ON USING AND IMPROVING GRADIENT DOMAIN PROCESSING FOR IMAGE ENHANCEMENT

DENG FANBO

NATIONAL UNIVERSITY OF SINGAPORE

2013

ON USING AND IMPROVING GRADIENT DOMAIN PROCESSING FOR IMAGE ENHANCEMENT

DENG FANBO

(*B.E., Harbin Institute of Technology, 2008*)

A DISSERTATION SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

NATIONAL UNIVERSITY OF SINGAPORE

2013

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Signature: _____   **Date:** _____

To my parents.

Acknowledgements

First and foremost I would like to express my sincerest gratitude to my advisor Prof. Michael S. Brown for his consistent guidance and support throughout the past five years, for his brilliant inspiration to my research problems, for his thoughtful encouragement when I met with difficulties, for his great patience when helping with writing and polishing all my papers also including this dissertation, and much more. I could not imagine a better or friendlier advisor and mentor for my Ph.D study.

I am heavily thankful to my collaborators Dr. Wu Zheng, Dr. Seon Joo Kim, Dr. Tai Yu-Wing, and Dr. Dilip Prasad for their great contribution to my research works. Without their valuable advice and enthusiastic guidance, my research works could not have been completed. Sincere thanks also go to my co-authors Dr. Lu Zheng, Dr. Zhuo Shaojie, Dr. Fu Chi-Wing, and Dr. Moshe Ben-Ezra for their comments and suggestions on the writing of papers. Extra thanks to Dr. Dilip Prasad for his careful proofreading and further polishing of this dissertation.

I also want to thank the committee members of my dissertation, Prof. Mohan S. Kankanhalli and Prof. Ping Tan for their patience in reading and helpful, insightful comments on this dissertation.

I thank all my labmates in NUS Computer Vision Group: Cheng Yuan, Gao Junhong, Lin Haiting, and Liu Shuaicheng, for the inspirational discussions, for those sleepless nights we were fighting for upcoming deadlines, and for the wonderful five years we have spent together. Also I thank all the friends I got to know in Singapore. Our valuable friendship has made my life here very graceful and colorful.

Last but not the least, I would like to express my deepest gratitude to my parents, for their kindly understanding, unreserved support, and

unselfish love. I would like to thank my girlfriend, Chen Qi, who has always loved, encouraged, and supported me through all the ups and downs in my life.

Contents

Summary

Gradient domain image processing is a type of image manipulation that directly processes the derivatives of an image (i.e. gradient) instead of its pixel values. This involves a two step procedure where the image gradients are first processed in a task-specific manner based on the desired enhancement, followed by a reconstruction step that estimates the new pixels values from the modified gradient field. Since its adoption nearly a decade ago, there has been several successful examples of using gradient domain processing for image enhancement tasks ranging from texture transfer, gradient boosting to saliency sharpening and data fusion. This dissertation continues this trend of gradient domain image enhancement and offers three contributions in this area.

Our first contribution is focused on enhancing images of old and damaged documents. Specifically, we show how gradient domain processing can be used to effectively combine information from visible and non-visible spectral bands to significantly improve the visual quality of old documents suffering from age-related effects such as ink-bleed, corrosion, and decay.

Our second contribution proposes a new method to reduce noticeable compression artifacts that arises from tone-adjustment. Tone-adjustment is a fundamental image editing operation that can significantly enhance image quality but can also boost undesirable compression artifacts that are otherwise not noticeable in the original image. In particular, we propose a novel method to detect and correct compression errors in the gradient domain. We show that this gradient domain strategy that can produce more compelling results than those obtained with existing methods.

Our third contribution targets the reconstruction step of gradient domain processing. In particular, we present a color-aware regularization method that can avoid color shift artifacts that often occur in existing gradient domain reconstruction methods. Key to this work is a novel regularization technique which uses an anisotropic Mahalanobis distance for restricting to the image's color distribution while applying gradient domain processing. The effectiveness of this regularization method is illustrated using three common image enhancement approaches including gradient transfer, gradient boosting and saliency sharpening.

These collective contributions help to advance the state-of-the-art in image enhancement techniques within a gradient domain context.

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Not long after the invention of photography over one hundred years ago, pho-
tographers started to use photo editing (or photo retouching) techniques to alter
printed photographs or undeveloped negatives with the purpose of enhancing
their visual appearance. Early photo editing techniques were quite limited and
were only possible within a small crowd of professionals who were equipped with
specific domain knowledge and film processing hardware and chemicals. This sit-
uation has changed significantly since personal computer and digital photography
technologies were invented. Digital photography is now popular and inexpensive
and digital photo is accessible to most people.

From a computer scientist's point of view, photo editing is only a part of our
interest. In a more general sense, we concentrate on developing image processing
techniques that enhance or extend the capabilities of digital photography. Over the
years, image filters have been widely used by researchers of the computer vision
community in many image processing pipelines such as image sharpening [72],

image denoising [39], pseudo-relighting [80], and so on. Image filters developed in the last few decades usually directly manipulate pixel values in the *spatial domain* (a.k.a. *color domain* or *0th-order domain*), or modify frequencies in the *frequency domain* [74, 62, 87]. However, a particular form of spatial domain filtering, which is referred to as *gradient domain* (a.k.a. *1st-order domain*) filtering or gradient domain image processing, has recently become the new cornerstone of numerous image processing algorithms [24, 59, 45, 54, 9].

Attneave's [4] and Barten's [7] studies on human visual system show that our visual system perceives local contrast (correlated to image gradients) instead of absolute pixel intensities. Motivated by this observation, more and more gradient domain image processing methods have been developed. These methods manipulate pixel differences (e.g. 1st-order image gradients) in addition to pixel intensities in order to better resemble the way how humans perceive images and achieve some enhancement effects that are difficult to be done in spatial domain, such as reflection removal [1], shadow removal [25], drag-and-drop pasting [31], etc.

The major difference between the traditional image processing pipeline and gradient domain image processing pipeline is illustrated in Figure 1.1. Assume we need to enhance the contrast level of an input image. Using traditional image processing technique, input pixel values are directly modified by applying a specific tone curve to the input image. However, gradient domain processing introduces changes to the 1st-order image gradients – scaling (i.e. boosting) the input gradient field to enhance the contrast. After the gradient has been modified, a reconstruction step is applied that estimates the pixel values from the modified gradient field. As such, gradient domain image processing usually involves a two step procedure: 1) the image gradients are first manipulated/modified in a task-specific manner
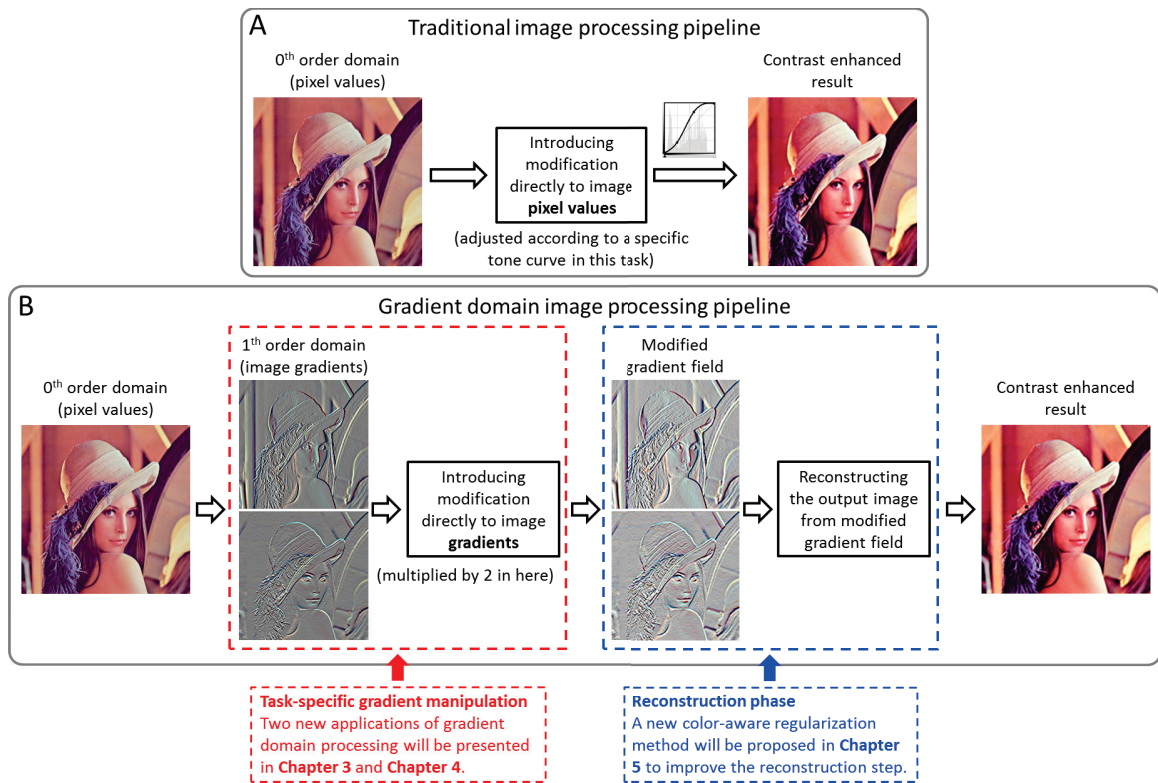
Figure 1.1: An example using a contrast enhancement task to illustrate the difference between the traditional image processing pipeline, shown in (A), and gradient domain image processing pipeline shown in (B).

to obtain the desired gradient field, and 2) a reconstruction step is carried out to estimate the new pixel values from the modified gradient field.

Within the past decade, gradient domain processing has been successfully applied for image enhancement tasks including texture transfer, gradient boosting and saliency sharpening. This dissertation continues the trend of gradient domain image enhancement and explores three unsolved problems in this area: visual enhancement of old documents, compression artifact reduction and color-aware regularization. The first two problems are related to the first step (task-specific gradient manipulation) of the gradient domain processing procedure, while the last

(a)   (b)   (c)

Figure 1.2: An example of the visual enhancement of old documents: (a) original RGB image; (b) 850nm NIR band image; (c) enhanced result.

problem can be considered as an issue of the reconstruction step. These problems are briefly introduced in the following section.

1.2   Problems to Be Solved

Visual enhancement of old documents   Archives and other related institutions such as libraries and museums serve as the custodial record keepers of our collective memories. One important role of these institutions is the management and preservation of historically significant documents. These very old documents often suffer from various kinds of deterioration including paper yellowing effect, ink bleed and corrosion, biological and physical damage, etc. For example, Figure 1.2(a) shows a cropped region of a line drawing that suffers from the low contrast issue caused by ink corrosion.

In recent years, some libraries and archives (e.g. our collaborator, the Nationaal Archief of the Netherlands) start to use the hyperspectral imaging (HSI) technique to image their collected documents and drawings. HSI can capture a densely sampled spectral response of a document over a broad spectrum including invisible

spectra such as ultra-violet (UV) and near-infrared (NIR). An advantage of HSI is that the invisible spectral bands (especially NIR bands) usually provide less artifacts or more details for the document being imaged. Figure 1.2(b) shows the 850nm NIR band image of the same region as Figure 1.2(a), which exhibits more details on the sailing boats.

Our goal is to design a visual enhancement framework for degraded historical documents based on the gradient domain fusion of normal RGB images and hyperspectral images. We mainly focus on how to fuse gradients of the invisible spectra, most notably NIR bands, with the normal RGB image to visually enhance the appearance of historical documents. As shown in Figure 1.2(c), after transferring the gradients of the NIR band into the RGB image, we can greatly improve the legibility of this line drawing. Using a similar idea, we also demonstrate how to improve the visual quality of text-based document corrupted with undesired artifacts such as ink-bleed, ink-corrosion, and foxing. Chapter 3 provides more detailed discussion and experimental results of this work.

Compression artifact reduction The JPEG compression standard is a commonly used *lossy* compression format for digital images. The degree of compression can be adjusted as a trade-off between storage size and image quality. When the degree of compression goes higher, some distinct artifacts start to appear in the compressed image, including blocking artifact (generally in homogeneous regions like the sky or walls), color distortion, staircase noise along curving edges, and ringing artifact. In the last few decades, the computer vision community has made some excellent progress in analyzing and reducing compression artifacts for heavily compressed images. Some representative works can be found in [84, 81, 55, 29, 37, 3, 2, 47, 69, 26, 86, 23], however, we try to deal with this problem from a different point of view.
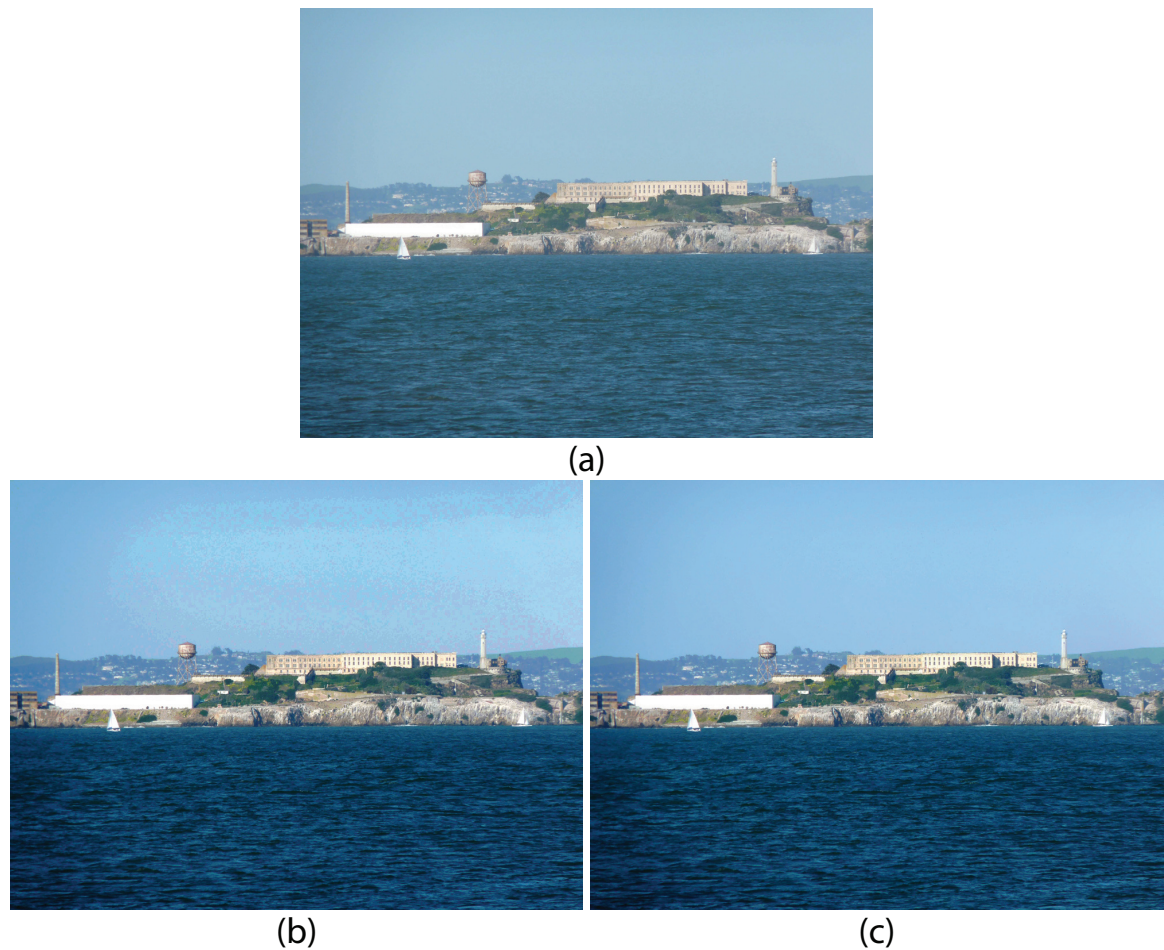
Figure 1.3: An example of compression artifact removal: (a) original JPEG image with good quality; (b) histogram equalized result of (a), suffering from blocking and color distortion artifacts; (c) restored result of (b).

