign a high weight to the regularization term in these regions. In addition, since the backside image is slightly blurred, the back-side edges affect not only the front-side pixels at their exact location, but also pixels in their vicinity.

Based on these guidelines we propose the following weighting scheme that uses both front and back-side reconstructed images y_1 and y_2. Two weight maps λ^1 and

λ^2 are created. Conforming with (5.15) each map has two components one at vertical half-pixel shifts $(i+\frac{1}{2}, j)$ and one at horizontal half-pixel shifts $(i, j+\frac{1}{2})$.

Weighting Algorithm:

1. Create dilated absolute difference images:

$$y_{1x} = dilate(|y_1(i, j+1) - y_1(i, j)|)$$
$$y_{1y} = dilate(|y_1(i+1, j) - y_1(i, j)|)$$
$$y_{2x} = dilate(|y_2(i, j+1) - y_2(i, j)|)$$
$$y_{2y} = dilate(|y_2(i+1, j) - y_2(i, j)|)$$

   The gray-scale dilation uses a small flat structuring element such as a 3 by 3 square [19].

2. For each location (i, j):

   (a) If $y_{2x}(i, j) < c * y_{1x}(i, j)$:

$$\lambda^1_{i+\frac{1}{2},j} = \frac{1}{y_{1x}(i, j)/255 + \epsilon}$$

   where c is an estimated attenuation factor.

   (b) else (edge is a result of show-through):

$$\lambda^1_{i+\frac{1}{2},j} = y_{2x}(i, j)$$

3. Repeat previous step for derivatives in y direction to compute $\lambda^1_{i,j+\frac{1}{2}}$.

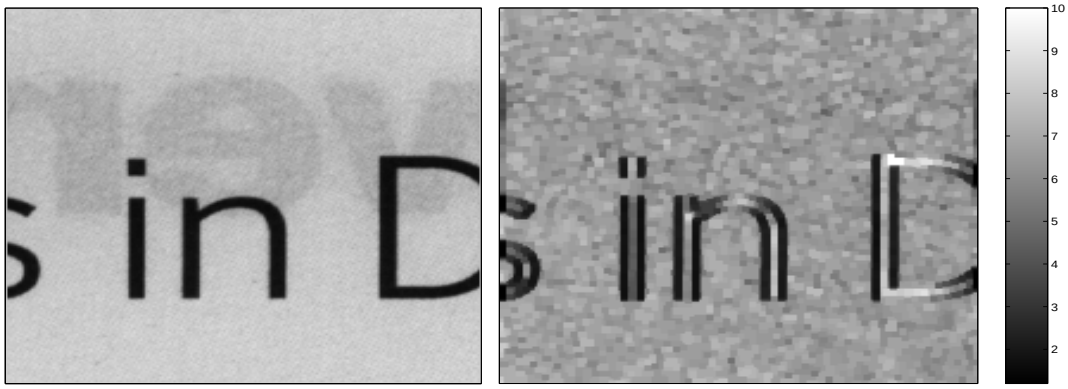
4. Repeat previous two steps for image y_2.

5. Clean weight maps by morphological closing and opening using a small flat structuring element (done separately for components in x and y directions).

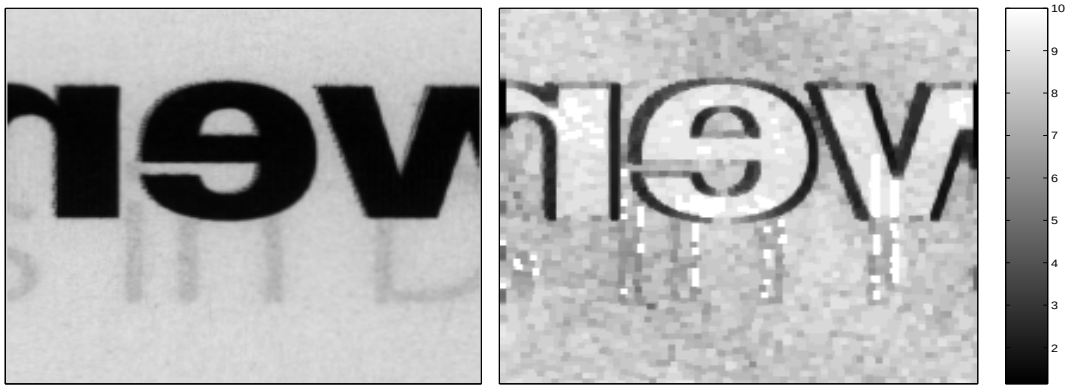
6. Truncate weight values to a reasonable range (1-10).

When dealing with linear mixtures the factor c can be set to be inversely proportional to a_{12} and a_{21} (once reasonable estimates of these elements are achieved). In the case of non-linear mixtures, determining this factor is more complex. We found that values in the range of 5-10 proved adequate in our examples.

Fig.5.2 depicts an example of the horizontal weighting maps (λ_x). Notice the low (dark) values along edges with vertical components, such as along the edges of the letters "i" and "n" in the front-side image (5.2(b)). Notice however, that the show-through edges do not have the same effect. Areas affected by show-through, as well as smooth areas, all have relatively higher (brighter) values, as can be seen in 5.2(d).



(a) Front-side image y_1 (b) λ_x^1 - front-side weighting map



(c) Back-side image y_2 (d) λ_x^2 - back-side weighting map

Figure 5.2: Weighting maps example.

In the following chapter we discuss optimization schemes for the solution of the image separation problem and propose a show-through removal algorithm.

Chapter 6

Solving the Show Through Problem by MAP Estimation

6.1 Optimization Framework

MAP estimation (5.3) in the joint solution space $\{\mathbf{Y}, \mathcal{H}\}$, discussed in 5, is a very difficult problem. The optimization problem is both non-linear and non-convex in the joint solution space, requiring specialized optimization tools. Therefore, we do not solve the problem directly. Instead, we approach the joint MAP estimation by an alternating minimization scheme [21, 41, 42]. This popular complexity reduction approach consists of iterative alternating estimation steps with respect to the different sets of variables.

The optimization process alternates between the Maximum-a-Priori problems:

$$\widehat{\mathbf{Y}} = \arg\max_{\mathbf{Y}} P(\mathbf{X}|\mathbf{Y}, \mathcal{H}) P(\mathbf{Y}) \tag{6.1}$$

$$\widehat{\mathcal{H}} = \arg\max_{\mathcal{H}} P(\mathbf{X}|\mathbf{Y}, \mathcal{H}) P(\mathcal{H}) \tag{6.2}$$

where $\mathbf{Y}$ are the estimated images, $\mathcal{H}$ the estimated mixing operator and $\mathbf{X}$ the measured images.

We alternate between the minimization of the image and the mixing operator such that the following free energies are minimized:

$$\widehat{\mathbf{Y}} = \arg\min_{\mathbf{Y}}\{\| \mathcal{H}(\mathbf{Y}) - \mathbf{X} \|_2^2 + \lambda Reg\{\mathbf{Y}\}\} \tag{6.3}$$

$$\widehat{\mathcal{H}} = \arg\min_{\mathcal{H}}\{\| \mathcal{H}(\mathbf{Y}) - \mathbf{X} \|_2^2\} \tag{6.4}$$

where $Reg\{\mathbf{Y}\}$ is the regularization function describing the image prior.

Equations (6.3) and (6.4) can also easily be expanded to incorporate additional terms such as priors on $\mathcal{H}$ and penalty terms.