In fact, with the popularization of modern mass storage devices, we usually do not compress JPEG images a lot in order to keep good image quality, but this does not mean compression artifacts have disappeared, actually they just have been well-hidden. Figure 1.3(a) shows a JPEG image with fairly good quality (above medium) that has no visible artifact. However, when we modify this image

using a histogram equalization (or other applicable tone-adjustment operations) to enhance its contrast level, blocking and color distortion artifacts immediately become apparent as shown in Figure 1.3(b). This means that tone-adjustment not only enhances the image contrast, but also boosts undesirable compression artifacts that are otherwise not noticeable in the original image. This is because schemes such as JPEG optimize their compression in a scene dependent manner. As such, low-contrast images exhibit few perceivable artifacts even for relatively high-compression factors. After tone-adjustment, however, subtle artifacts are magnified. While there exists numerous approaches aimed at reducing compression artifacts (typically called "deblocking"), they are generic in nature and tend to blur the image. Our goal is to propose a new method, with the help of gradient domain enhancement, to restore the appearance of tone-adjusted JPEG images that suffer from blocking artifacts. The term "tone-adjusted JPEG image" stands for the image that is enhanced (e.g. increasing contrast) by applying a pre-defined tone curve. A restored example can be found in Figure 1.3(c). More details and results of this work will be discussed in Chapter 4.

Color-aware regularization Reconstructing the final enhanced image from the composite (or modified, in general) gradient field is a crucial step in the gradient domain image enhancement pipeline. This can be usually done by using a suitable optimization approach (e.g. constructing and regularizing a bi-objective function with suitable data/color constraints and smoothness/gradient constraints). According to the study of Omer [56] in 2004, colors of objects in natural images typically follow distinct distributions (forming elongated clusters) in the RGB space. However, the conventional regularization method that uses L_2 norm to formulate the data/color constraint is blind to these color distributions, which may cause undesir-
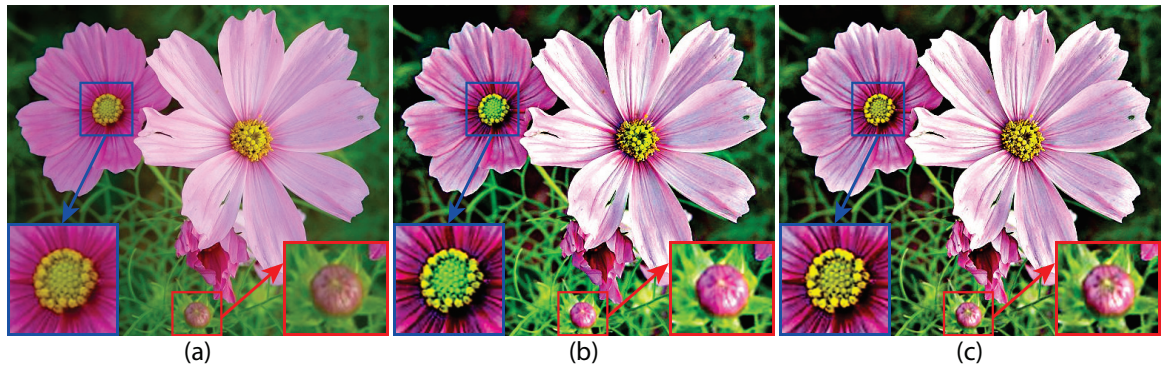
Figure 1.4: An example of our color-aware regularization applied to an image that has had its gradient boosted for a better contrast level: (a) original RGB image; (b) gradient boosting result of conventional regularization method; (c) gradient boosting result of our color-aware regularization method. Note the subtle color-shifting exhibited in (b).

able colors appearing in the final reconstructed image (see Figure 1.4(b), the color in the center region of the flower on the left changes from yellow to green). This is known as the color shift problem.

As a follow-up study of gradient domain image enhancement, we aim to propose a color-aware regularization method by using an anisotropic Mahalanobis distance to control the output colors to better fit original input color distributions, so as to avoid the color shift artifact in the reconstruction step. An example of our method's result is shown in Figure 1.4(c). More details and results of this work will be discussed in Chapter 5.

1.3   Contributions

In this dissertation, three research works amounting to three major contributions have been proposed to advance the state-of-the-art image enhancement techniques

within a gradient domain context and have offered three major contributions in this area. Two of them target new applications of gradient domain processing and the other one contributes on how to improve the reconstruction step of gradient domain processing.

Visual enhancement of old documents We propose a visual enhancement framework for degraded historical documents based on gradient domain fusion. The key of our framework is to take the desired "good" gradients (with more details or less artifacts) from hyperspectral images of the document, and fuse them into the normal RGB image by reconstructing a new image from the composited gradient field. For both text-based documents corrupted with various kinds of artifacts and drawing-based documents with low contrast regions, our framework can effectively enhance their visual quality and legibility. This work has been published in Pattern Recognition'2011 [35]. In addition, our enhancement framework has been integrated as part of a comprehensive hyperspectral image visualization tool used by the Nationaal Archief of the Netherlands. In this visualization tool, the user can interactively select the NIR band to provide the desired gradients. This work has been published in IEEE Visualization'2010 [36].

Compression artifact reduction We present a new method to reduce the blocking artifact that arises from tone-adjustment when applied to JPEG images. Our approach first introduces a simple detection step based on the histogram of oriented gradients (HoG) to find the regions in the tone-adjusted image that exhibit noticeable blocking artifacts. Then we use a dictionary learning method to replace gradients in corrupted regions using a training set of images to which we have applied the same compression and tone-adjustment too. Finally, we obtain the new image using gradient-domain reconstruction technique from its enhanced gradient

field. Our proposed method can produce compelling results that are superior to those obtained with existing deblocking methods.

Color-aware regularization We introduce a color-aware regularization method for use with gradient domain image enhancement to avoid color shift artifact that is very likely to arise in the reconstruction phase. We formulate the color-aware regularization as an anisotropic Mahalanobis distance [20] which can be expressed as a linear system, so that it can be easily incorporated into the existing optimization frameworks. Our color-aware regularization is simple, easy to implement, and does not introduce significant computational overhead. We demonstrate the effectiveness of this regularization method on a variety of inputs using three selected applications, including gradient transfer, gradient boosting and saliency sharpening. This work has been published in ACCV'2012 [17].

1.4 Outline

The rest of this dissertation is organized as follows. Chapter 2 provides some background information on gradient domain image processing techniques. Chapter 3 describes the visual enhancement framework for degraded historical documents. Chapter 4 presents the compression artifact reduction method that is tailored specifically for tone adjustment. In Chapter 5, we propose the color-aware regularization method for general gradient domain image processing methods. Finally, Chapter 6 concludes the dissertation along with a short discussion on possible future research directions.

Chapter 2

Background

Gradient domain processing has been adopted by researchers during recent years to achieve various image processing tasks. As discussed in the last chapter, gradient domain image processing generally involves two steps: 1) task-specific gradient manipulation based on the targeted enhancement task, and 2) a reconstruction phase to obtain the new pixel values from the modified gradient field. This chapter gives some brief background information on these two steps by reviewing several representative gradient domain processing approaches.

2.1   Task-specific Gradient Manipulation

The first step of gradient domain processing is to obtain the desired gradient field according to the given specific task. In this section, we mainly review how the existing gradient domain approaches manipulate image gradients to achieve this goal. Gradient domain manipulation can be generally grouped into two categories: per pixel manipulation and corresponding gradients manipulation in two images.
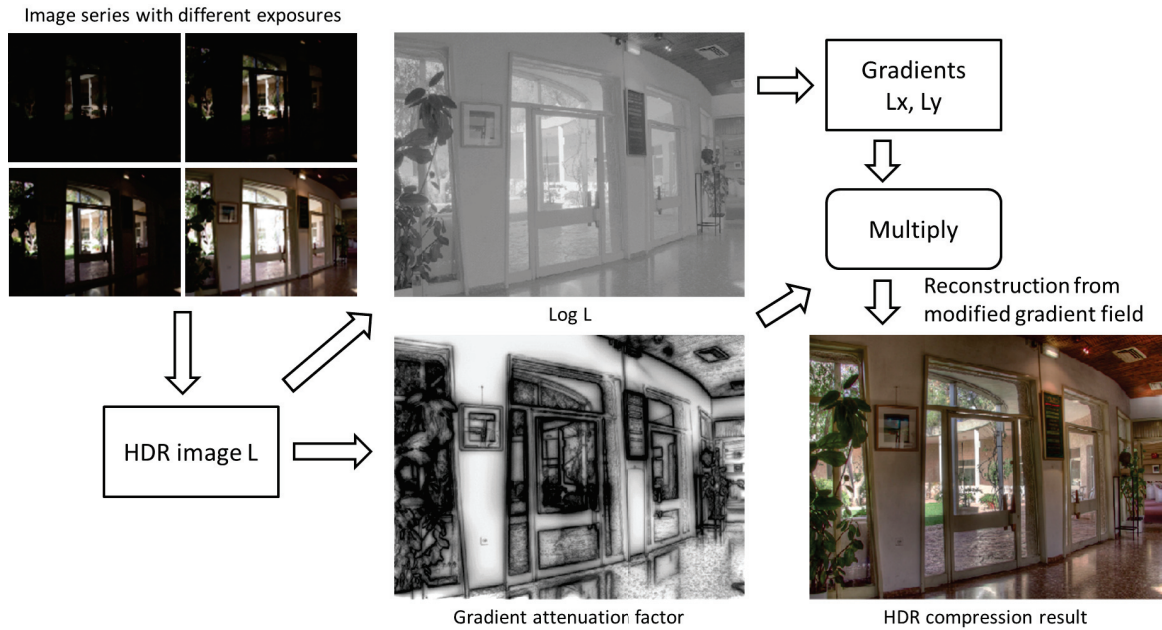
Figure 2.1: An overall workflow of HDR compression method.

2.1.1   Per pixel manipulation

Per pixel manipulation is mostly used by image processing algorithms that take single image as input, and can be done in the following manners:

- Set to zero (shadow removal, texture de-emphasis)

- Non-linear operations (HDR compression, local illumination change)

- Poisson matting

We select the HDR compression application as the representative work to briefly review its gradient manipulation operation/technique, since we also manipulate gradients in a similar way for the saliency sharpening application which will be discussed in Chapter 5.

HDR compression Fattal et al. proposed a high dynamic range (HDR) compression method [24] to render HDR images on conventional low dynamic range (LDR) displays. Their approach manipulates the gradient field of the luminance image by attenuating the magnitudes of HDR image gradients by a factor of $\Phi(x, y)$ at each pixel. The key idea is to progressively attenuate the HDR gradients by shrinking gradients of large magnitude more than small ones. An overall workflow of their HDR compression method is shown in Figure 2.1. The HDR image L is first obtained from a series of photographs taken under different exposures. The gradient attenuation factors (darker shades indicate smaller scale factors and strong attenuation) are computed and then multiplied with the gradients of the logarithm of HDR image L to compress the HDR radiance map in the gradient domain. Finally, the LDR image can be reconstructed from the attenuated gradient map.

2.1.2 Corresponding gradients manipulation in two images

Corresponding gradients manipulation is usually used by image processing approaches that take two (or more) images as input, and can be done in the following manners:

- Binary choose or copying operation (Poisson image editing, seamless cloning)

- Max operator (day/night image fusion, visible/NIR image fusion)

- Projection tensors (reflection removal)

- Vector operations (flash/no-flash image combination)

We select two representative applications from the above list, Poisson image editing and day/night image fusion, and review their gradient manipulation manners as follows.
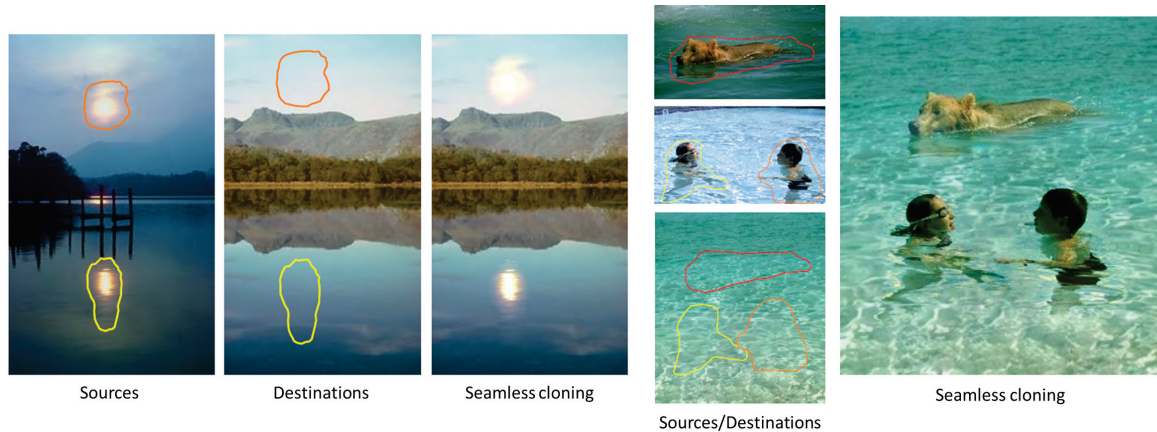
Figure 2.2: Seamless cloning examples using Poisson image editing.

Poisson image editing Poisson image editing (PIE) [59] is a seamless cloning method that can seamlessly blend a region of interest (ROI) from the source image onto the target image. The gradient manipulation operation/technique performed by PIE is just a simple copy-and-paste operation. The gradients of the ROI from the source image are copied and then pasted to an appropriate position in the gradient field of target image. To achieve the seamless cloning effect, a hard boundary constraint is enforced to make the boundary color of the pasted ROI agree with that of the target image. Finally, the cloning result is reconstructed from the composite gradient field by solving a Poisson equation. Two examples of seamless cloning using PIE is shown in Figure 2.2. A similar copy-and-paste gradient manipulation operation/technique is used for the gradient transfer application which will be discussed in Chapter 5.

Day/night image fusion Raskar et al. proposed a gradient domain image fusion method [63] to automatically combine images of a scene captured under day-time and night-time for context enhancement purpose. An overall workflow of this
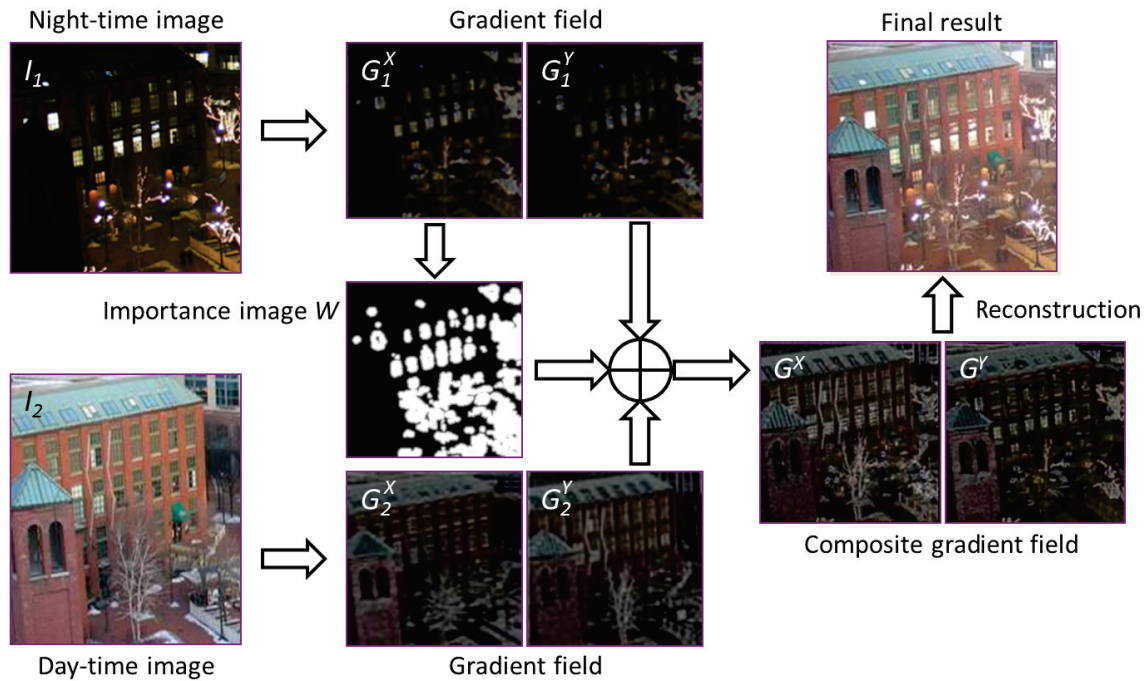
Figure 2.3: An overall workflow of day/night image fusion method.

approach is shown in Figure 2.3. First, the gradient fields of day-time and night-time images are computed in both x and y directions. The next step is to find the locally-important areas, which are considered as regions of high variance in the night-time image. This can be done by simply thresholding the gradient field of the night-time image to keep large scale gradients. With the importance image W, a new composite gradient field can be easily constructed. For those white pixels in W, their gradients are taken from the gradient field of the night-time image, and gradients of the rest pixels are taken from the gradient field of the day-time image. Lastly, the final result is reconstructed by integrating the composite gradient field. Such kind of gradient manipulation operation/technique (with certain importance/saliency masks) is also adopted by our visual enhancement

framework (Chapter 3) and compression artifact reduction method (Chapter 4).

2.2 Reconstruction from Modified Gradient Field

Another crucial procedure of gradient domain processing is to estimate the new pixel values from the composite/modified gradient field obtained in the last stage, which is usually referred to as the reconstruction phase. In this section, we introduce two common strategies that are widely used to reconstruct the new image from its gradient field.

2.2.1 Poisson equation

Early gradient domain processing approaches [24, 59, 31, 63, 54] were formulated using the Poisson equation. Taking the Poisson image editing method as example, we discuss how to use the Poisson equation to formulate gradient domain problems. Given the target gradient field G, we look for an image I with gradient field closest to G in the least squares sense. More formally, the final result I can be solved by minimizing the following equation:

$$F(\nabla I, G) = \|\nabla I - G\|^2 = \left(\frac{\partial I}{\partial x} - G_x\right)^2 + \left(\frac{\partial I}{\partial y} - G_y\right)^2, \tag{2.1}$$

where $\nabla. = [\frac{\partial .}{\partial x}, \frac{\partial .}{\partial y}]$ is the gradient (partial derivative) operator. To minimize Eq. 2.1, I must satisfy the associated Euler-Lagrange equation:

$$\frac{\partial F}{\partial I} - \frac{d}{dx}\frac{\partial F}{\partial I_x} - \frac{d}{dy}\frac{\partial F}{\partial I_y} = 0, \tag{2.2}$$

after substituting F in Eq. 2.2 we can get:

$$2\left(\frac{\partial^2 I}{\partial x^2} - \frac{\partial G_x}{\partial x}\right) + 2\left(\frac{\partial^2 I}{\partial y^2} - \frac{\partial G_y}{\partial y}\right) = 0, \tag{2.3}$$

from Eq. 2.3 we can obtain the final Poisson equation:

$$\nabla^2 I = div(G_x, G_y) = \frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y}, \tag{2.4}$$

where $\nabla^2 . = \frac{\partial^2 .}{\partial x^2} + \frac{\partial^2 .}{\partial y^2}$ is the Laplacian operator and div is the divergence operator.

In order to enforce the boundary color of the pasted ROI to agree with the color of target image, a hard boundary constraint is needed. The Dirichlet boundary condition has been adopted by the Poisson image editing method:

$$I(x, y) = I_0(x, y), \forall(x, y) \in \partial\Omega, \tag{2.5}$$

where I_0 is the target image and $\partial\Omega$ is the boundary of pasted ROI. To obtain the final reconstructed image I, we solve the Poisson equation Eq. 2.4, subject to the boundary condition in Eq. 2.5. Eq. 2.4 and Eq. 2.5 can be expressed as a large sparse linear system and solved by many numerical methods such as direct solvers, multigrid, preconditioned conjugate gradients, etc. These numerical methods are out of the scope of this dissertation. Readers may refer to related literatures for more details if interested.

2.2.2 Optimization scheme with L_2 norm regularization

While generally sufficient for most gradient domain processing approaches, the Poisson equation formulation can, from time to time, exhibit very noticeable color

shifts inside the processed region. This is because only the color of boundary pixels in the processed region has been constrained against certain color constraints (e.g. the Dirichlet boundary condition). Recent gradient domain processing approaches [9, 85, 40, 82, 79, 67] impose color constraints over the entire 0th order domain (color domain) of the solution. This is typically done using an L_2 norm regularization on one or more of the 0th order color channels. This solution results in a bi-objective function that tries to manipulate the image gradients while minimizing the Euclidean error (i.e. L_2) between the original and output 0th order domains, which can be considered as an optimization scheme with L_2 norm regularization.

Taking the gradient transfer application as example, we review the conventional optimization framework based on an L_2 regularization term. The purpose of gradient transfer is to transform gradients from the NIR image g (source image, with more desired details) to the RGB image f (target image) while preserving the original look-and-feel of the RGB image (f and g are precisely aligned). That is, we seek a new image u whose colors (from one or more color channels) are as close as possible to f, and at the same time, has gradients that are as close as possible to g. More formally, the final result u is generated by minimizing the following bi-objective cost function

$$E(u) = \sum_{p \in u} (\lambda E_d(p) + E_s(p)), \tag{2.6}$$

where p is the pixel index of image u; E_d is the 0th order color constraint term and E_s is the 1st order gradient constraint term; λ is used for the balance between E_d and E_s. These two terms are defined as:

$$E_d(p) = (u_p - f_p)^2 \tag{2.7}$$

and

$$E_s(p) = \left(\frac{\partial u}{\partial x} - c \cdot \frac{\partial g}{\partial x}\right)_p^2 + \left(\frac{\partial u}{\partial y} - c \cdot \frac{\partial g}{\partial y}\right)_p^2, \qquad (2.8)$$

where $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ denote the partial derivative operators in x- and y-direction; c is a scaling factor used to control the strength of target gradient field.

Using vector and matrix notation we may rewrite Eq. 2.6 as:

$$\begin{aligned} E(\mathbf{u}) \;&=\; \lambda E_d(\mathbf{u}) + E_s(\mathbf{u}) = \lambda(\mathbf{u} - \mathbf{f})^T(\mathbf{u} - \mathbf{f}) \\ &+ (G_x\mathbf{u} - c \cdot G_x\mathbf{g})^T(G_x\mathbf{u} - c \cdot G_x\mathbf{g}) \\ &+ (G_y\mathbf{u} - c \cdot G_y\mathbf{g})^T(G_y\mathbf{u} - c \cdot G_y\mathbf{g}), \end{aligned} \qquad (2.9)$$

where G_x and G_y are 1st order forward differentiation operators. Here $\mathbf{u}$, $\mathbf{f}$, and $\mathbf{g}$ are all single channel images represented by column vectors (for RGB image, we can minimize Eq. 2.9 respectively for each color channel). Minimizing Eq. 2.9 amounts to taking its derivative, setting it to zero, and solving for vector $\mathbf{u}$ that is uniquely defined as the solution of the following linear system:

$$(\lambda I + G_x^T G_x + G_y^T G_y)\mathbf{u} = \lambda\mathbf{f} + c \cdot (G_x^T G_x\mathbf{g} + G_y^T G_y\mathbf{g}). \qquad (2.10)$$

To solve this linear system, many numerical methods can be used. In this dissertation, we adopt the conjugate gradient method [6] to iteratively solve Eq. 2.10 and obtain the reconstructed output image $\mathbf{u}$. Note that for this specific gradient transfer task, our target gradient field is exactly the gradient field of the NIR image $\mathbf{g}$, so we can simply use its gradients $G_x\mathbf{g}$ and $G_y\mathbf{g}$ as the gradient constraint in Eq. 2.9. However, for other gradient domain processing approaches, the task-specific target gradient field may vary from case to case. We can replace the target

gradient field $G_x\mathbf{g}$ and $G_y\mathbf{g}$ with more general notations $\mathbf{TG}_x$ and $\mathbf{TG}_y$ in Eq.2.9. $\mathbf{TG}_x$ and $\mathbf{TG}_y$ may be any appropriate gradient field (e.g. combined gradient field from two or more images) depending on the requirement of the specific task.

2.3 Summary

This chapter provided background on several gradient manipulation tasks (HDR compression, Poisson image editing, and day/night image fusion) and described how the gradient was manipulated in a task-specific manner for each task. In addition, we also provided background on two conventional reconstruction methods (Poisson and an optimization scheme) that are used to estimate new pixel values from the gradient field. As the first step of gradient-domain processing is task specific, Chapters 3 and 4 will introduce additional related works specific to the gradient domain image enhancement addressed in those topics. Chapter 5 will directly work on the bi-objective function of the optimization framework presented in Section 2.2.2.

Chapter 3

Visual Enhancement of Documents using Gradient Domain Fusion

In this chapter, we describe a visual enhancement framework for old documents that suffer from various kinds of artifacts based on gradient domain fusion technique. The key idea is to replace the gradients in degraded regions with the desired "good" gradients (with less artifacts or more details) so as to compose an enhanced gradient field, and then reconstruct the final image from gradients using an optimization scheme. These "good" gradients are provided by hyperspectral images of the document, especially the NIR images. We start the introduction of this chapter with some background knowledge of hyperspectral imaging.

3.1   Introduction

Hyperspectral imaging (HSI) captures a densely sampled spectral response of a scene object over a broad spectrum including invisible spectra such as ultraviolet (UV) and near-infrared (NIR). Hyperspectral imaging has been employed

in various scientific disciplines to provide valuable data for fields such as astronomy [46, 13], earth science and remote sensing [51, 19], biological and medical data [68, 65], and computer vision [58]. With the advances in technology and cost reductions, hyperspectral imaging of historical art works and documents can now be used in national libraries and archives [57, 14].

One advantage of HSI in document imaging over the standard 3-channel imaging (i.e. RGB) is that HSI provides a detailed quantitative measurement of the document spectral response. Traditional RGB imaging, on the other hand, contains only a subset of the information available by combining the response of all visible electro-magnetic (EM) radiation into three bands. This makes HSI more suitable for tasks that require accurate quantitative measurement such as conservation, detecting damage, and the analysis of the original features (like ink and pigments) and the changes over time (due to ageing or light exposure) in a document. In addition, hyperspectral imaging provides measurements in the invisible spectrums (NIR, UV) which further enrich the available details and enable richer analysis and enhancement of the data. Measurements in the invisible spectral bands provide more useful information about the document being imaged by sometimes seeing more details than the visible range and by sometimes seeing less artifacts than the visible range. This is demonstrated by two examples in Figure 3.1. For the first example, the NIR band at 900nm (Figure 3.1 (b)) provides more salient gradient details than the document in the visible band at 500nm (Figure 3.1 (a)). Conversely, for the second example, the NIR band at 800nm (Figure 3.1 (d)) is better for guiding enhancement than the 450nm visible band (Figure 3.1 (c)) since artifacts such as ink-bleed and ink-corrosion are less prevalent.

The goal of this work is to take advantage of hyperspectral images of historical

(a) 500 nm	(b) 900 nm
(c) 450 nm	(d) 800 nm
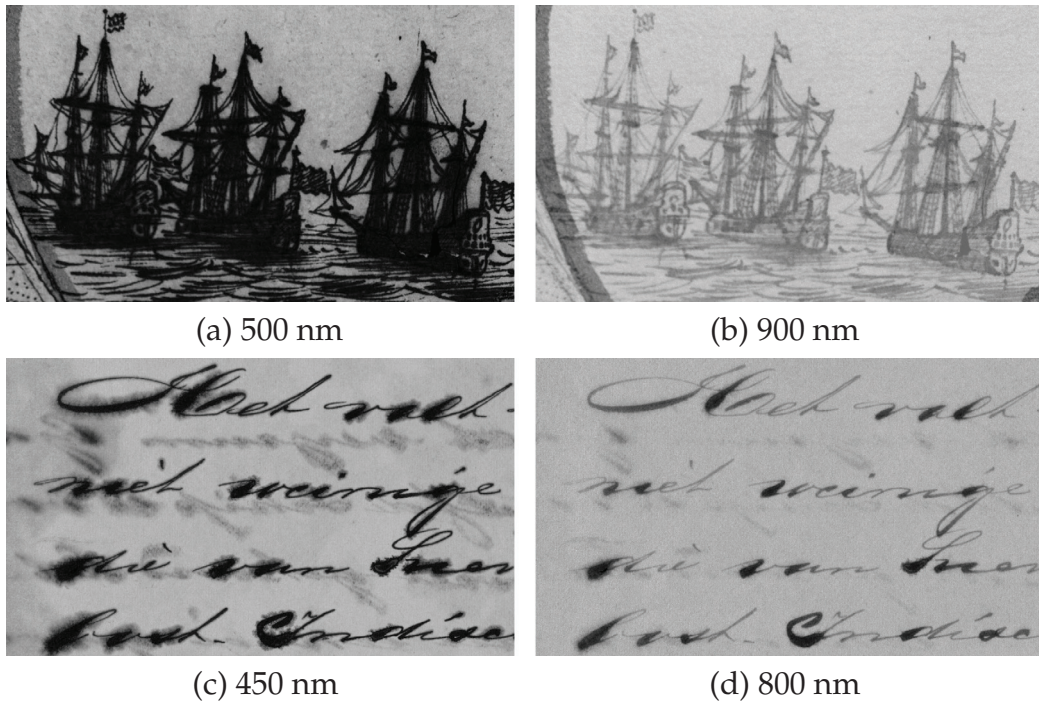
Figure 3.1: Hyperspectral imaging provides measurements in invisible spectral ranges which helps to improve data analysis. In the first example, the image in the NIR band (b) captures more details of the image content which is barely seen in the visible band (a). In the second example, the NIR image (d) is useful because it does not exhibit as many undesired artifacts as the visible bands (c).

documents to visually enhance the document's content by exploiting additional information provided by the NIR bands. The visual enhancement in this work is applied to the RGB image of the hyperspectral data as the RGB image is the most natural visualization of the data. In this work, we are interested in two tasks. For the text-based documents that are corrupted with artifacts such as ink-bleed, corrosion, and foxing, we use the invisible bands which capture much less artifacts than the visible bands to clean up the artifacts in the documents while preserving the look and feel of the original document. For drawing-based documents that contain low contrast regions, we use NIR bands which capture more details than the visible bands to enhance the contrast in the documents. The data are enhanced in the

gradient domain which has been shown to be effective for many computer vision tasks such as image editing [59], contrast adjustment [24], image stitching [45], and intrinsic image computation [78]. The key components of our algorithm include detecting regions that can be enhanced by the additional NIR spectral images, composing the enhanced gradient map from NIR images, and reconstructing the final image from gradients using an optimization scheme.

This work is a part of an ongoing collaborative effort with the Nationaal Archief of the Netherlands (NAN), one of Europe's leading research archives, and Art Innovation, a manufacturer of hyperspectral imaging hardware designed for historical documents. The documents presented in this chapter, which are indicative of the type of artifacts common to historical documents, are imaged at NAN using the SEPIA Quantitative Hyper-Spectral Imager (QHSI) device developed by Art Innovation [38]. The device performs hyperspectral imaging by capturing a very narrow spectral band of EM radiation one at a time by placing a bandpass filter in front of the light source to block out all but a selected band of the EM spectrum. A monochromatic camera is then used to capture the amount of light that is reflected by the document at that selected band. The filter is changed for each image, thus capturing different parts of EM spectrum to build up the HSI (Figure 3.2).

The QHSI device captures images at different wavelength bands from 365 nm (UV) to 1100 nm (near infrared (NIR)) with the step size of 10nm in most cases except the bands in 300nm's and 1000nm's. The images have the resolution of 4 mega pixels (2048×2048) for a physical surface area of 125 mm × 125 mm and are captured at 16 bit per pixel. Such high-resolution (approximately 256 pixels per mm^2) provides a reliable spatial measurement suitable for even thin lines of handwriting and printed text.
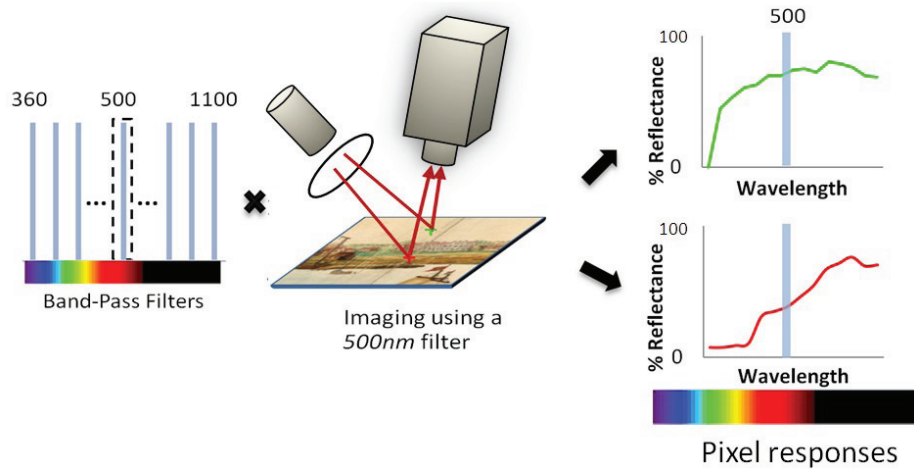
Figure 3.2: Hyperspectral imaging process. At each scan, a monochrome camera measures the reflected light from the document surface. The document reflects a very narrow band of EM radiation due to the bandpass filter positioned in front of the light source (500nm in this example). This process is repeated using 70 different bandpass filters to build the HSI.