While this reduction makes the optimization more feasible (the fidelity term at least is convex, with regard to $\mathcal{H}$ and $\mathbf{Y}$ separately), the optimization process is still burdensome. The $\mathbf{Y}$ optimization stage is especially problematic. The large number of interrelated variables (all the pixels in both images) poise a computational challenge. Furthermore, non-convex MRFs require the use of special optimization techniques.

6.2 Optimization Methods

6.2.1 Iterated Conditional Modes

As explained above, the many dependent variables (all the image pixels for both front and back sides) cause the image estimation stage to be cumbersome. Furthermore, image priors such as MRFs assume certain local image characteristics, while a global optimization process may induce undesirable large scale properties of the random field. To solve these problems Besag [7] offers the Iterated Conditional Modes (ICM) algorithm, an iterative method for image reconstruction that does not depend on

large scale characteristics.

In the ICM method, pixel locations are visited sequentially. At each step the pixel value is updated as to maximize the image probability (minimize the image cost) with all other pixel values held constant. When applied sequentially to all the pixels, this procedure defines a single cycle of the algorithm. Typically, convergence occurs after only a few such iterations to a local image probability maximum.

Applying this method to the image separation problem, each location consists of a pixel pair (front and back side pixels at that location) to be changed.

An advantage of this method is that during each pixel optimization we need not calculate the complete image probability (or cost function). Instead, due to the local nature of the pixel relations, we need to apply minimization only to the parts of the probability/energy functional that are affected by the current location.

In pointwise mixtures each pixel affects only the opposite side pixel at the same location. Thus, each optimization need only include the fidelity and regularization terms of that location. In convolutive mixtures, each pixel affects a group of pixels on the opposite side. In this case, each pixel optimization includes the fidelity terms for all locations affected.

A detailed derivation of the probability functionals, maximized by the ICM method for pointwise and convolutive image mixtures, is given in Appendix A.

6.2.2 Graduated Non-Convexity

Non-convex functions require the use of specially tailored optimization methods, such as Graduated Non-Convexity (GNC) [8, 28]. The GNC method is a deterministic annealing method for approximating the global solution for non-convex unconstrained, continuous minimization problems. This method finds good solutions at considerably less cost than stochastic Simulated Annealing which is commonly used for such problems.

In the GNC method, a parametric family of functions is created, the first of which is convex, and the last is the desired regularization function. Optimization starts by using the convex function and as the optimization progresses it is morphed into increasingly non-convex functions.

The family of functions has to be created specifically for the regularization function chosen (and is not necessarily unique). An example of such a family for the following non-convex MRF:

$$\phi(\Delta s) = \frac{|\Delta s|/T}{1 + |\Delta s|/T} \tag{6.5}$$

is given by [5]:

$$\phi^{(p)}(\Delta s) = \begin{cases} \frac{|\Delta s|/T}{1+|\Delta s|/T} & |\Delta s| \geq p \\ r(\Delta s)^2 + q & otherwise \end{cases} \quad ; \quad p = p_{max}, .., 0 \tag{6.6}$$

where r and q are appropriate constants, assuring a continuous and smooth function. Parameter p controls the convexity of the function. During the optimization process this parameter is gradually reduced until the function $\phi^{(p)}$ is equal to the original MRF at $p = 0$. The effect of different values of p is shown in Fig.6.1 for the above example. $p = 255$ gives the topmost dashed line, and $p = 100, 50, 25, 10$ the dashed lines approaching the MRF (solid line).

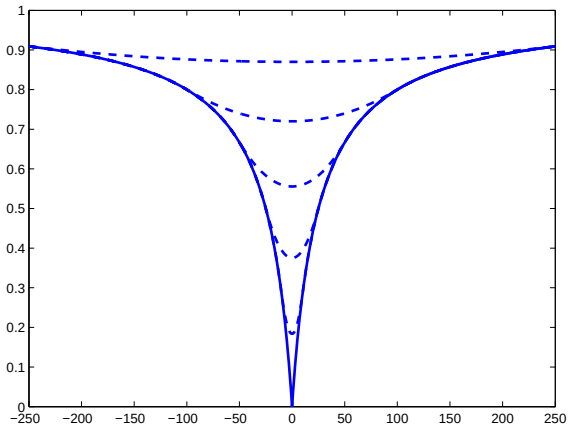


Figure 6.1: Function ϕ (solid line) and $\phi^{(p)}$ for various values of p (dashed lines).

6.3 Derivation of Proposed Algorithm

Several options for regularization functions were examined:

- Non-convex MRFs
- Convex MRFs
- Total Variation

The non-convex, edge preserving, MRFs were are our first choice as they provide the best available image model. Under the alternating minimization optimization process, each $\mathbf{Y}$ stage consisted of a series of GNC iterations, each of which consisted of ICM iterations. The computational requirements of such a process proved too high for all but very small images. Furthermore, the number of options and free parameters in such a process is very large: the specific MRF used (both shape and support), the choice of the parametric function family and the morphing strategy.

In trying to simplify the process we tested convex MRFs with small support. Using convex MRFs implies that the energy functional minimized in the $\mathbf{Y}$ stage is convex and can be minimized using standard gradient based techniques. The simplest MRF we tested was the absolute difference of adjacent pixel values (actually a smoothed approximation for numerical stability $|a| \approx \sqrt{a^2 + \epsilon}$). The L^1 norm is related to

sparsification, in this case the sparsification of image edges, which is a good measure of image separation.

While the results were pretty good, an obvious progression of this approach was to use Total-Variation regularization (actually the L^1 norm on pixel differences constitutes a crude implementation of Total-Variation).

The full algorithm developed includes an outer layer of alternating minimization (Eqs.(6.3) and (6.4)). Utilizing the cascaded approach (4.2.2) we start with a small filter support, increasing the filter support in later iterations. At the start of each $\mathbf{Y}$ iteration, the fidelity/regularization tradeoff maps (see 5.1.3.1) are recomputed with the minimization itself being implemented by the ICM method.

In order to avoid the directional preference that only one pass of the images generates, each of our ICM iterations consists of two passes, scanning once from top left to bottom right, and then in the opposite direction. In each scan the TV is computed based on the "causal" pixels. The ICM process typically converges in a few (3-5) iterations.

Each pixel (pair) optimization can be done using any standard multi-dimensional optimization technique. However, in order to reduce the computational cost, we are satisfied with finding new pixel values that reduce the cost but do not necessarily minimize it. A case can even be made that because of the complex (and non-convex) nature of the solution space and the interrelated variables, we should prefer that each pixel value optimization will not be too strong, so that the whole optimization process will not get stuck in a local minima.

We used a simple process that finds a "V" combination[1] of points in the direction of the gradient. Thus, instead of exact two-dimensional minimization, we perform a single (non-exact) one dimensional line-search. This optimization process is simple to implement at relatively low cost. While this process does not find a global minimum of the cost function (but only a lower point than the current one) we found it adequate.

The global background values (4.2.1) are used both in computing the fidelity term (the background is used in the non-linear transformation between reflectance and density/absorptance values). Local background values can be used for clipping pixel values updated in each optimization step, not allowing pixel values to be brighter than the background.

The $\mathcal{H}$ optimization step is much simpler due to the smaller number of variables. Given no prior knowledge on the mixing parameters this problem becomes a Least-Squares problem minimizing only the fidelity term (Eq.(6.4)). Using all pixel information (the whole two images) may hinder the solution (as in the decorrelation algorithms) as many non-informative pixels introduce noise to the process. As before we want to evaluate the PSFs by only using locations with single-side activity (as defined in 4). This is done by adding a mask to the fidelity term so only the relevant pixels contribute to the cost function.