The remainder of the chapter is organized as follows: we begin by reviewing the related work in Section 3.2. In Section 3.3, we introduce our algorithm for visually enhancing old documents using the hyperspectral data. We show experimental results in Section 3.4 and summarize our enhancement framework in Section 3.5.

3.2 Related Work

Existing approaches that are most relevant to our work are the image fusion techniques. There are several different categories in image fusion. Image fusion plays a vital role in remote sensing where the goal is to fuse different types of images from satellites or aircrafts to provide an increased visual saliency of the area being imaged. The types of imagery provided cover different portions of the electromagnetic spectrum at different spatial, temporal, and spectral resolution [12]. In

many cases, the focus of fusion techniques in remote sensing is to assist with the interpretation of the data by simple false-coloring. A comprehensive review of image fusion in remote sensing is offered in [61].

In computer vision and graphics, image fusion of flash/no-flash photographs has gained interest to assist in imaging in dark environments. In [21] and [60], flash images were used to significantly enhance the details and reduce noise in ambient images. Both of these approaches use joint-bilateral filters to decompose images into detail and base layers, and reconstruct the image by combining the large scale of the ambient image and the details of the flash image. In [1], a gradient projection scheme for flash/no-flash image fusion was introduced with the goal of removing the flash artifacts. In related work, an image fusion technique for combining images captured under different illuminations for context enhancement was introduced in [63].

The closest work to ours is the multispectral image fusion methods introduced in [8, 88, 40]. In [8], a video taken in a low light environment is enhanced by fusing the visible-spectrum video with the video simultaneously captured with an infrared sensor. This work introduced a modified bilateral filter suited for multispectral imagery to essentially decompose the images to low frequency contents and details. In [88], low contrast photographs were enhanced with NIR images by incorporating texture information from NIR. They applied wavelet decomposition to decompose the image into low frequency and high frequency details similar to other methods. The details of the NIR image were transferred by histogram matching. An extension of flash/no-flash techniques to multispectral imaging was introduced in [40]. In their work, a prototype of camera and flash that uses infrared and ultra-violet also was presented. They exploit the correlations between images at different spectral

26

bands to reduce noise and add fine details in the ambient image.

Our work can be viewed as the extension of the multispectral image fusion to hyperspectral image fusion[1]. In multispectral image fusion methods [8, 88, 40], images consist of three visible-spectrum images (RGB) and one extra image that integrates information in NIR (an extra UV image is included in [40]). In contrast, we deal with much more images densely sampled over a broad range of spectrums. Each hyperspectral image example shown in this chapter contains 70 spectral images sampled from 365nm to 1100nm. This brings more challenges and advantages to the fusion problem since there is significantly more information available. Hence, one of the main contributions of this work is to present methods to detect regions in the visible-spectrum images that can be enhanced by using the information from NIR images and to extract a single map containing information to be fused from all the NIR images.

3.3 HSI Document Enhancement Algorithm

As mentioned earlier, there are two types of enhancements that are targeted in this work. With text documents, our algorithm aims to remove the undesired artifacts, notably ink-bleed, ink-corrosion, and foxing (age related spots). The final results are enhanced documents that still maintain the look of the original with the undesired artifacts significantly reduced. For this task, the images in the invisible range provide the source for the background of the enhanced image since invisible range images are much less affected by the corrosive artifacts. For the second task pertaining the documents with line drawings and having low contrast due

---

[1]The distinction between multispectral vs. hyperspectral is typically related to the number of bands as well as the manner in which the data is collected. We refer to our data as hyperspectral as it provides a densely sampled (10nm intervals) spectral response from a single sensor.

to ink-corrosion, our algorithm increases the contrast and adds image details not present in the visible range. For this task, the invisible range images are sources of foreground ink with salient gradient details.

While the data can be enhanced by modifying each spectral band image in the visible spectrum and then constructing an RGB image, we chose to enhance the RGB image of the hyperspectral data in order to reduce computation[2]. The enhancement is performed in the gradient domain and both tasks follow similar procedures. In the first step, pixels are segmented into two groups: a group that should remain the same (group A) and a group that needs to be enhanced (group B). A new gradient map is then composed by combining gradients from different spectral band images. For the pixels in the group A, the gradients remain unchanged. For the pixels in the group B, the gradients are replaced by the gradients from the invisible band images that suit the purpose of the task. Finally, output images are reconstructed from the gradient maps by an optimization scheme.

The following notations will be used throughout the chapter. The term I_λ indicates the image of the data at a spectrum with wavelength λ, $\mathbf{x}$ indicates a pixel location, and $\mathbf{s_x}$ is the spectral response of the point $\mathbf{x}$:

$$\mathbf{s_x} = [I_{\lambda_1}(\mathbf{x}), I_{\lambda_2}(\mathbf{x}), \cdots, I_{\lambda_m}(\mathbf{x})]^T.$$

In the following, we describe our algorithm starting with the details from compositing gradient maps to reconstructing images from gradients.

---

[2]Note that there is no RGB image in the HSI data. The RGB image must be computed by applying synthetic lighting and integrating the visible spectral response.

<div align="center">(a)   (b)   (c)   (d)</div>
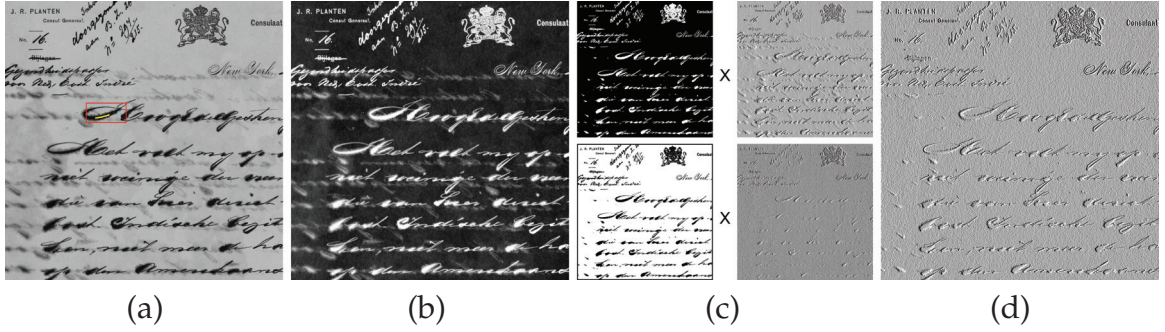
Figure 3.3: Gradient map construction for text documents: (a) input image and an user mark-up, (b) similarity map S, (c) gradients for foreground ∇I_i and background ∇I_λ, (d) gradient composite G.

3.3.1 Gradient map composite for artifact removal

Figure 3.3 summarizes the algorithm for composing a composite gradient map for the artifact removal tasks in text documents. The first step is to segment the input into foreground (texts) and background. In our system, we employ a simple user-assisted strategy where the user needs to only provide a mark-up on a small area that belongs to the foreground (highlighted using yellow color in the red box in Figure 3.3 (a)). After the segmentation, the mean spectrum of the foreground pixels ($\bar{\mathbf{p}}$) is computed, then a similarity map S for all the pixels is computed (Figure 3.3 (b)):

$$S(\mathbf{x}) = 1 - \frac{\left\| \bar{\mathbf{p}} - \mathbf{s_x} \right\|^2}{m}, \tag{3.1}$$

where m is the number of spectral samples. After applying thresholding, we get a binary mask M which has 1's only in the foreground region. To composite a new gradient map (G) for the enhancement, we use the gradient of the original image (∇I_i) for the foreground and the gradient of an image from the invisible range (∇I_λ, $700 < \lambda <= 1100$) for the background. The band with the smallest variation in the

| (a) image at 500nm | (b) image at 900nm | (c) spectral plots |
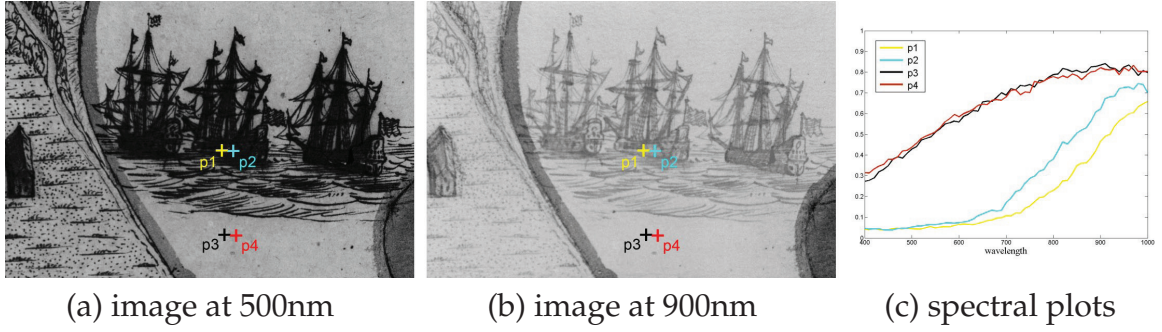
Figure 3.4: We detect regions where the local contrast is much higher in the NIR bands than the visible bands to apply enhancement using the NIR bands.

background region is chosen for the I_λ. Hence, the gradient map G is generated as follows :

$$G = M \circ \nabla I_i + (1 - M) \circ \nabla I_\lambda. \tag{3.2}$$

The operator $\circ$ is the Hadamard product (the entry-wise product of two matrices). This procedure is applied individually to each of the RGB channel of the input image.

3.3.2 Gradient map composite for contrast enhancement

For the contrast enhancement task, the procedure is similar to the procedure for the artifact removal, but can be performed automatically. The first step is to determine which regions can be enhanced by using the additional information from the NIR bands. To do this, we use the observation that the local contrast of regions that can be enhanced with additional spectral bands is much higher in the NIR bands than the local contrast in the visible bands. In Figure 3.4 (a), the local contrast inside the drawings of the ships is extremely low in the visible bands (for example, points $p1$ and $p2$). The local contrast inside these regions greatly increases in the NIR bands

(a)　　　　　　　　(b)　　　　　　　　(c)　　　　　　　　(d)
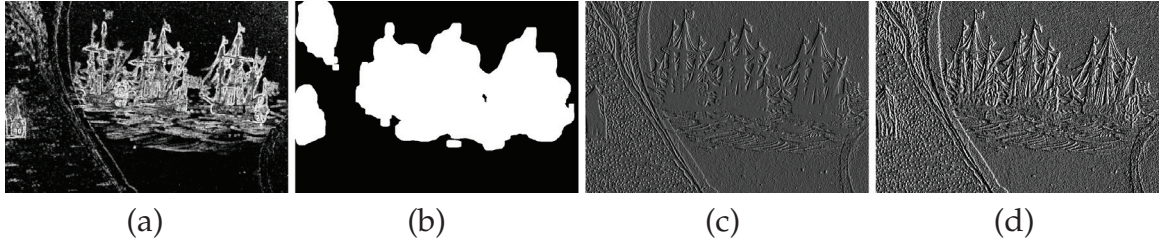
Figure 3.5: Gradient map composition for enhancement : (a) saliency map S (Eq.3), (b) binary mask M, (c) original gradient map ∇I_i, (d) new gradient composite G.

as can be seen in Figure 3.4 (b), (c). In contrast, the local contrasts remain constant throughout the spectrum in some other regions in the data (points *p3* and *p4*). It is unnecessary to enhance these regions with additional bands and risk amplifying the sensor noise. Using the insight explained above, we compute a saliency map (*S*, Figure 3.5 (a)) as follows :

$$S(\mathbf{x}) = \frac{1}{m'} \sum_{\lambda \in \text{NIR}} \max_{\mathbf{y} \in N(\mathbf{x})} \left| I_\lambda(\mathbf{x}) - I_\lambda(\mathbf{y}) \right| - \max_{\mathbf{y} \in N(\mathbf{x})} \left| I_i(\mathbf{x}) - I_i(\mathbf{y}) \right|, \quad (3.3)$$

where I_i is the input image, $N(\mathbf{x})$ refers to the neighbors of $\mathbf{x}$, and m' is the number of spectral bands in the NIR. We tried different neighborhood sizes (3, 5, 7, 9, 11) and found that 7×7 gives relatively good results. Hence, in this work, we use the size of 7×7 for the neighborhood.

After thresholding the saliency map S with a morphological operation to increase its size slightly, we again obtain a binary mask M as the previous task (Figure 3.5 (b)). The foreground mask M has 1's only in the region where the enhancement is necessary. A new gradient map G for the enhancement is computed as follows :

$$G = M \circ G' + (1 - M) \circ \nabla I_i. \quad (3.4)$$

Determining the new gradient G' in Eq. 3.4 is the key to enhancing the details in the original image. One option for computing G' is to select a band in NIR with high contrast similar to choosing a band with lowest variation as the background in Sec. 3.3.1. This may not be the best choice since we have observed that different regions in the image have higher contrast in different NIR bands. Hence one suitable option for computing G' is to integrate information from all available NIR bands and choose different bands to provide the strongest gradients for each pixel in these regions. To maintain spatial consistency in this gradient assignment we formulate the problem as a Markov Random Field (MRF) where each pixel $\mathbf{x}$ is assigned a label $l_\mathbf{x} \in$ NIR. To solve the MRF, the following energy is minimized in order to find optimal pixel labels :

$$E = E_d + \omega E_s,\qquad(3.5)$$

where E_d is the data-cost energy reflecting the likelihood of assigning a label to each pixel, E_s is the smoothness energy representing the cost of assigning different labels to adjacent pixels, and ω is the weight that controls the strength of the smoothness term. The data cost E_d is computed as follows :

$$E_d(l_\mathbf{x} = \lambda) = -|\nabla I_\lambda(\mathbf{x})|, \quad \lambda \in \text{NIR}.\qquad(3.6)$$

We enforce smoothness on adjacent pixels $(\mathbf{p}, \mathbf{q})$ with the following smoothness cost :

$$E_s(l_\mathbf{p}, l_\mathbf{q}) = |l_\mathbf{p} - l_\mathbf{q}| .\qquad(3.7)$$

Our MRF is optimized using the Middlebury MRF library [70] with the graph-cuts solver [10]. After the labeling is complete, the gradient map G' is constructed as

follows :

$$G'(\mathbf{x}) = \nabla I_{l_x}(\mathbf{x}). \tag{3.8}$$

3.3.3   Image reconstruction from a gradient map

We now have to reconstruct an image from the gradient map computed in Eq. 3.2 or Eq. 3.4 (refer to Section 2.2.2 in Chapter 2 for background information on the reconstruction stage). The goal is to reconstruct an image R in which the intensities are close to the input image I_i and the gradients are close to the computed gradient map G. We use the following cost function for the reconstruction which is similar to the one used in [40]:

$$\underset{R}{\operatorname{argmin}} \sum_{\mathbf{x}} \left[\gamma (R(\mathbf{x}) - I_i(\mathbf{x}))^2 + \left| \nabla_x R(\mathbf{x}) - \beta G_x(\mathbf{x}) \right|^\alpha + \left| \nabla_y R(\mathbf{x}) - \beta G_y(\mathbf{x}) \right|^\alpha \right], \tag{3.9}$$

The first term in Eq. 3.9 forces R to be close to the input image I_i under an ℓ^2 norm. The second and the third term make the gradients of the reconstructed image R to be close to the gradient map G under a sparse norm ($\alpha \leq 1$) where sub-indices x,y refer to the gradient directions. Using the sparse norm on the gradient terms encourages the edge structures in R to align spatially with those in G in contrast to the ℓ^2 norm where ∇R will be matched closely with G [40]. The parameter γ controls the balance between the effect of the first term and the gradient terms. The parameter β controls the strength of the gradient composite on the final image.