Like all ICA methods our algorithm may suffer from two ambiguities:

1. Permutation

2. Scale

The Permutation ambiguity, while theoretically possible, is in practice not an issue. This is due to the sequential scan by the ICM method. The probability of a large

[1]Three points $a < b < c$ so that $f(a) > f(b) < f(c)$.

number of pixels, not to mention the whole image, simultaneously being "switched" is negligible. This has not happened even once in all our tests.

The Scale ambiguity is also not of major concern. While mixing parameters h_{11} and h_{22} do tend to drift from their expected value of 1, this drift is only noticeable after a large number of iterations and can be easily rectified by normalizing the images and mixing parameters. The density images are multiplied by the estimated value of h_{11} and h_{22}, which are then set to 1. Alternatively, we can fix the values of h_{11} and h_{22} to 1 during the optimization process, thus avoiding the need to normalize. However, we found that this hinders the optimization process somewhat.

The last issue that concerned us is that of the initialization. In view of the complex nature of the solution space, a good initial guess can greatly aid the optimization. Given no prior knowledge of the images, the input image mixtures X are used as an initial guess for the image estimates Y. However, a reasonable guess can be made regarding the mixing operator. Elements h_{11} and h_{22} are known to be 1 while the PSFs h_{12} and h_{21} are known to have a strong attenuation factor. Thus, we initialize these operators to small scalar values ϵ.

6.4 Proposed Algorithm

To summarize, the proposed algorithm is:

1. Estimate the global white paper reflectance, R_{pg}, using the mean-shift process on the whole image for both front and back sides (see 4.2.1.1).

2. Estimate the local white paper reflectance images, R_{pl}, for both front and back: local mean-shift process coupled with activity estimate (see 4.2.1.2).

3. Initial guesses $\mathbf{Y} = \mathbf{X}$ and $\mathcal{H} = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix}$, $\epsilon \sim 0.1$

4. Alternating minimization process:

   (a) $\mathbf{Y}$ optimization:

      i. Compute tradeoff maps $\lambda_{1,2}(m,n)$ (see 5.1.3.1)

      ii. ICM iterations until convergence - each iteration includes two scans: forward and reverse.

         A. For each pixel location minimize:

$$\{y_1(m,n), y_2(m,n)\} = \arg \min_{y_{1,2}(m,n)} \{\| \mathcal{H}(\mathbf{Y}) - \mathbf{X} \|_2^2 + \lambda TV\{\mathbf{Y}\}\}_{(m,n)} \tag{6.7}$$

            minimization is done by a line-search process and is done only on the elements of the cost function dependent on location (m,n).

         B. Clip values to $[0, R_{pwl}(m,n)]$.

   (b) $\mathcal{H}$ optimization:

      i. Compute activity masks

      ii. Least-Squares minimization

   (c) Normalize images and PSFs.

5. Repeat stage 4 with larger PSF supports.

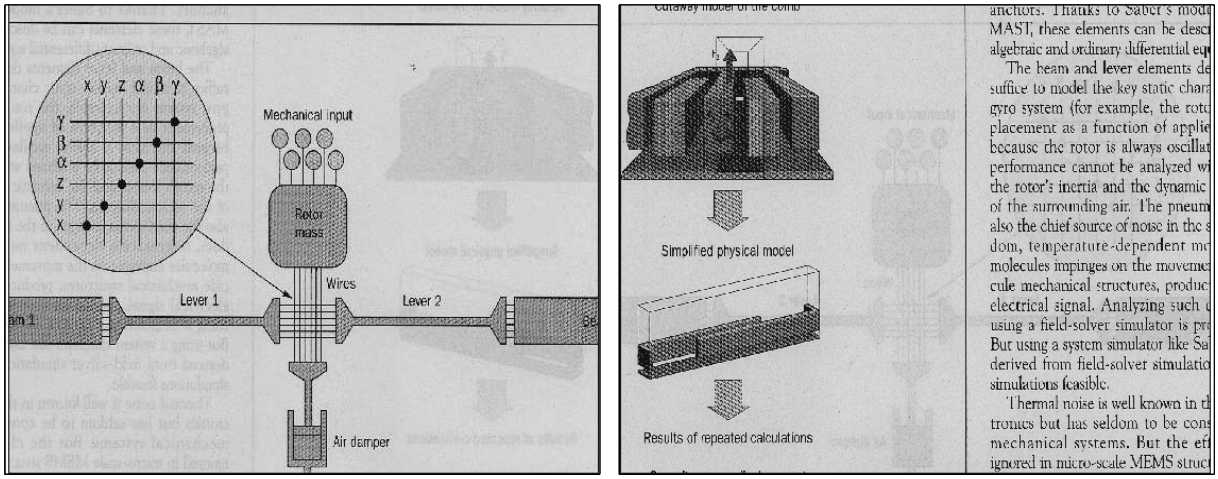
6.5 Simulation Results

The following example is for a pair of 256 gray level images of 420×560 pixels. For convenience, the images are shown in their correct orientation but obviously during registration and throughout the cleaning process the mirror image (flip left-right) of one of the images is used.

Fig.6.2 depicts the original scans. Notice that this example contains not only text (binary image) but also contains various gray graphical elements.

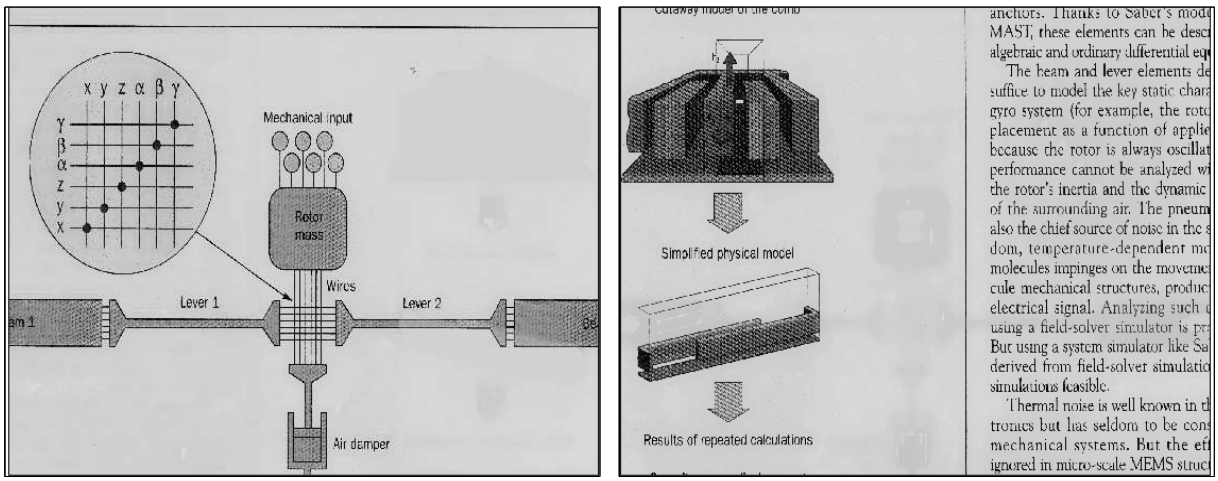
Fig.6.3 depicts the output of our BSS algorithm. The separated images are the result of only six alternating minimization stages. The first stage uses a 1×1 filter support, and all the following stages a 5×5 filter support.

While the separation (and subsequent Show-Through removal) is not perfect, we can see a big improvement. Additional minimization stages will further improve these results, however, we had to severely limit the number of iterations due to the high computational cost. Several improvements of our algorithm that could allow for better Show-Through removal are discussed in Section 7.2.



(a) Front (b) Back

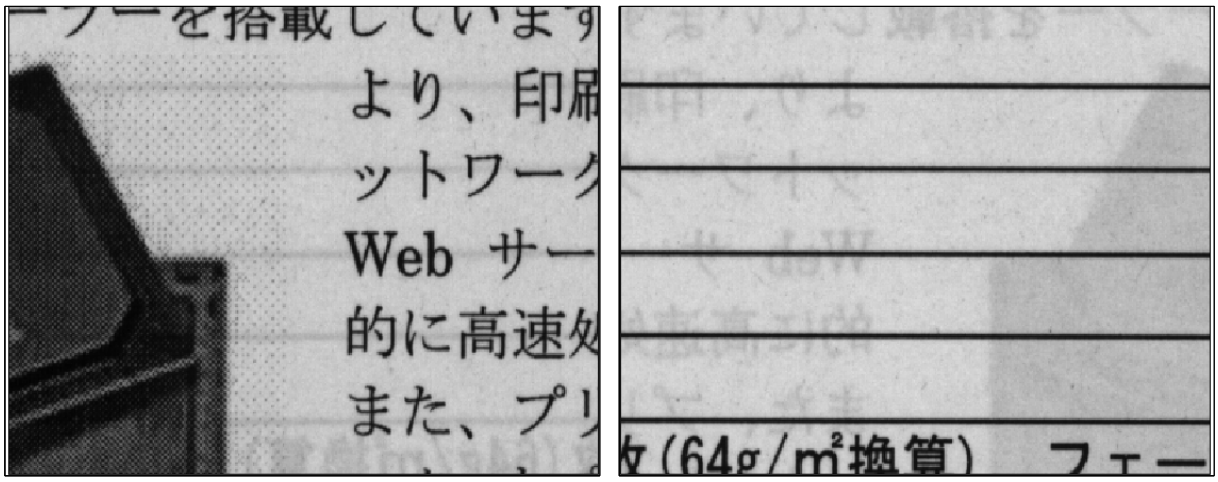
Figure 6.2: Example 1 : Original Images



(a) Front (b) Back

Figure 6.3: Example 1 : Separated Images

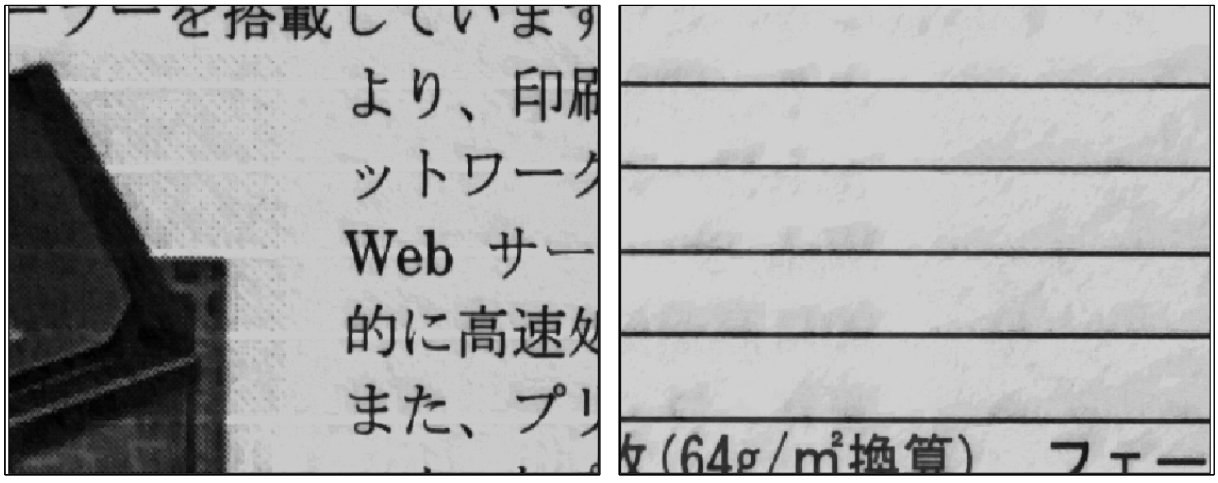
Fig.6.4 depicts another example for a pair of 420×560 pixel images. The cleaned images are shown in Fig.6.5.



(a) Front (b) Back

Figure 6.4: Example 2 : Original Images

Notice in the left part of the separated front-side image, in Fig.6.5, the smoothing of the textured area. This type of artifact is typical of Total-Variation based image processing techniques.



(a) Front (b) Back

Figure 6.5: Example 2 : Separated Images

In conclusion, in this chapter we presented a Blind Source Separation algorithm for Show-Through removal. The algorithm attempts to minimize a cost functional which incorporates the non-linear and convolutive image mixture model in the fidelity term, and Total-Variation image regularization. The fidelity/regularization tradeoff is set via a location dependent scheme, preserving only the desired image edges. Optimization is done by applying an alternating minimization scheme using the ICM approach to adjust pixel values.

Chapter 7

Conclusion

7.1 Summary

In this work we presented the Show-Through problem in scans of duplex printed documents, showed how it can be modelled as a non-linear convolutive mixture of images and developed two algorithms for solving the problem.

Our literature search revealed only one significant work, by Sharma [37], relating to this specific problem. The main contribution of Sharma's work is in modelling the Show-Through phenomenon and in partially linearizing it. However, the image processing tools utilized by Sharma to remove the Show-Through were fairly basic and the solution suffered from several flaws.

The first part of our work focused on analyzing Sharma's Decorrelation algorithm, understanding it's shortcomings, and offering improvements. Thus, our first algorithm is an improved Decorrelation based algorithm. We significantly improved on previous results by incorporating three main components:

- Local background estimation : Use of a global background value can cause over-whitening artifacts. We offered an algorithm for estimating local background

values using an adaptive mean-shift process.

- Cascaded filtering scheme : We propose using a cascaded adaptive filtering scheme, using filters with increasing support, to achieve better signal decorrelation.

- Adaptive post-processing stage : The cross-contamination of the signals causes the decorrelation process's output to be distorted. A post-processing stage is required to cancel the distortion. We offer a method to estimate the post-processing filter using adaptive filtering techniques.

The developed algorithm gives good results at relatively low computational costs. We found the algorithm to be robust, both in handling complex documents and in not being strongly affected by choice of parameters.

In the second part of our work we tackled the Show-Through problem from an ICA/BSS approach. The second algorithm we propose uses Blind Source Separation techniques to estimate both the clean images and the mixing parameters.

Image separation (resulting in Show-Through cancellation) is achieved by minimization of a cost functional. The non-linear convolutive image mixture model is incorporated into a mean-squared fidelity term. The independence assumption allows us to use an image regularization approach to the separation problem, applying the regularization separately to each image. We chose Total-Variation as our image prior due to it's convexity and edge preserving features.

Fidelity/Regularization tradeoff is set by a location dependent scheme that takes into account that the interfering signal is itself an image. Thus, only desired edges are preserved, while Show-Through induced edges are removed.

Minimization is achieved by an alternating minimization scheme, minimizing the cost functional by optimizing the image pixel values $\mathbf{Y}$ for given mixing parameters $\mathcal{H}$, and vice versa, for several iterations. This complexity reduction allows us to break the non-linear non-convex optimization problem into a set of simpler linear convex stages. The image minimization stage is done by the ICM method adapted for image separation. The algorithm gives good results but at a very high computational cost.