To optimize Eq. 3.9, we modified the fast optimization scheme used for image deconvolution recently introduced in [41]. With this optimization scheme, we were able to process our reconstruction in 20 seconds per channel, a significant improvement over the suggested solver in [44, 40] which required 8 minutes per
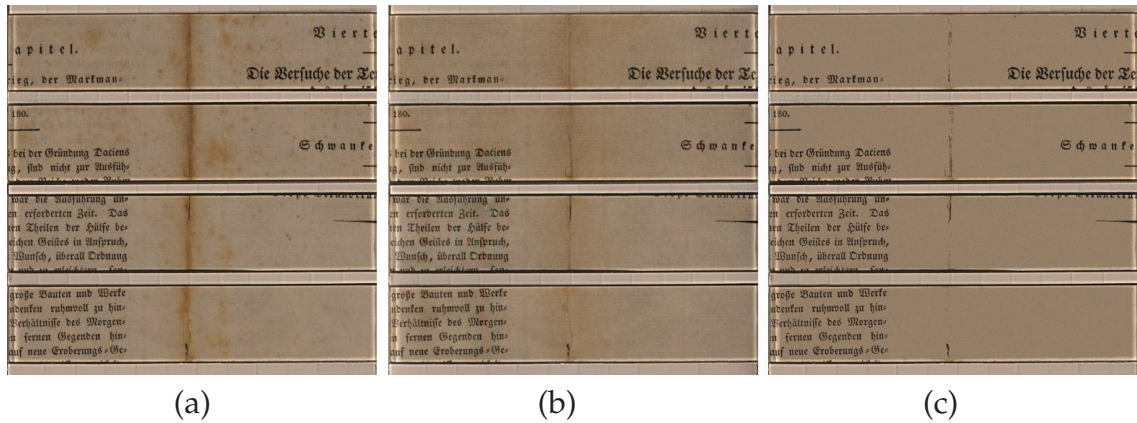
Figure 3.6: (a) The original RGB image is visually enhanced by reducing the foxing artifact. (b) With the hyperspectral data, the enhanced image preserves the texture and the look of the original image. (c) Image reconstructed by replacing the background with the mean value does not look natural.

channel.

3.4   Experiments

The first set of experiments target the removal of artifacts on the documents. The document in the first example (Figure 3.6) is visually corrupted with foxing. The result of our enhancement algorithm is shown in Figure 3.6 (b). The foxing artifact is greatly reduced in the enhanced image while the texture and the look of the original image is preserved. The ability to preserve the look and feel is one significant advantage offered by having the additional gradients in the NIR information. In related document processing work [66, 71, 30, 52], the output of the artifact removal is a binary image with a uniform color background and a color for the foreground. While the binarization enhances the ability to interpret the data, the texture and the look of the original document is completely lost. This is shown in Figure 3.6
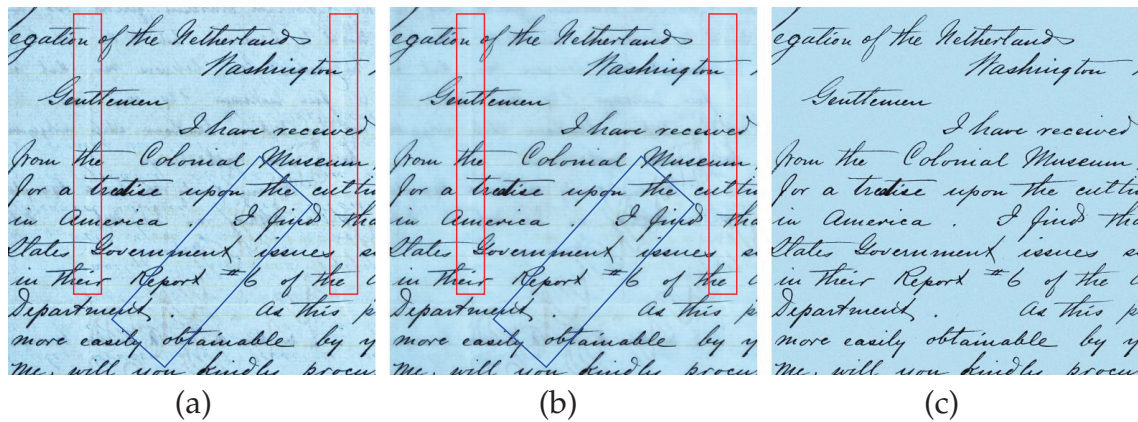
Figure 3.7: (a) The original RGB image is visually enhanced by removing the ink-bleed artifact. (b) With the hyperspectral data, the enhanced image preserves the texture and the look of the original image. Note that the watermark (blue rectangle) and the fold lines (red rectangle) on the image are preserved. (c) Image reconstructed by replacing the background with the mean value completely loses the look of the original document.
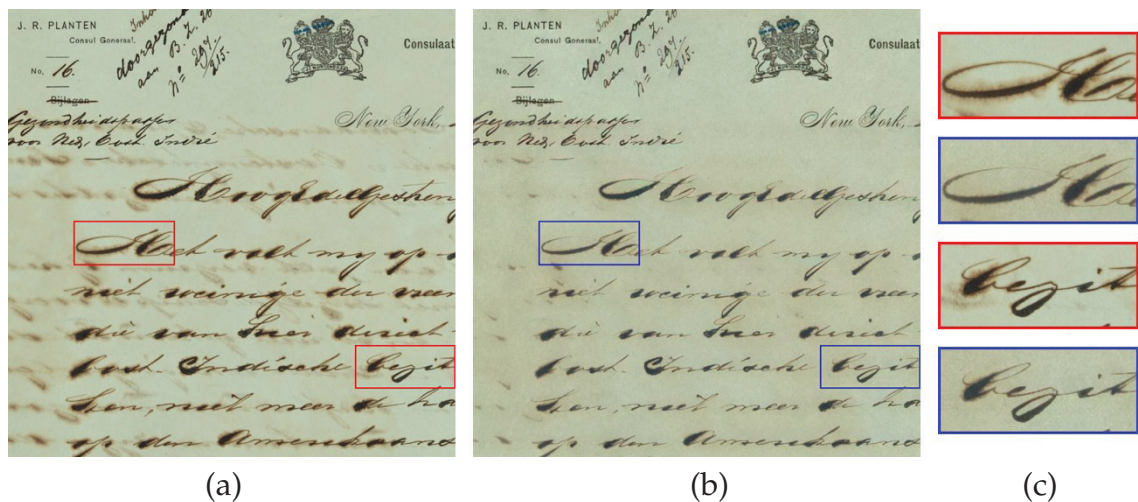
Figure 3.8: (a) The original document contains severe ink-bleed and corrosion artifacts. (b) The artifacts are reduced and the image is visually enhanced with our algorithm. A close views of selected regions are shown in (c).

(c) where the background is filled with the mean RGB of the background in the original image with a Gaussian noise (to simulate the paper texture). As can be seen, this looks unnatural compared with the HSI enhancement.

Another example of artifact removal is shown in Figure 3.7. In this example, a document is corrupted with mild ink-bleed. Using our method, the ink-bleed artifact is removed while the textures and the lines in the original image are preserved. Note that the watermark in the original image is also preserved in our reconstruction. While thresholding could be used to remove the ink-bleed for this input, replacing the background with the mean RGB completely loses the feel and important features such as lines and the watermark of the document (Figure 3.7 (c)). The document in Figure 3.8 (a) is significantly more affected by ink bleed and corrosion. The artifacts are greatly reduced using algorithm as can be seen in Figure 3.8 (b). Some artifacts can be still seen in this example because the spectrum of the pixels with strong ink-bleeds have the same spectral properties as the foreground text. Additional user assistance such as in [52] could be used for further enhancement of the document in this case.

Next, we show results on the contrast enhancement of documents with line drawings. Figure 3.9(a) shows a part of early map of Syracuse drawn circa 1700. Due to corrosion, contrast in some parts of the map decreased resulting in the loss of details (see close-up view in Figure 3.9(a)). We first test contrast enhancement by using just one NIR band as the source for the gradient map as the artifact removal examples. Figure 3.9 (b),(c),(d) show enhancement results using NIR images at 750nm, 850nm, and 950nm respectively. Since images at different wavelength show different contrast in different regions, one image is not enough to yield enhancement in all areas. For example, while the house region is enhanced well

<div align="center">(a) original image in RGB</div>

<div align="center">(b) image enhanced with 750nm</div>

<div align="center">(c) image enhanced with 850nm</div>

<div align="center">(d) image enhanced with 950nm</div>
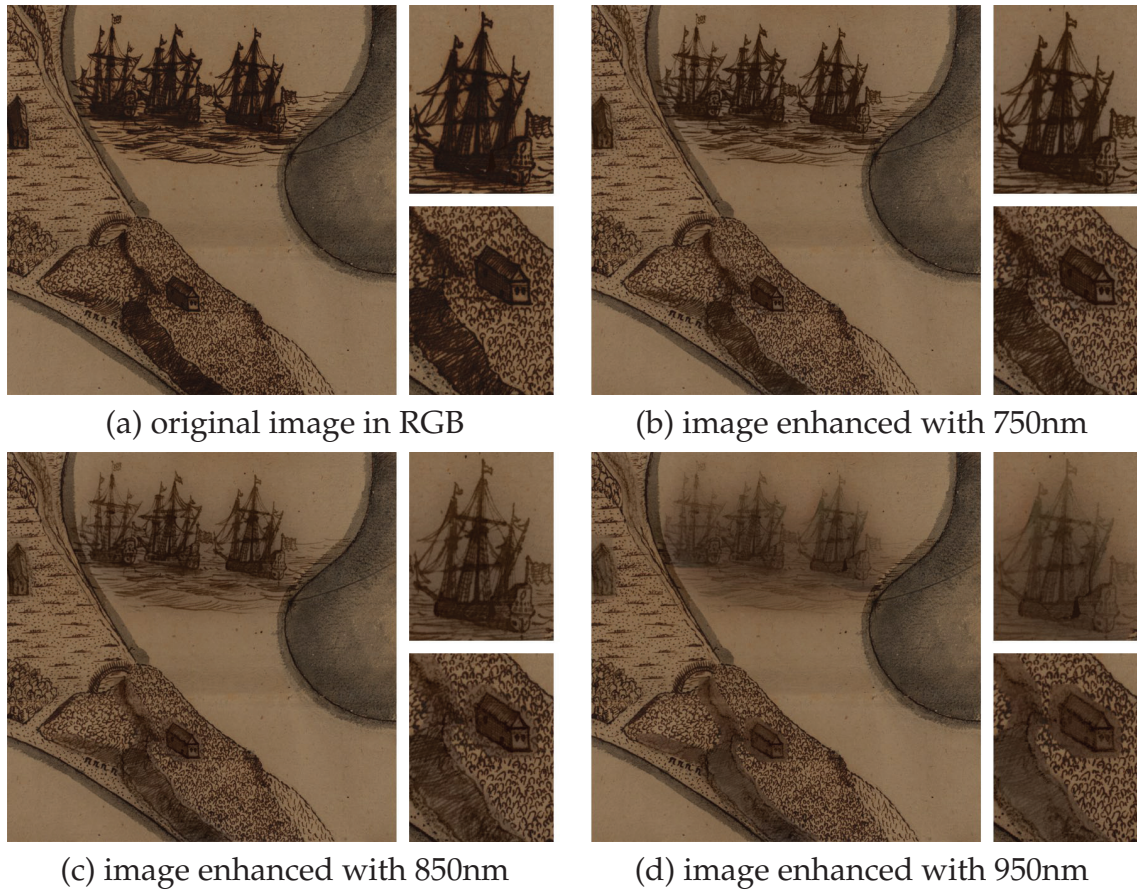
Figure 3.9: Original image with low contrast in some parts (RGB, (a)) is enhanced using images in NIR range. Using just one NIR band does not give satisfactory results since one band does not capture the best contrast for all regions. Hence a scheme for integrating information from all NIR bands is necessary.

in Fig 3.9 (b), the details of the textures on the ships are still not apparent in this image. While the details of the ships are recovered with the image at 850nm (Fig 3.9 (c)), the area with the house is blurred due to low gradients in this region at this wavelength. Both areas are washed out at 950nm, but this image provides the most clear view of the hole in one of the ships (see the closed-up view in Fig 3.9 (d)). Hence, the best strategy for enhancing contrast in documents with line drawings would be to integrate information from all available NIR bands as we proposed in

(a)   (b)
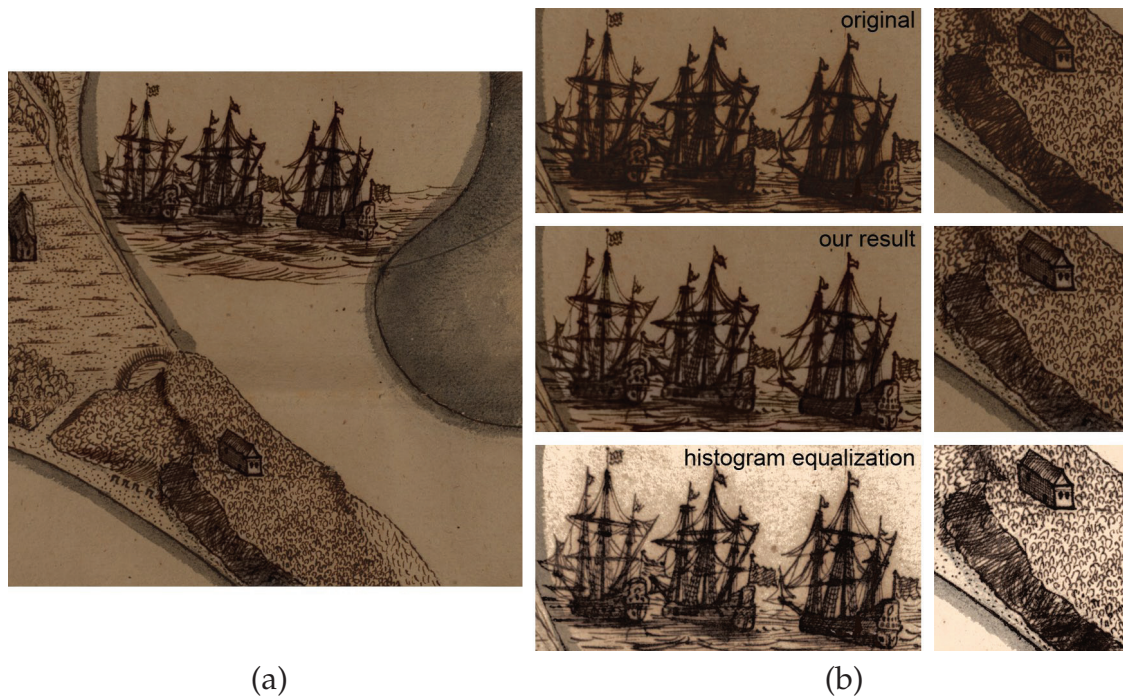
Figure 3.10: (a) The enhancement result using our algorithm. The contrast is greatly enhanced and the details on the ships and on the houses is now recovered. (b) Close-up views of the original RGB image (top), our enhancement result (middle), and histogram equalization result (bottom).

Sec. 3.3.2.

Figure 3.10 shows the result of our contrast enhancement. Details lost in the original RGB image are recovered and all regions are equally enhanced in contrast to the results from using just one band (Figure 3.9). Figure 3.11 shows some of the labeling results by our MRF framework. As can be seen from the labeling map, information from different bands are integrated. For example, lower NIR band images were used to provide strong boundaries of the ships, mid NIR band images were used to provide details inside the ships, and the highest NIR band image was used to distinguish the hole in one of the ships. Our enhancement result is compared with the enhancement using local histogram equalization in Figure 3.10.