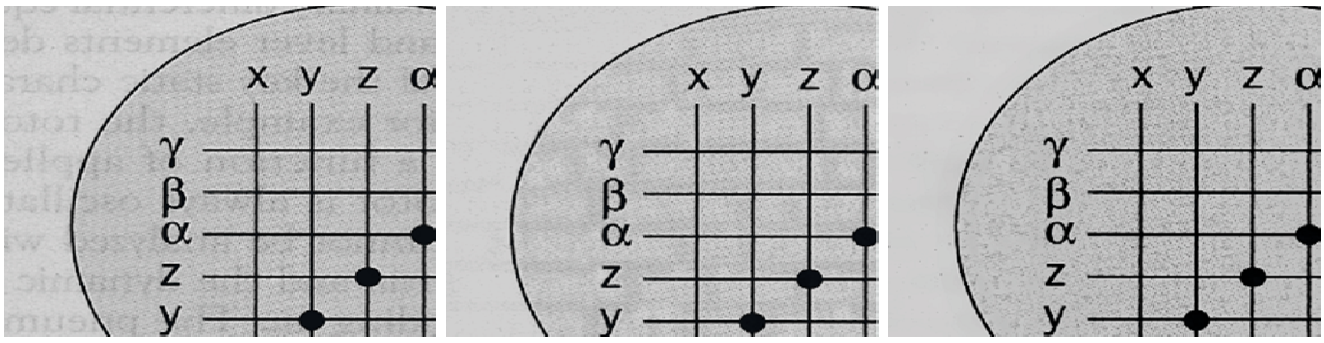
Several directions for improving the BSS algorithm are discussed in section 7.2.

7.1.1 Algorithm Comparison

Fig.7.1 shows a comparison of the two algorithm's output. It is evident that both algorithms significantly reduce the Show-Through although each approach suffers from different types of artifacts.

In the decorrelation based approach, areas of the image not affected by Show-Through remain virtually unchanged. Areas where Show-Through is removed may however suffer from over-cleaning effects, as shown in Fig.7.1(b), due to the algorithm's dependance on the background estimation.

In the BSS approach, the Show-Through is removed just as effectively, however artifacts typical to Total-Variation (or similar regularization functions) such as over-smoothing may occur. This is evident in Fig.7.1(c) where the texture of the gray background is smoothed. The BSS algorithm also comes at a much higher computational cost.



(a) Scanned (b) Decorrelation (c) BSS

Figure 7.1: Output of Decorrelation and BSS algorithms

7.2 Future Directions

While we feel that we have, more or less, exhausted the possibilities for improvement using the Decorrelation approach, much work can still be done under the BSS approach.

We divide future work into two main categories:

1. Technical improvements
2. Theoretical extensions

7.2.1 Technical Improvements

Under the heading of Technical Improvements we include various computational improvements to the numerical approximations used in our work. Specifically, both the Gradient and the Total-Variation estimations can be performed in various ways, usually with a clear tradeoff between computational complexity and the quality of the estimation. We expect better results with improved numerical approximations.

Another category of technical improvements goes to the optimization scheme used. Various other schemes can be applied to the problem. For example, Pyramidal and Multigrid methods have been used in many image processing problems, and may

be appropriate for the Show-Through problem as well. Variational methods may also hold promise for image separation, in general, and the Show-Through problem, specifically.

7.2.2 Theoretical Extensions

The technical improvements, listed above, attempt to minimize the same cost functional we used in our work, however better image models/priors can be proposed. A model that may be more suitable for our problem is the Structure-Texture model [2]. In this model an image is decomposed into structural and textural parts. Due to the attenuating and smoothing nature of the Show-Through process, most of the show through is probably due to the structural parts of the interfering image. Proper use of such a model may both lead to better Show-Through removal and lessen the smoothing artifacts inherent in the TV model.

Another problem, which exceeds the scope of this work, is that of Show-Through in color images. Even the modelling of the problem in color images is subject to further research. Are the different color channels affected in a similar way by Show-Through? How do they interact? Also, given a Show-Through model for color images, what tools are most appropriate for handling the problem?

Appendix A

Iterated Conditional Modes

The ICM method [7] aims to reconstruct degraded images taking into account the relations between adjacent pixels. The ICM method consists of an iterative process, where each iteration consists of sequential simple optimizations. Thus computational heavy large scale optimization is avoided as are undesirable large scale properties of the MRF.

We applied the ICM methodology to separation of image mixtures. The derivation of the ICM method to image separation is given below.

A.1 Single Image Reconstruction

Denoting $\mathbf{y}$ an estimate of scene $\mathbf{s}$, the algorithm aims to update point $\mathbf{y}(n)$ in light of all available information. $\mathbf{y}(n)$ is updated so as to maximize the conditional probability, given the scanned image $\mathbf{x}$ and the current estimate at all other points $\mathbf{y}(\Omega\backslash n)$:

$$\mathbf{y}(n) = \arg\max_{\widetilde{\mathbf{y}}(n)} P(\widetilde{\mathbf{y}}(n)|\mathbf{x}, \mathbf{y}(\Omega\backslash n)) \tag{A.1}$$

Assuming a locally dependent Markov random field model:

$$P(\mathbf{y}(n)|\mathbf{y}(\Omega\backslash n)) = P(\mathbf{y}(n)|\mathbf{y}(\partial_n)), \tag{A.2}$$

where ∂_n denote the neighbors of n (directly related to n through the MRF).

Assuming also that each $\mathbf{x}(n)$ has the same density function dependent only on $\mathbf{s}(n)$ (as in the case of additive i.i.d. noise), and that $\mathbf{x}(n)$ are conditionally independent for given $\mathbf{s}$, yields:

$$P(\mathbf{x}|\mathbf{s}) = \prod_n p(\mathbf{x}(n)|\mathbf{s}(n)) \tag{A.3}$$

It follows that :

$$P(\mathbf{y}(n)|\mathbf{x}, \mathbf{y}(\Omega\backslash n)) \propto p(\mathbf{x}(n)|\mathbf{y}(n))P(\mathbf{y}(n)|\mathbf{y}(\partial_n)) \tag{A.4}$$

When applied sequentially to each pixel in turn, this procedure defines a single cycle of an iterative algorithm for estimating $\mathbf{s}$.

The algorithm uses a "greedy" maximization strategy. During the process $P(\mathbf{y}|\mathbf{x})$ never decreases ($P(\mathbf{y}|\mathbf{x}) = P(\mathbf{y}(n)|\mathbf{x}, \mathbf{y}(\Omega\backslash n))P(\mathbf{y}(\Omega\backslash n)|\mathbf{x})$), so eventual convergence is assured. Typically convergence occurs after only a few iterations to a local maximum of $P(\mathbf{y}|\mathbf{x})$.

A more complicated scenario is that where $\mathbf{x}(n)$ depends not only on $\mathbf{s}(n)$ but also on points in it's vicinity v_n. This is the the case in deconvolution problems, where v_n is defined by the support of the blur kernel. The $\mathbf{x}(n)$'s are still considered conditionally independent given $\mathbf{s}$. In this scenario the conditional probability can be written as:

$$P(\mathbf{x}|\mathbf{s}) = \prod_n p(\mathbf{x}(n)|\mathbf{s}(n), \mathbf{s}(v_n)) \tag{A.5}$$

The probability to be maximized is given by:

$$\begin{aligned} P(\mathbf{y}(n)|\mathbf{x}, \mathbf{y}(\Omega\backslash n)) \propto\ & p(\mathbf{x}(n)|\mathbf{y}(n), \mathbf{y}(v_n)P(\mathbf{y}(n)|\mathbf{y}(\partial_n))\cdot \\ & \prod_{m\in\rho_n\backslash n} p(\mathbf{x}(m)|\mathbf{y}(n), \mathbf{y}(m), \mathbf{y}(v_m\backslash n)) \end{aligned} \tag{A.6}$$

where $\rho_n = \{m : n \in v_n\}$, i.e., all the points that point n falls within their vicinity.

A.2 Iterated Conditional Modes for Image Mixtures

We applied the ICM methodology to image mixtures. Denoting $\mathbf{y_i}$ an estimate of scene $\mathbf{s_i}$, the algorithm aims to update the point pair $(\mathbf{y_1}(n), \mathbf{y_2}(n))$ in light of all available information. $(\mathbf{y_1}(n), (\mathbf{y_2}(n))$ is updated so as to maximize the conditional probability, given the scanned images $\mathbf{x_i}$, the current estimate at all other points $\mathbf{y_i}(\Omega \backslash n)$:

$$(\mathbf{y_1}(n), \mathbf{y_2}(n)) = \underset{(\mathbf{y_1}(n), \mathbf{y_2}(n))}{\arg\max} P(\mathbf{y_1}(n), \mathbf{y_2}(n) | \mathbf{x_1}, \mathbf{x_2}, \mathbf{y_1}(\Omega \backslash n), \mathbf{y_2}(\Omega \backslash n)) \tag{A.7}$$

Assuming a locally dependant Markov random field model:

$$P(\mathbf{y_i}(n) | \mathbf{y_i}(\Omega \backslash n)) = P(\mathbf{y_i}(n) | \mathbf{y_i}(\partial_n)) \tag{A.8}$$

Where ∂_n denote the neighbors of n.

The sources $\mathbf{s_1}$ and $\mathbf{s_2}$ are assumed to be statistically independent so:

$$P(\mathbf{y_1}, \mathbf{y_2}) = P(\mathbf{y_1}) P(\mathbf{y_2}) \tag{A.9}$$

A.2.1 Pointwise Model

Assuming that each $\mathbf{x_i}(n)$ has the same density function dependent only on $\mathbf{s_{i,j}}(n)$ (instantaneous model), and that $\mathbf{x_i}(n)$ are conditionally independent for given $\mathbf{s}$, yields the following conditional probability:

$$P(\mathbf{x_1}, \mathbf{x_2} | \mathbf{s_1}, \mathbf{s_2}) = \prod_{n=1}^{N} p(\mathbf{x_1}(n), \mathbf{x_2}(n) | \mathbf{s_1}(n), \mathbf{s_2}(n)) \tag{A.10}$$

It follows that the probability to be maximized is given by:

$$\begin{aligned} P(\mathbf{y_1}(n), \mathbf{y_2}(n) | \mathbf{x_1}, \mathbf{x_2}, \mathbf{y_1}(S \backslash n), \mathbf{y_2}(S \backslash n)) \propto & p(\mathbf{x_1}(n), \mathbf{x_2}(n) | \mathbf{y_1}(n), \mathbf{y_2}(n)) \\ & P(\mathbf{y_1}(n) | \mathbf{y_1}(\partial_n)) P(\mathbf{y_2}(n) | \mathbf{y_2}(\partial_n)) \end{aligned} \tag{A.11}$$

A.2.2 Convolutive Model

In the convolutive mixture scenario $\mathbf{x_i}(n)$ depends not only on $\mathbf{s_j}(n)$ but also on it's vicinity v_n. The $\mathbf{x_i}(n)$'s are still considered conditionally independent given $\mathbf{s_{i,j}}$. The conditional probability is given by:

$$P(\mathbf{x_1}, \mathbf{x_2}|\mathbf{s_1}, \mathbf{s_2}) = \prod_{n=1}^{N} p(\mathbf{x_1}(n), \mathbf{x_2}(n)|\mathbf{s_1}(n), \mathbf{s_2}(n), \mathbf{s_1}(v_n), \mathbf{s_2}(v_n)) \tag{A.12}$$

The probability to be maximized is given by:

$$\begin{aligned} P(\mathbf{y_1}(n), \mathbf{y_2}(n)|\mathbf{x_1}, \mathbf{x_2}, &\mathbf{y_1}(S\backslash n), \mathbf{y_2}(S\backslash n)) \propto \\ & p(\mathbf{x_1}(n), \mathbf{x_2}(n)|\mathbf{y_1}(n), \mathbf{y_2}(n), \mathbf{y_1}(v_n), \mathbf{y_2}(v_n))\cdot \\ & P(\mathbf{y_1}(n)|\mathbf{y_1}(\partial_n))P(\mathbf{y_2}(n)|\mathbf{y_2}(\partial_n))\cdot \\ & \prod_{m\in\rho_n\backslash n} p(\mathbf{x_1}(m), \mathbf{x_2}(m)|\mathbf{y_1}(n), \mathbf{y_2}(n), \mathbf{y_1}(m), \mathbf{y_2}(m), \mathbf{y_1}(v_m\backslash n), \mathbf{y_2}(v_m\backslash n)) \end{aligned} \tag{A.13}$$

where $\rho_n = \{m : n \in v_m\}$, i.e., all the points that point i falls within their vicinity.

References

[1] A. Hyvärinen and J. Karhunen and E. Oja. *Independent component analysis*. John Wiley and Sons, NY, 2001.

[2] J-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection. *accepted to International Journal of Computer Vision (IJCV)*.

[3] L. Bar, N.A. Sochen, and N. Kiryati. Variational Pairing of Image Segmentation and Blind Restoration. In *ECCV*, volume 2, pages 166–177, May 2004.

[4] L. Bedini, G.M. Del Corso, and A. Tonazzini. Preconditioned Edge-Preserving Image Deblurring and Denoising. *Pattern Recognition Letters*, 22(10):1083–1101, Aug. 2001.

[5] L. Bedini, I. Gerace, and A. Tonazzini. A GNC algorithm for constrained image reconstruction with continuous-valued line processes. *Pattern Recognition Letters*, 15(9):907–918, Sept. 1994.

[6] A.J. Bell and T.J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[7] J. Besag. On the Statistical Analysis of Dirty Pictures. *Journal Royal Statistical Society (B)*, 48(3):259–302, 1986.

[8] A. Blake and A. Zisserman. *Visual Reconstruction*. MIT Press, Cambridge,MA, 1987.

[9] A. Bronstein, M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Separation of Reflections via Sparse ICA. In *Proc. IEEE International Conference in Image Processing*, volume 1, pages 313–316, Sept. 2003.

[10] A. Bronstein, M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Blind Deconvolution of Images Using Optimal Sparse Representations. *IEEE Transactions on Image Processing*, 14:726–736, 6 2005.

[11] T.F. Chan and C. Wong. Total Variation Blind Deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375, March 1998.

[12] D. Comaniciu and P. Meer. Robust Analysis of Feature Spaces: Color Image Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 750–755, June 1997.

[13] D. Comaniciu, V. Ramesh, and P. Meer. The Variable Bandwidth Mean Shift and Data-Driven Scale Selection. In *IEEE International Conference on Computer Vision*, volume 1, pages 438–445, July 2001.

[14] S. Cruces and L. Castedo. Stability Analysis of Adaptive Algorithms for Blind Source Separation of Convolutive Mixtures. *Signal Processing*, 78(3):265–275, 1999.

[15] T. Darrell and E. Simoncelli. Separation of Transparent Motion into Layers using Velocity-Tuned Mechanisms. In *Proc. European Conf. Computer Vision*, pages 306–313, May 1994.

[16] H. Farid and E.H. Adelson. Separating Reflections from Images by use of Independent Component Analysis. *J. Optical Soc. America A*, 16(9):2136–2145, Sept. 1999.

[17] I. Gerace, F. Cricco, and A. Tonazzini. An Extended Maximum Likelihood Approach for the Robust Separation of Autocorrelated Images from Noisy Mixtures. In *ICA 2004*, pages 954–961, Sept. 2004.