(a)

(b)   (c)   (d)
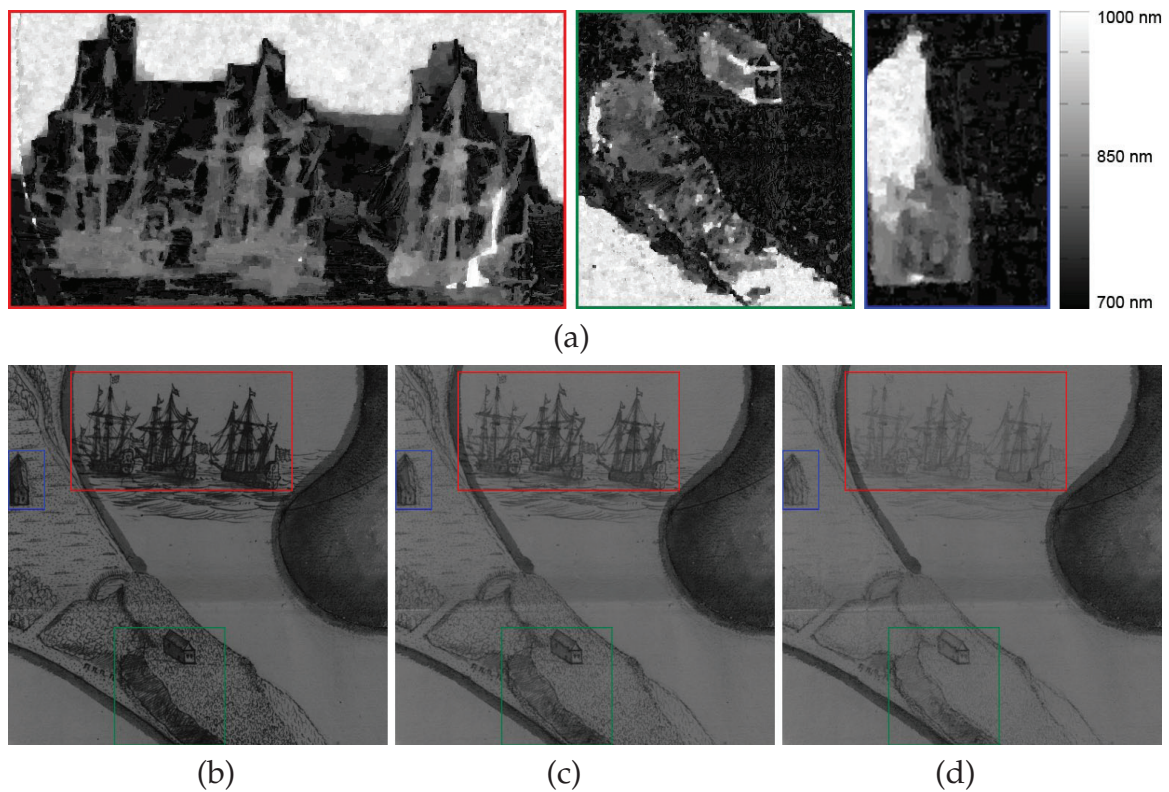
Figure 3.11: Labels for the gradient map composite: (a) labeling of pixels indicates which band image (nm) to use to get gradients, (b) image at 720nm, (c) image at 880nm, (d) image at 1000nm.

While the contrast is enhanced and the details are revealed by applying histogram equalization locally to each region, the noise is also amplified and the original color is lost in the process. We also applied global histogram equalization to the whole image which did not give a satisfactory result.

3.5 Summary

We have described how to take advantage of hyperspectral imaging, most notably using images in near-infrared to assist in the visual enhancement of old documents.

Specifically, we demonstrated how to improve the visual quality of text-based documents corrupted with artifacts such as ink-bleed, ink-corrosion, and foxing, by using the invisible bands to help remove these undesired artifacts. For documents with line drawings that suffer from low contrast, we use the invisible bands to provide more details to enhance legibility. The key components of our framework included detecting regions that can be enhanced by NIR range images, compositing the enhanced gradient map from NIR images, and reconstructing the final image from gradients using an optimization scheme.

Since our visual enhancement framework is subjective in nature, it's quite hard to find an appropriate quantitative metric to measure the improvement of results. However, the feedback from our collaborators at the NAN has been highly positive. Our algorithms have been integrated as part of a comprehensive HSI visualization tool used by the NAN. In particular, they state that this algorithm can be used to produce results that augment physical exhibitions, where a printout of the enhanced version produced by our algorithm is displayed near the original document. The ability to maintain the look and feel of the original document was especially lauded, as previous attempts by themselves was done with less success in Photoshop.

Chapter 4

Reducing Compression Artifacts Arising from Tone Adjustments

In this chapter, we introduce a new gradient domain based compression artifact reduction method that is tailored specifically for tone-adjustment. As far as we are aware this is the first approach to consider removing compression artifacts that arise due to the application of image enhancement. As discussed in Chapter 1, such gradient domain processing is task specific. In this chapter we describe how we implement our method through a combination of several state-of-the-art techniques including dictionary and example-based learning and corrupted regions detection. The experimental results show that the proposed method can effectively suppress the compression artifact.

4.1 Introduction

One of the most common image-editing task is simple tone adjustment that is used to modify the contrast of an image. This is performed by either manually

specifying or selecting a pre-defined tone-curve T, or via histogram equalization which computes T based on the input image's histogram. While tone-adjustment can help bring out details in an image, it can also amplify noises and artifacts due to compression.

This is a real problem because the vast majority of image are compressed, most often using JPEG compression. The JPEG compression scheme breaks an input image into 8×8 pixel blocks and applies a discrete cosine transformation (DCT) to each block individually. To reduce storage space the DCT coefficients are quantized at various levels – more quantization gives higher compression but lowers image quality (for more details see [77]). JPEG assigns a quality factor, QX, to link the amount of quantization with subjective quality, e.g. Q100-Q85 are considered high-quality images, while Q85-75 are mid-quality, and so on. Because of the 8×8 processing strategy used in JPEG, lower-quality images often exhibit a characteristic blocking effect.

Early JPEG compression methods used fixed quantization tables for different quality settings, however, most JPEG schemes now use what is referred to as *optimized JPEG* where quantization tables are customized based on the image's content [76]. This allows relatively high compression rates, e.g. Q75, with little noticeable visual artifacts. However, when post-processing techniques like tone-mapping or histogram equalization are applied, the blocking effects become more prominently visible as shown in Figure 4.1-(A), since the tone-adjustment operation also amplifies the sharp edges of blocking artifacts.

To address this issue, we introduce a new method to reduce the blocking artifacts that arise from tone-adjustment. Our approach first determines regions in the tone-adjusted image that exhibit prominent blocking artifacts based on the histogram
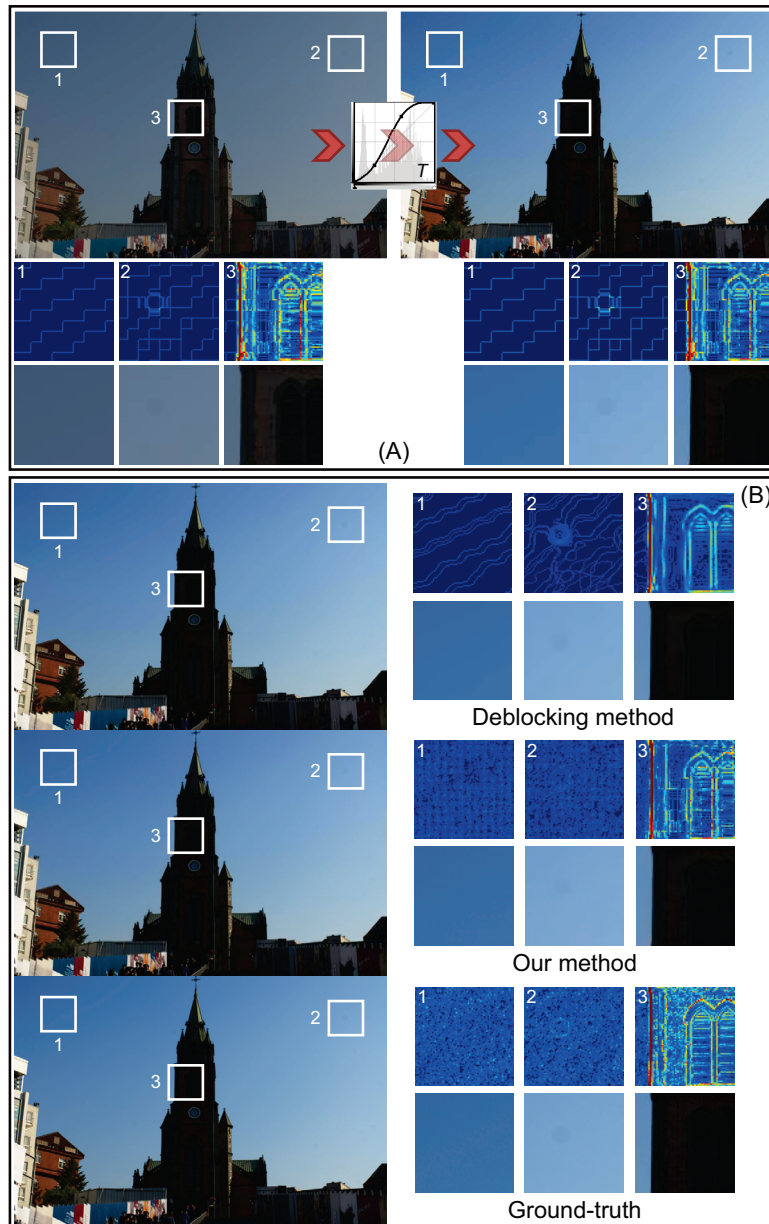
Figure 4.1: (A) Example of noticeable artifacts appearing after tone-mapping. The insets show some selected patches in various regions and their underlying gradient. The characteristic blocking artifact is distinctive in the gradient image. (B) Shows a comparison of our result and the one obtained by state-of-the-art deblocking [26]. Also shown is the ground-truth. Note that our method produces image gradient that better resembles the ground-truth. (Please see the electronic version for better visualization.)

of oriented gradients (HoG). Next, we use a learning-based method to replace the image gradient in these corrupted regions using a training-set of images to which we have applied the same compression and tone-adjustment too. The final deblocked image is then estimated using gradient-domain reconstruction with the new gradients. As shown in Figure 4.1-(B), this approach can produce compelling results that are better than standard deblocking algorithms. The details of our algorithm as well as comparisons with other methods are discussed in the following sections.

4.2 Related Work

Image artifact/blocking has long been recognized in the image processing community [42, 84]. However, this image enhancement problem remains unsolved and an active area [18, 83, 33, 28, 32]. Varied approaches have been used which can broadly be categorized in four general classes. The first is based on global compensation (spectral or otherwise) for the local blocking effects [48, 50, 34]. The second is based on block level or localized identification of artifact and application of local filter [15, 26, 42, 49, 50, 53]. The third is based on projection onto convex sets. Finally, the fourth are dictionary/prior learning based identification and compensation of artifacts [18, 33, 22, 69]. Such methods often use sparse and/or kernel based representations and perform *apriori* learning of the dictionary entries and the corresponding compensators.

Most of these methods target at improving the value of quality assessment metrics like reducing mean square error (MSE) or an energy/entropy function [69], or increasing peak signal to noise ratio (PSNR), PSNR-B [83] (PSNR with blocking

effects), or structural similarity (SSIM) index. In addition, some of the methods create a blocking artifact mask (succinctly called blocking mask) in order to identify and represent the regions with blocking artifacts [50, 48].

Our proposed approach also uses dictionary learning. However, in our approach we are able to dynamically change the dictionary such that it is learnt specifically for a given tone-curve and compression quality. We also introduce a simple detection step based on HoG to find regions that exhibit blocking artifacts. In Section 4.4, we show that our method is able to improve the visual quality of the image in comparison with other two representative works (they have made their code available on-line), one local filtering based approach [26] and one prior learning based approach [69], even though the quality assessment metrics of our method may not be the best.

4.3 Proposed Method

Figure 4.2 shows a high-level overview of our method. An input image I^c with compression quality QX (e.g. Q1 – Q100) is modified by a tone-mapping function T resulting in a new image I_T^c. Here, c denotes the image as being compressed and T is the tone-mapping function which is either user-supplied or computed from a method such as histogram equalization. Our goal is to reduce the blocking artifacts in I_T^c. Our method builds a dictionary from a small set of uncompressed training images that have been compressed and tone-mapped using the same compression quality QX, and tone-mapping function T. We also build a mask of where blocking artifacts exist in I_T^c based on the HoG feature. A new gradient field is computed using the dictionary and regions of the original I_T^c based on the mask. Our restored
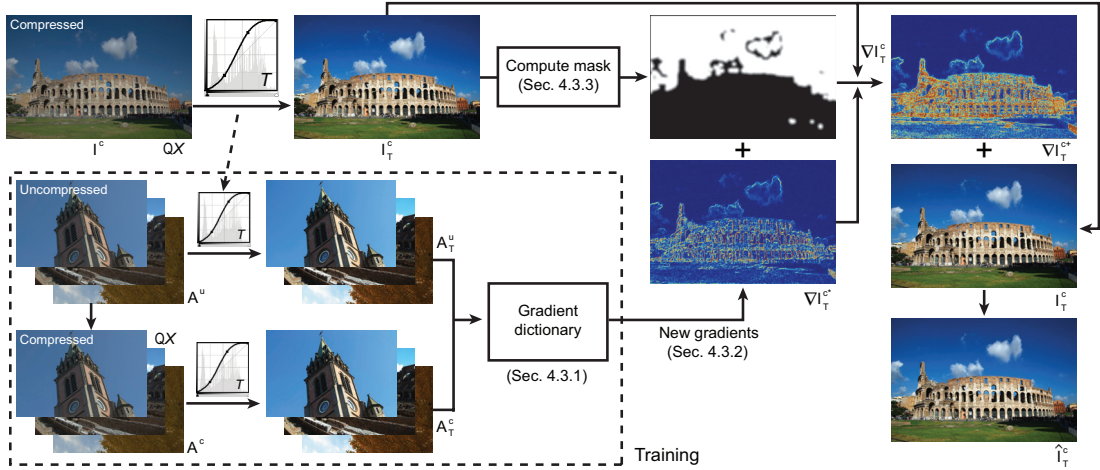
Figure 4.2: A high-level overview of our proposed method. A dictionary is learned from uncompressed and compressed training images which have undergone the same tone-mapping curve as input image. A simple HoG analysis is applied to detect regions with artifacts. Gradients from the learned dictionary are then transferred to replace the gradients in regions with artifacts. A result image is reconstructed from these new gradients to reduce compression artifacts.

image, $\hat{I}_T^c$ is obtained from a gradient-domain reconstruction using the new gradient and tone-mapped image I_T^c. Details are given in the following sections.

4.3.1 Dictionary Construction

We describe here how to construct the dictionary used in our approach. As shown in the bottom-left part of Figure 4.2, our method uses a set of uncompressed training images (35 in total), $\mathbf{A}_i^u$, where u denotes that the image is uncompressed, i.e. high-quality. JPEG compression with quality factor QX is applied to the images by extracting the compression tables directly from the input image I^c. This gives us a set of compressed images $\mathbf{A}_i^c$. The same tone-mapping function T is also applied to the compressed and uncompressed training images resulting in $\mathbf{A}_T^u$ and $\mathbf{A}_T^c$, where the subscript i has been omitted for clarity. Since our approach works in the
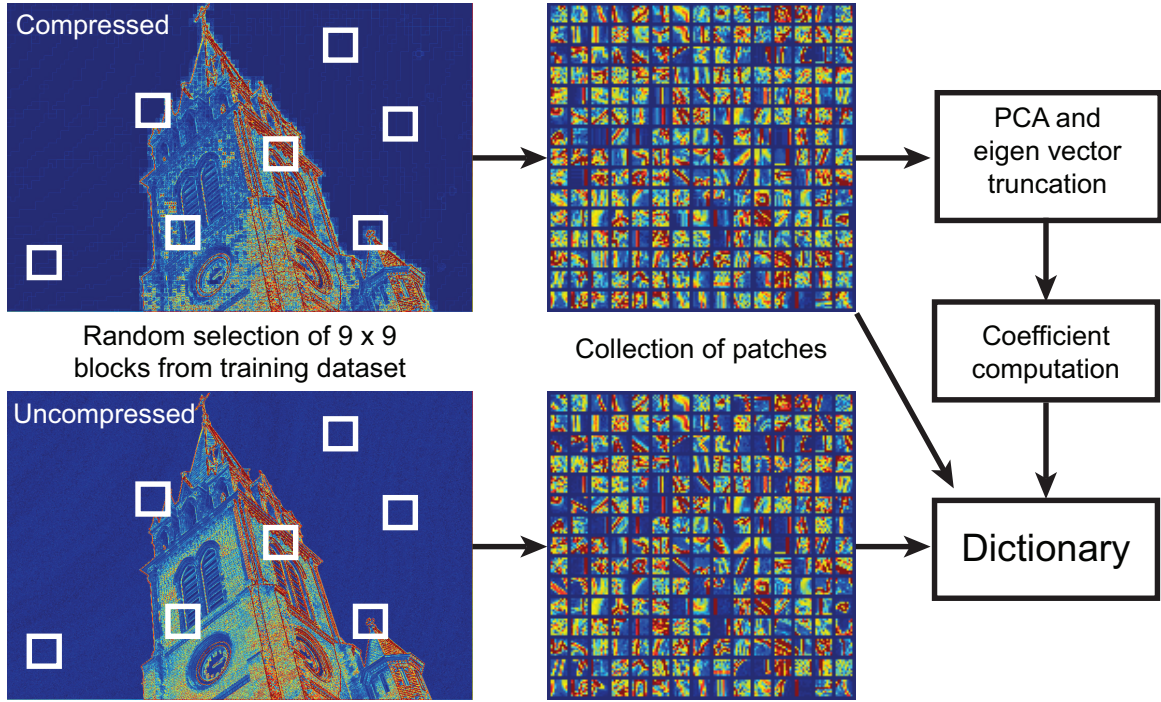
Figure 4.3: This figure shows a diagram of how to construct the dictionary. Image patches are mean-shifted so that the training dataset contains only high frequency structures. We apply PCA to select the top 50K image patches as our learnt dictionary.

gradient domain, our training data is in the form of $\nabla \mathbf{A}_T^u$ and $\nabla \mathbf{A}_T^c$.