[18] I. Gerace, R. Pandolfi, and P. Pucci. A New Estimation of Blur in the Blind Restoration Problem. In *International Conference on Image Processing*, volume 1, pages 261–264, Sept. 2003.

[19] C.R. Giardina and E.R. Dougherty. *Morphological Methods in Image and Signal Processing*. Prentice Hall, NJ, first edition, 1988.

[20] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, NJ, second edition, 1991.

[21] R. Kaftory, N. Sochen, and Y.Y. Zeevi. Variational Blind Deconvolution of Multi-Channel Images. *International J. Imaging Science and Technology*, 15(1):56–63, 2005.

[22] K.T. Knox. Show-through correction for two-sided documents. U.S. Patent 5,646,744, July 1997.

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

[24] A. Levin and Y. Weiss. User Assisted Separation of Reflections from a Single Image Using a Sparsity Prior. In *Proc. European Conf. Computer Vision*, pages 602–613, May 2004.

[25] A. Levin, A. Zomet, and Y. Weiss. Separating Reflections from a Single Image Using Local Features. In *Proc. IEEE Computer Vision and Pattern Recognition*, volume 1, pages 306–313, June 2004.

[26] S.Z. Li. *Markov Random Field Modelling in Computer Vision*. Springer-Verlag, Tokyo, 1995.

[27] J. Míguez and L. Castedo. Maximum Likelihood Unsupervised Source Separation in Gaussian Noise. *Journal of VLSI Signal Processing*, 31(1):7–18, May 2002.

[28] M. Nikolova. Markovian Reconstruction Using a GNC Approach. *IEEE Transactions on Image Processing*, 8(9):1204–1220, Sept. 1999.

[29] L. Parra and C. Spence. Convolutive Blind Separation of Non-Stationary Sources. *IEEE Transactions on Speech and Audio Processing*, 8:320–327, May 2000.

[30] L. Parra, C. Spence, and B. de Vries. Convolutive Source Separation and Signal Modelling with ML. In *International Symposium on Intelligent Systems*, Sept. 1997.

[31] D.T. Pham and J. Cardoso. Blind Separation of Instantaneous Mixtures of Non Stationary Sources. *IEEE Trans. on Signal Processing*, 49(9):1837–1848, Sept. 1997.

[32] D.T. Pham and P. Garat. Blind Separation of Mixture of Independent Sources Through a Quasi-Maximum Likelihood Approach. *IEEE Trans. on Signal Processing*, 45(7):1712–1725, July 1997.

[33] F.R. Ruckdeschel and O.G. Hauser. Yule-Nielsen effect in printing: A physical analysis. *Appl. Opt.*, 17:3376–3383, Nov. 1978.

[34] L. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259–268, 1992.

[35] B. Sarel and M. Irani. Separating Transparent Layers through Layer Information Exchange. In *Proc. European Conf. Computer Vision*, pages 328–341, May 2004.

[36] Y.Y. Schechner, N. Kiryati, and R. Basri. Separation of Transparent Layers Using Focus. *International Journal of Computer Vision*, 39(1):25–39, Aug. 2000.

[37] G. Sharma. Show-Through Cancellation in Scans of Duplex Printed Documents. *IEEE Transactions on Image Processing*, 10(5):736–754, May 2001.

[38] D.M. Strong, P. Blomgren, and T.F. Chan. Spatially Adaptive Local Feature-Driven Total Variation Minimizing Image Restoration. In *Proceedings of SPIE., Statistical and Stochastic Methods in Image Processing II*, volume 3167, pages 222–233, Oct. 1997.

[39] H.N. Thi and C. Jutten. Blind Source Separation for Convolutive Mixtures. *Signal Processing*, 45(3):209–229, 1995.

[40] A. Tonazzini, L. Bedini, E.E. Kuruoglu, and E. Salerno. Blind Separation of Time-Correlated Sources from Noisy Data. Technical Report TR-42-2001, IEI-CNR, 2001.

[41] A. Tonazzini, L. Bedini, E.E. Kuruoglu, and E. Salerno. Blind Separation of Auto-Correlated Images from Noisy Mixtures Using MRF Models. In *ICA 2003*, pages 675–680, April 2003.

[42] A. Tonazzini and I. Gerace. Bayesian MRF-Based Blind Source Separation of Convolutive Mixtures of Images. In *EUSIPCO*, Sept. 2005.

[43] K. Torkkola. Blind Separation of Convolved Sources Based on Information Maximization. In *IEEE Workshop on Neural Networks for Signal Processing*, pages 423–432, Sept. 1996.

[44] Y. Tsin, S.B. Kang, and R. Szeliski. Stereo Matching with Reflections and Translucency. In *Proc. IEEE Computer Vision and Pattern Recognition*, volume 1, pages 702–709, June 2003.

[45] C.R. Vogel and M.E. Oman. Iterative Methods for Total Variation Denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, Jan. 1996.

[46] E. Weinstein, M. Feder, and A.V. Oppenheim. Multi-Channel Signal Separation by De-Correlation. *IEEE Transactions on Speech and Audio Processing*, 1(4):405–413, Oct. 1993.

[47] B. Widrow et al. Adaptive noise cancelling: principles and applications. *Proc. IEEE*, 63:1692–1716, Dec. 1975.

[48] D. Yellin and E. Weinstein. Criteria for Multichannel Separation. *IEEE Transactions on Signal Processing*, 42(8):2158–2168, Aug. 1994.

[49] D. Yellin and E. Weinstein. Multichannel Signal Separation: Methods and Analysis. *IEEE Transactions on Signal Processing*, 44(1):106–118, Jan. 1996.

[50] Y. You and M. Kaveh. A Regularization Approach to Joint Blur Identification and Image Restoration. *IEEE Transactions on Image Processing*, 5(3):416–428, Mar. 1996.

ביטול השתקפות צד-אחורי בתמונות סרוקות

בעז אופיר

ביטול השתקפות צד-אחורי בתמונות סרוקות

חיבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת תואר

מגיסטר למדעים

בהנדסת חשמל

בעז אופיר

הוגש לסנט הטכניון — מכון טכנולוגי לישראל

חשון תשס"ז חיפה אוקטובר 2006

חיבור על מחקר נעשה בהדרכת פרופ' דוד מלאך

בפקולטה להנדסת חשמל

הכרת תודה

ברצוני להודות לפרופ' דוד מלאך על רעיונותיו הפוריים ועל הזמן היקר שהקדיש לי.

אני רוצה להודות גם לדר' אהוד קרנין ולדר' יוג'ין וואלך ממעבדות י.ב.מ. בחיפה על שהציגו בפנינו את הבעיה וסיפקו את הנתונים ששימשו אותנו בניסויים.

לבסוף, אני מודה למשפחתי על סבלנותם ותמיכתם. עבודה זו מוקדשת להם.

אני מודה לטכניון על התמיכה הכספית הנדיבה בהשתלמותי

תוכן עניינים

רשימת איורים

תקציר

בעבודה זו אנו מציגים את בעיית ההשתקפות של הצד-האחורי (Show-Through) במסמכים סר-וקים. תופעה זו מתרחשת כאשר סורקים מסמך המודפס משני צידי הדף. הדפוס בצידו האחורי של הדף משתקף דרך הדף ומקלקל את תמונת הצד הקדמי. התופעה קורת במהופך כאשר סורקים את צידו השני של הדף. הדבר אינו מטריד במיוחד כאשר מדובר בסריקות באיכות נמוכה (כמו בסריקת מסמכים טיפוסית בסורק שולחני) שם נושא האיכות אינו מהותי. אולם, כאשר איכות הסריקה חשובה, פתרון בעיית ההשתקפות נהיה חיוני. זהו המקרה בתעשיית הדפוס הדיגיטאלי בעת יצירת עותק master.

גישה נאיבית לפתרון הבעיה, ע"י הפעלת רמת סף, פותרת את הבעיה עבור מסמכים "פשו-טים". כאלה הם למשל מסמכים המכילים טקסט בלבד (למעשה, תמונות בינאריות). גישה כזאת תכשל בטיפול במסמכים מורכבים יותר, המכילים אלמנטים גראפיים בדרגות רמות אפור שונות. כך גם תכשל כל גישה המתייחסת לתמונה בודדת בלבד, שאינה מנצלת סריקות של שני צידי הדף, במשותף.

עבודות מעטות בלבד התייחסו לבעיה זו. העבודה המשמעותית היחידה שמצאנו מתמקדת בניתוח תהליך היצירה של תופעת השתקפות הצד האחורי ע"י מעקב אחר מסלול מעבר האור דרך מנגנוני הסורק, בעת סריקת המסמך. האור יוצא ממנורת הסורק, עובר דרך שכבת הדיו בצדו הקדמי של הדף ומוחזר מהדף, חזרה דרך הדיו, אל הגלאי, ליצירת תמונת הצד הקדמי. בעיית ההשתקפות נוצרת כיוון שרכיב מסוים של האור אינו מוחזר ישירות מהדף, אלא עובר דרך הדף, דרך שכבת הדיו בצידו האחורי של הדף, ומוחזר ממכסה הסורק (דרך שתי שכבות הדף, והדף עצמו) אל הגלאי. כך נוצר ערבוב לא ליניארי של תמונות הצד הקדמי והצד האחורי. כל מודל המתאר את התופעה צריך לכלול בין היתר פונקציית טשטוש המייצגת את פיזור האור בעברו דרך הדף, וכך

מטשטשת את השתקפות הצד האחורי. ניתן לבצע ליניאריזציה חלקית של המודל כך שהתופעה תתואר כערבוב ליניארי ומטשטש של התמרות לא ליניאריות של התמונות הרצויות.

נקודת המוצא שלנו, לפתרון הבעיה, היא שיש בידנו שתי תמונות רמות אפור שהתקבלו מסריקת שני צידי הדף. אנו מניחים כי אין בידנו כל מידע מוקדם לגבי תוכן המסמך הסרוק, או לגבי פרמטרי תהליך הערבוב הנובעים מתכונות פיסיקליות של הדף, הדיו והסורק. הנחת המוצא היחידה אותה אנו מניחים היא שהתמונות עברו התאמה מרחבית (רגיסטרציה) כתהליך מקדים.

מטרתנו ליצור שתי תמונות (צד קדמי וצד אחורי), מנוקות מהשתקפויות, הדומות ככל האפשר לתמונות אותן היינו מקבלים מסריקת אותו מסמך לו היה מודפס מצד אחד של הדף, בלבד.

בעבודה זו אנו מציגים שני אלגוריתמים לפתרון בעיית השתקפות הצד האחורי.

האלגוריתם הראשון מתבסס על קריטריון הדה-קורלציה. בעבודה קודמת, של Sharma, נעשה שימוש בקריטריון זה אך האלגוריתם שהוצע היה בסיסי מאד וסבל ממספר פגמים וחולשות. אנו מנתחים את עבודתו של Sharma, ומציעים מספר שיפורים ותוספות. באלגוריתם שלנו, דה-קורלציה מושגת באמצעות תהליך רב שלבי (קסקדה) של מסננים מסתגלים (אדפטיביים). בכל שלב של הקסקדה משוערכים מסננים (אחד לכל תמונה) בעלי תמך גדול יותר מהשלב הקודם, כאשר השערוך מתבצע בשיטת LMS. בנוסף, האלגוריתם המוצע מסתגל לשינויי בהירות בתמונה ע"י שיערוך רמת האפור של הרקע באופן מקומי. הדבר נעשה ע"י תהליך Mean-Shift. אנו מבטלים את העיוות, שנוצר עקב כך שהאות המרעיש (אות הייחוס - שהוא תמונת הצד השני) מכיל בעצמו אלמנטים מהאות הרצוי (עקב השתקפות הצד הקדמי בצד האחורי), ע"י תהליך סינון נוסף ע"י מסנן מסתגל.

התוצאה היא אלגוריתם עמיד (רובוסטי) ויעיל חישובית המטפל בהצלחה בסריקות של מסמכים מורכבים המכילים אלמנטים גראפיים מורכבים, זאת ללא ארטיפקטים בולטים שמאפיינים את האלגוריתם של Sharma.

האלגוריתם השני משתמש בשיטות של הפרדה עיוורת של מקורות (Blind Source Separation). האלגוריתם משערך בו זמנית את התמונות ואת פרמטרי הערבוב. הפרדת התמונות (ועקב כך ביטול

ההשתקפות) מושגת באמצעות תהליך מינימיזציה של פונקצית מחיר. פונקציה זו מורכבת משני רכיבים עיקריים. הרכיב הראשון הוא רכיב של שגיאה ריבועית ממוצעת ומתאר את ההתאמה (fidelity) של השערוך לתמונות הסרוקות. הרכיב השני מייצג מודל א-פריורי של התמונות ודואג להסדרה (regularization) של הפתרון. מודל הרגולריזציה שנבחר על ידינו הוא Total-Variation. מודל זה נבחר בשל תכונת שימור השפות כאשר בו בזמן פונקציה זו היא קעורה, מה שמקל על תהליך האופטימיזציה. הנחת האי-תלות בין התמונות מאפשרת לנו לבצע הסדרה בנפרד על כל תמונה. האלגוריתם קובע את יחס הגומלין בין שני הרכיבים באופן תלוי מיקום. במקומות בהם קיימת שפה (edge) הנובעת בסבירות גבוהה מההשתקפות, ניתן משקל גבוה להסדרה. לעומת זאת, במקומות בהם קיימת שפה "אמיתית", ניתן משקל נמוך להסדרה. כך האלגוריתם משמר את השפות הרצויות בעודו מסיר שפות הנובעות מהשתקפות הצד האחורי.

מורכבות הבעיה לא מאפשרת ביצוע אופטימיזציה ישירה על פונקצית המחיר. אי לכך אנו מבצעים פישוט של הבעיה ומבצעים את המינימיזציה באופן איטרטיבי, כאשר המינימיזציה של פונקצית המחיר נעשית לסירוגין, פעם על ערכי רמות האפור של התמונות, כאשר ערכי פרמטרי הערבוב מוחזקים קבועים, ופעם על פרמטרי הערבוב, כאשר ערכי התמונות מקובעות. כך מתקבל כי כל שלב בתהליך הוא ליניארי וקעור. צעד האופטימיזציה בערכי התמונות נעשה בעזרת אד-פטציה של שיטת ICM (Iterated Conditional Modes) לבעיית הפרדת תמונות. בשיטה זו מפורק תהליך האופטימיזציה לסדרה של צעדים, כשבכל צעד מתעדכנים רק ערכי הפיקסלים בנקודה אחת. התהליך כולל מספר מעברים כאלה על התמונות עד להתכנסות. בשיטה זו נמנע הצורך באופטימיזציה בקנה מידה גדול.

האלגוריתם משיג תוצאות טובות אך במחיר חישובי גבוה מאד. עם זאת, אנו מאמינים בפוט-נציאל של גישה זו לפתרון הבעיה ומציעים מספר שיפורים והרחבות אפשריים לאלגוריתם זה.