Figure 4.3 provides an accompanying diagram of how to construct the dictionary. To reduce the training data to a compact dictionary we randomly select a large number, M, of 9×9 blocks denoted as $\mathbf{B}_i$ from $\nabla \mathbf{A}_T^c$. These blocks are mean shifted such that each block has zero mean $\tilde{\mathbf{B}}$. The patches vectorized $\bar{B}$ compact data matrix D is formed:

$$\mathbf{D} = \hat{\mathbf{B}}\hat{\mathbf{B}}^{\mathbf{T}}; \ \hat{\mathbf{B}} = \left[\begin{array}{cccc} \bar{B}_1 & \bar{B}_2 & \cdots & \bar{B}_M \end{array} \right] \tag{4.1}$$

The PCA vectors of $\mathbf{D}$, $\bar{v}_e$, are computed, where the corresponding eigen values

λ_e are arranged in descending order, i.e. $\lambda_e > \lambda_{e+1}$ [27]. The PCA bases are robust and represent the patches well for large value of M, such that the random sampling ensures that all possible features and their relative occurrence is captured well in the collection of patches. The top E PCA vectors $\bar{v}_e, e = 1$ to E are selected and the PCA coefficients of each patch $\bar{B}$ is computed such as:

$$a_T{}^C = \bar{v}_e \cdot \bar{B}. \tag{4.2}$$

We then select the top $50K$ patches from $\nabla \mathbf{A}_T^U$ and $\nabla \mathbf{A}_T^C$ based on the PCA coefficient magnitude of the $\nabla \mathbf{A}_T^C$ corresponding $a_T{}^C$. We denote these patches as B_i^c and B_i^u respectively. Our approach is applied in the YCbCr space and thus six dictionaries are created, one for each channel and in each gradient direction. The reason for using YCbCr color space is to avoid the color shifting problem that usually occurs when processing images in the color dependent RGB space. We tried several commonly-used color independent spaces like HSV, LAB, YIQ, YCbCr, etc., and found that some of them gave quite similar results. Taking into consideration that JPEG compression is performed in the YCbCr space, we maintain consistency with the JPEG compression pipeline and use the YCbCr color space.

4.3.2 Synthesizing New Gradient

The learnt dictionary is used to synthesize a new gradient field ∇I_T^{c*} to replace the gradient of our tone-mapped compressed image. Our approach is fashioned after learning-based super-resolution methods (see [27]). A schematic overview is provided in Figure 4.4. The gradient of our tone-mapped image ∇I_T^C is divided into 9×9 overlapping patches, denoted as P_x. For each input patch, we first search the

Each test patch

Regularly placed overlapping pathces of size 9 x 9

∇I_T^{c*}

Selection of close matches in the dictionary

MRF with smoothness prior
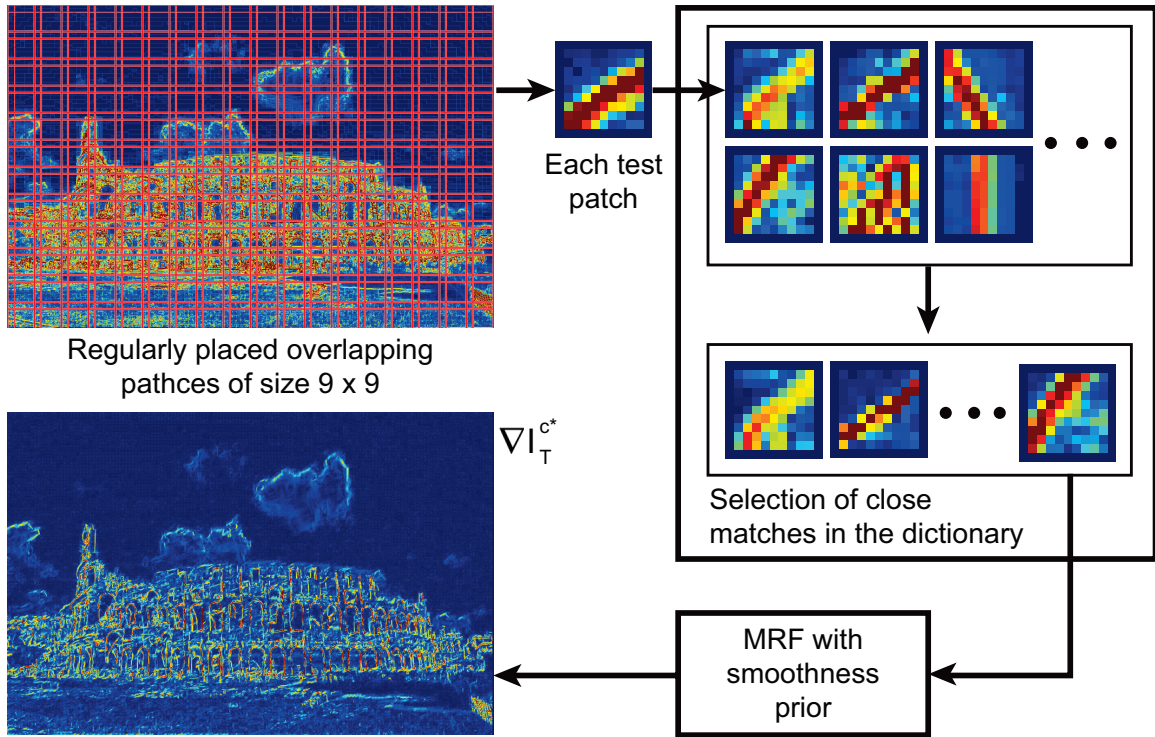
Figure 4.4: This figure shows a diagram of how to synthesize the new gradient field using the learnt dictionary. For each 9×9 patch, we find the top 20 closest matches from dictionary to infer the new gradients. A MRF is used to select the optimal matches based on structural similarity and neighboring connectivity.

dictionary to find the 20 closest matches as candidates, then we try to infer a new gradient patch by finding the most suitable gradient patch B_i^C in the 20 candidates, and using its corresponding artifact-free gradient patch, B_i^U, to replace the gradient in P_x.

To ensure structural similarity, a smoothness prior is imposed. This procedure can be formulated as labeling a Markov Random Field (MRF) with 20 labels (we set the number of labels to 20 by taking into consider both system performance and

running time) in the form:

$$\nabla I_T^{c*} = argmin_x \sum E_d + \sum_{x,y} E_s,$$ (4.3)

where the data cost, E_d of assigning patch B_i to patch P_x defined as $E_d(x) = \|P_x - B_i^c\|_2$ and the smoothness cost, E_s between two overlapping patch assignments is defined as $E_s(x, y) = \|O_{ij}(B_i^u) - O_{ji}(B_j^u)\|_2$, where $O_{ij}(B_i^u)$ and $O_{ji}(B_j^u)$ are the overlapping regions of the uncompressed patches assigned at neighbouring locations x and y.

We estimate the optimal solution to equation 4.3 using graph-cuts [11]. This gives us an index for each patch P_x to a patch in the dictionary B_i^u. The new gradient, ∇I_T^{c*}, is then formed using B_i^u with the overlapped regions being the average over the adjacent patches.

4.3.3   Error Mask

Although the new gradient field helps to find gradient with less blocking artifacts, it may not be able to capture the high resolution features (features with large spatial frequency) very well because of the smoothing performed in the overlap region between the test patches. To this end, we try to estimate the regions in the image that exhibit the most blocking artifacts. We observed that the HoG [16] feature can be used as a good predictor of blocking artifacts. The HoG is computed as a histogram of oriented gradients in a 9×9 window about each pixel and binned into 9 directions. We took several example patches for different tone-curves and compression factors and labeled regions as those exhibiting noticeable blocking artifacts and without noticeable artifacts. The mean HoG for both regions was estimated as shown in Figure 4.5-(B). The HoG for blocking artifacts has distinct

(A) Example patches for training

(B) Plot of mean HoG feature

(C) Probability map

(D) Final smoothed mask
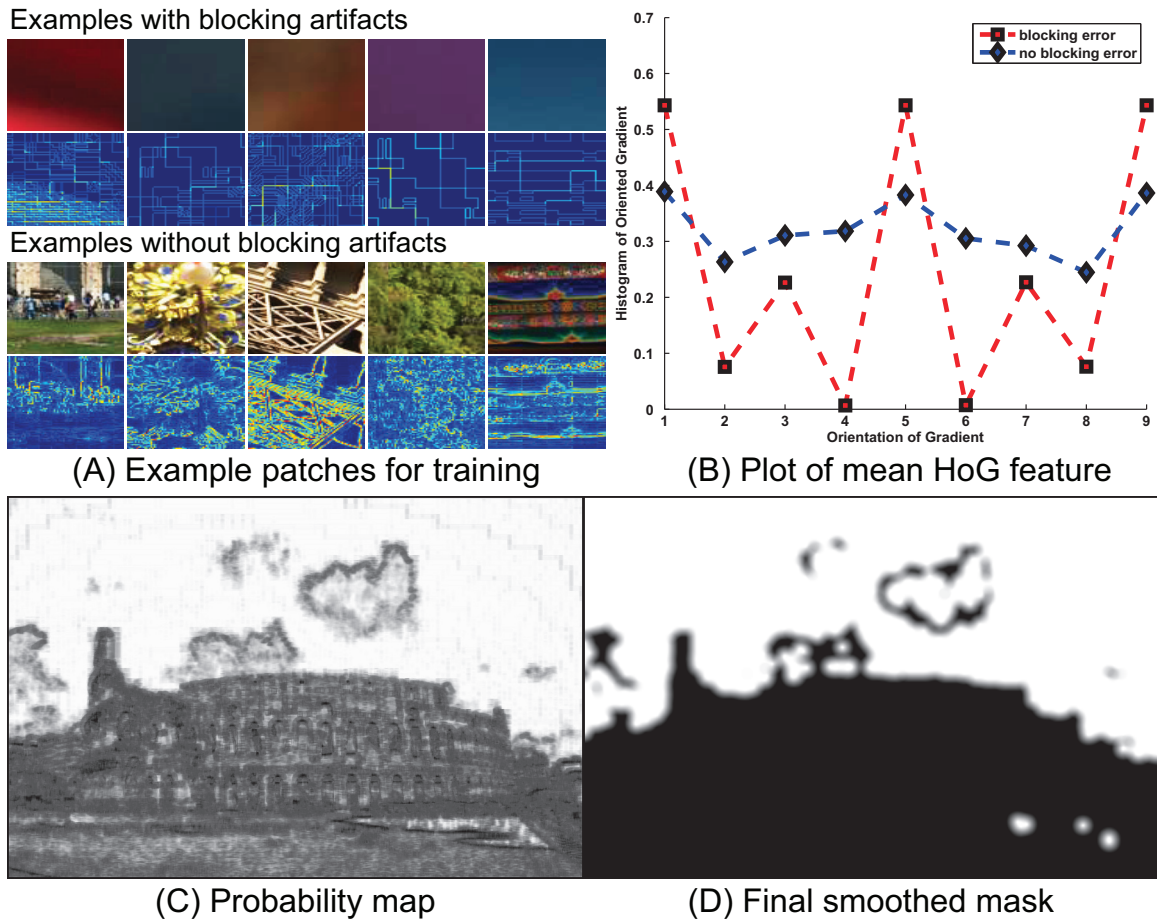
Figure 4.5: (A) some training patches with/without blocking artifacts; (B) the mean response of two types of HoG features: blocking or no blocking; (C) shows the probability of each pixel being the blocking error; (D) shows the final smoothed mask.

peaks at angles $0, \pi/2, \pi, 3\pi/2$ (since blocking artifacts consist of horizontal and vertical edges), while the HoG in other regions is relatively flat.

To estimate the error mask M, we simply compare the HoG on our image I_T^c to the mean blocking artifact HoG feature. A simple threshold t on HoG map is used to get a blocking artifact mask M, which has value 1 for the pixels with HoG distance less than t (i.e. the pixels should be corrected) and zero otherwise. The

map is smoothed slightly with a Gaussian filter. The new gradient field is adjusted based on this error map such that gradient in the regions not considered erroneous are not changed. This can be expressed as: $\nabla I_T^{c+} = M \nabla I_T^{c*} + (1 - M) \nabla I_T^c$.

4.3.4 Image Reconstruction

To reconstruct the final corrected image $\hat{I}_T^c$, we apply a gradient-domain reconstruction, which regularizes the solution against the original tone-mapped input while imposing the new estimated gradient. The formulation can be expressed as:

$$argmin_{\hat{I}_T^c} (\|\hat{I}_T^c - I_T^c\| + \|\nabla \hat{I}_T^c - \nabla I_T^{c+}\|). \tag{4.4}$$

This is a typical optimization problem and can be solved using the solver proposed by Krishnan et al. in their fast Fourier transform based deconvolution approach [41].

4.4 Results

In Figure 4.6 to 4.9, we present the results for four sample images with quality level Q70 and Q75, since these two quality levels are most appropriate for the situation we are dealing with: blocking artifacts are hard to notice before tone-adjustment, but become clearly noticeable after the tone-curve is applied. These figures present the uncompressed tone-mapped image (ground-truth), compressed tone-mapped image (test phase input), results of prior learning based method [69] (referred to as PLB), local filtering based method [26] (referred to as LFB), and the proposed method. In particular, the tone-curve used here is a pre-defined one as shown in Figure 4.2. Note that the compressed tone-mapped image I_T^c is saved using lossless

Figure 4.6: Sample 1 with quality Q75. From top to bottom: intensity image, insets of intensity image, difference map against the ground-truth, gradient image, insets of gradient image. Please see the electronic version for better visualization.

PNG instead of JPEG format to avoid introducing further compression artifact, while both PLB and LFB require the quantization table as an input to assist in estimating the blocking artifact. Hence, we extract the quantization table out of the compressed JPEG image I^c and pass it to these two approaches as the necessary input.

For each sample, we respectively show 1) the intensity image, 2) two rectangular insets (one with significant blocking effect and the other one with rich texture), 3) the difference map between the ground-truth and each other image, 4) the gradient map, and 5) the insets of gradient map. It is seen in the gradient insets of Figure 4.6 to 4.9, that significant bocking effect is present in homogeneous regions (e.g. the sky) of the input image. PLB fails to remove the blocking effect and the severe artifact can be directly identified on both the result image and its gradient map.
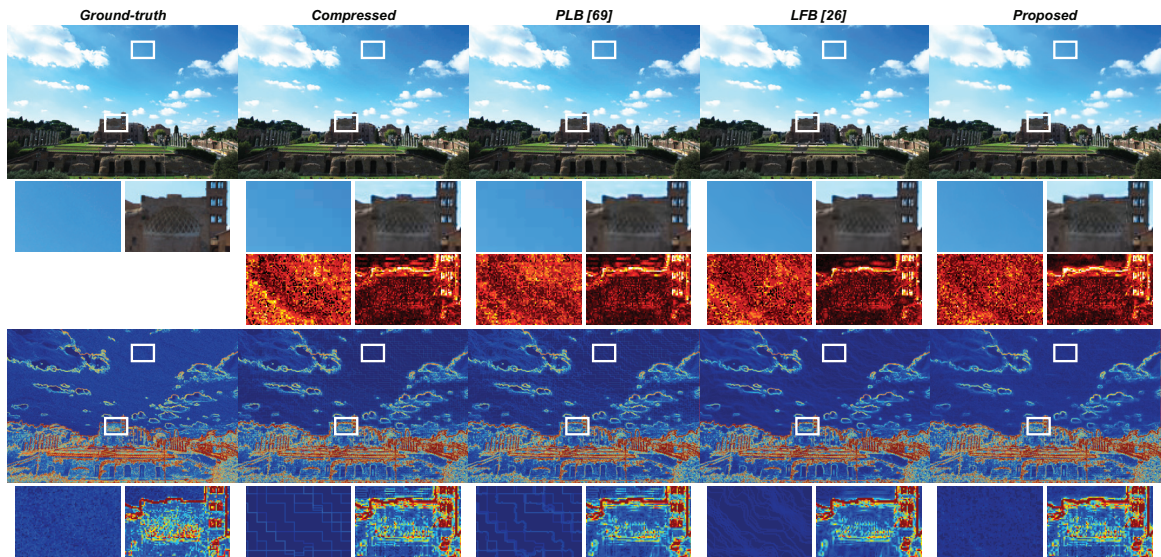
Figure 4.7: Sample 2 with quality Q70. From top to bottom: intensity image, insets of intensity image, difference map against the ground-truth, gradient image, insets of gradient image. Please see the electronic version for better visualization.

While LFB can visually ameliorate or diffuse the blocking effect, it does not solve the blocking issue completely. From the gradient maps, we can clearly see many artificial seams. Further more, both PLB and LFB tend to blur the input image so that some rich texture regions (refer to the bottom right inset) are partly smoothed out. In contrast, the proposed method can completely remove the blocking effect and result into a random gradient patch like the ground truth but differing in magnitude and the actual distribution. The observed deviation from the ground truth is not unexpected since the finer details of the patch present in the high spatial frequency components of the uncompressed images are lost completely after compression and even the dictionary based approach cannot exactly mimic the lost information. On the other hand, the proposed method does not introduce blurry effect to rich texture regions due to the usage of error mask.
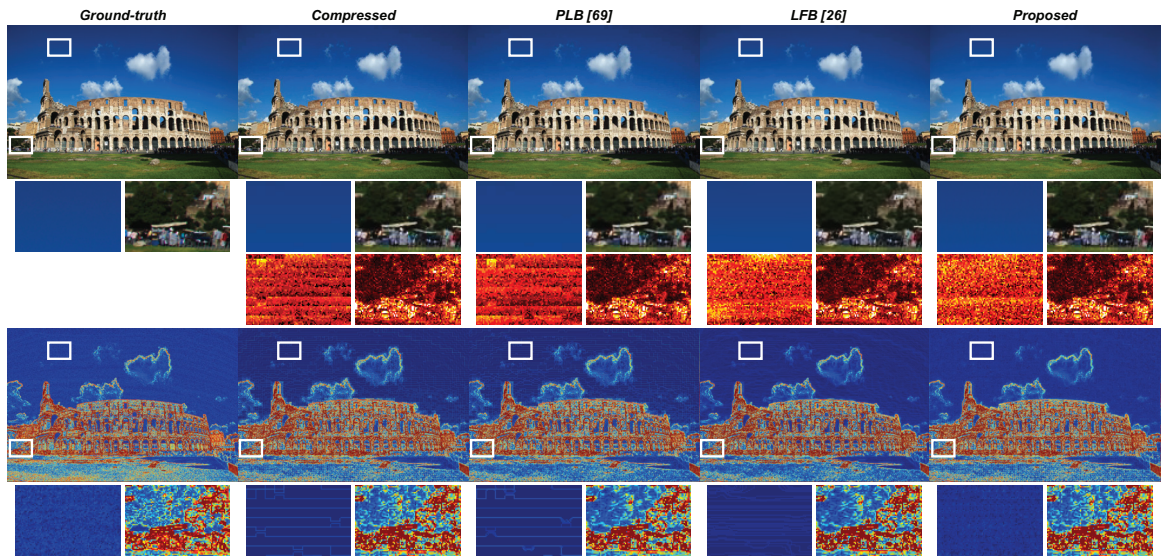
Figure 4.8: Sample 3 with quality Q75. From top to bottom: intensity image, insets of intensity image, difference map against the ground-truth, gradient image, insets of gradient image. Please see the electronic version for better visualization.

In stead of synthesizing the new gradient using the learnt dictionary as discussed in Section 4.3.3, we can also simply use the gradient map of compressed image I^c to compose the new target gradient map with the same error mask, since the blocking artifact is relatively unnoticeable (but still exists) in I^c before any tone-adjustment. We refer to this manner as the *simple* method. The visual quality of this method's results is in between LFB and the proposed method. We also did two "pre-correction" experiments using PLB and LFB. That is, instead of deblocking I^c_T, we can firstly deblock I^c and then apply the specific tone-curve to the deblocked result. We find that the pre-correction results are worse than results of the post-correction manner. They are smoothed out a lot, resulting in the distinct loss of fine details.

In addition to the qualitative comparison, the quantitative comparison is also
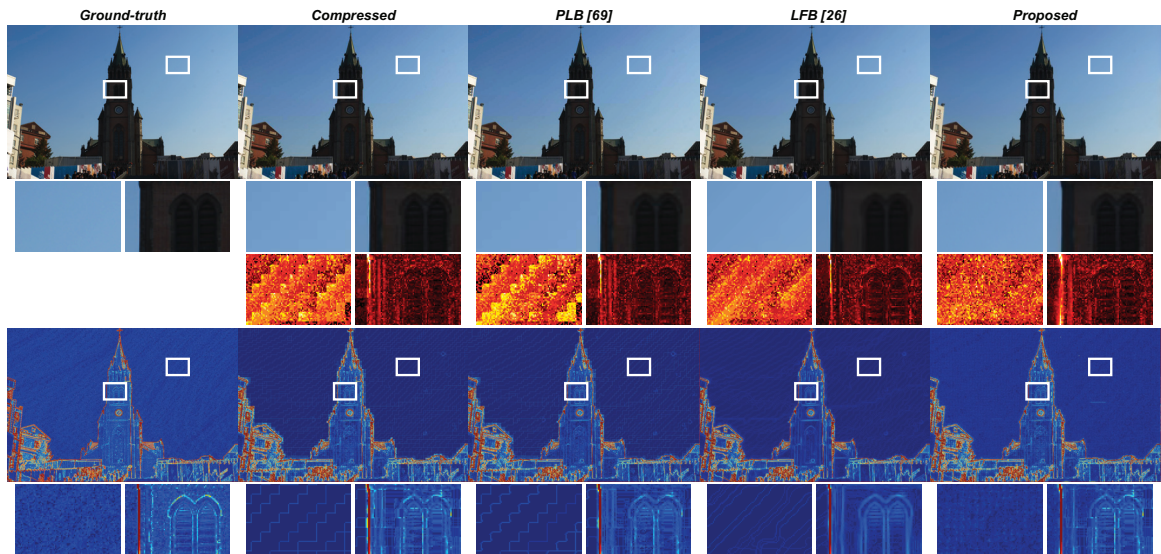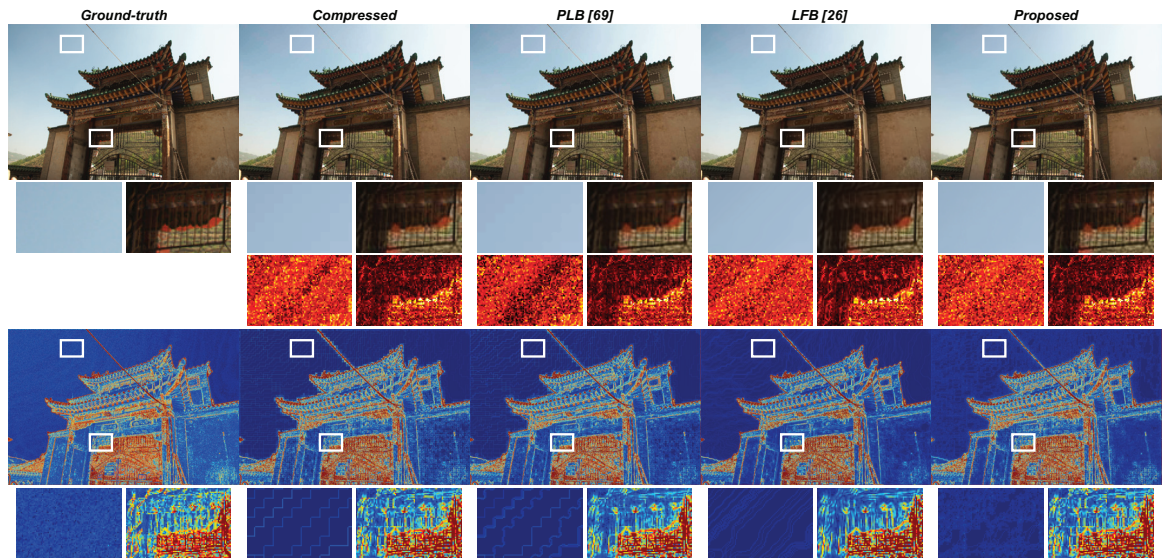
Figure 4.9: Sample 4 with quality Q70. From top to bottom: intensity image, insets of intensity image, difference map against the ground-truth, gradient image, insets of gradient image. Please see the electronic version for better visualization.

presented in Table 4.1. We select three commonly used quality assessment metrics, PSNR/MSE/SSIM, to evaluate the results of PLB, LFB, the simple method, and the proposed method. From Table 1 we can see that the quality assessment metrics of our method may not the best but comparable with other methods. However, the visual quality of our results is clearly better than other methods. The purpose of showing this quantitative comparison is to state that quality assessment metrics are not always reliable when they are used for ranking the image's visual quality.

Lastly, we performed a simple user study on user's preference of the results generated by 6 different methods (PLB, PLB-pre, LFB, LFB-pre, the simple method, and the proposed method). Thirty participants with average age around 27 were involved and not trained before the experiment. For each participant, we randomly select 7 examples and show our result against one of the other 5 results side by

	Metrics	Fig. 4.6	Fig. 4.7	Fig. 4.8	Fig. 4.9
PSNR	PLB	36.0883	33.6842	31.9571	**33.0604**
	LFB	**36.5471**	**33.7465**	**32.1644**	32.9419
	Simple	35.8764	32.9835	31.7029	32.5842
	Proposed	35.3928	32.7774	31.6734	32.4192
MSE	PLB	16.0050	27.8397	41.4349	**32.1395**
	LFB	**14.4004**	**27.4430**	**39.5044**	33.0285
	Simple	16.8049	32.7135	43.9326	35.8638
	Proposed	18.7845	34.3039	44.2321	37.2526
SSIM	PLB	0.9794	0.9738	0.9666	0.9657
	LFB	**0.9859**	**0.9760**	**0.9750**	**0.9682**
	Simple	0.9838	0.9686	0.9721	0.9661
	Proposed	0.9830	0.9702	0.9724	0.9660

Table 4.1: Quantitative evaluation of PLB [69], LFB [26], the simple method and the proposed method.

side. That means the participant needs to do 35 comparisons in total and decide which image (left side or right side) is visually better based on two criteria: 1) less blocking artifacts in homogeneous regions and 2) less blurry effect in rich texture regions. The order of examples, methods that we are comparing against, and the image's left and right positions are totally random. Lastly, we count the number of all user's choices for each method (total number is $30 \times 35 = 1050$) and plot the result in Figure 4.10. Note that we divide the count value of our method by 5 since our result appears 5 times more than other methods. The statistical result shows that our proposed method is preferred by most of users.

4.5 Conclusion

This chapter presents an approach to correct the blocking artifacts that appear in compressed images and become visibly annoying after tone-mapping operation.
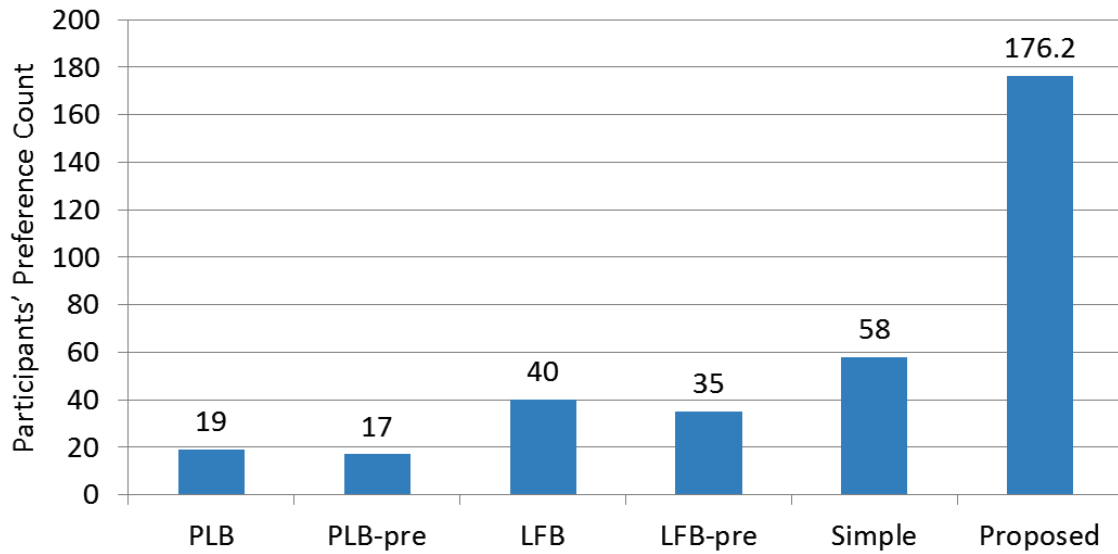
Figure 4.10: Participants preferred results of 6 different methods. The statistical result shows that our proposed method is preferred by most of users. Total number of choices made by users is $1050 = 19 + 17 + 40 + 35 + 58 + 176.2 \times 5$.

The method combines the strengths of several state-of-the-art techniques like dictionary, example learning, forming blocking artifact mask, and reconstructing the output image by optimizing an energy function. Each of the steps plays a crucial role in enhancing the performance of the proposed method. As a result, the proposed method is significantly more effective in suppressing blocking artifacts and providing more user-pleasing images in comparison to other recent methods. The performance of the method is demonstrated using qualitative and quantitative tools as well as using an extensive user study.

Chapter 5

Color-aware Regularization for Gradient Domain Manipulation

As discussed in Chapter 1, gradient domain manipulation techniques require a reconstruction step to compute the final pixel values from the modified gradient field. This reconstruction can sometimes lead to undesirable colors in the final result. In this chapter, we present a color-aware regularization method for use with gradient domain image manipulation to avoid color shift artifacts. The key idea is to use an anisotropic Mahalanobis distance to control output colors to better fit original distributions. Our approach is simple, easy to implement, and does not introduce significant computational overhead. The effectiveness of our color-aware regularization is evaluated on three gradient domain tasks, gradient transfer, gradient boosting, and saliency sharpening.

Before the detailed discussion of our method, we would like to precisely define the color shift problem by taking gradient boosting as an example. Gradient boosting is used to enhance the contrast of an image (i.e. make a flat image more vivid or colorful), but "color shift" occurs if a new hue that doesn't exist in the

input image shows up in the output image. For example, the flower center in Figure 5.2(C) changes from yellow to green (a completely new hue), resulting in the color shift problem; however, the petal changes from light purple to dark purple, which is not the color shift problem.

5.1 Motivation and Related Work

Gradient domain manipulation is the cornerstone of many image processing algorithms from image editing to texture transfer to image fusion. For an overview of gradient domain algorithms and applications we refer readers to [9]. As the name implies, gradient domain algorithms do not operate in the 0th order domain (i.e. color domain), but instead impose changes to the 1st order derivatives of the input image, i.e. the image gradient. When left unchecked, gradient domain processing can result in noticeable color shifts in the 0th domain output image. To ameliorate color-shifting artifacts, most gradient domain approaches impose an additional 0th order constraint either at the boundary of the processed region (early approaches, e.g. Poisson image editing [59]), or over the entire region (recent approaches, e.g. [9, 85, 40, 82, 79, 67]). Readers may refer to Section 2.2 in Chapter 2 for more information on these two regularization strategies.

Since imposing 0th order constraints over the entire processed region could have more control on the output color, most recent gradient domain processing approaches adopt this regularization strategy. This work also targets this regularization strategy which can be applied in one of two manners, either by 1) first converting the input space (presumably RGB) to a new color space (e.g. YUV or LAB) that separates the luminance and chrominance components and processing
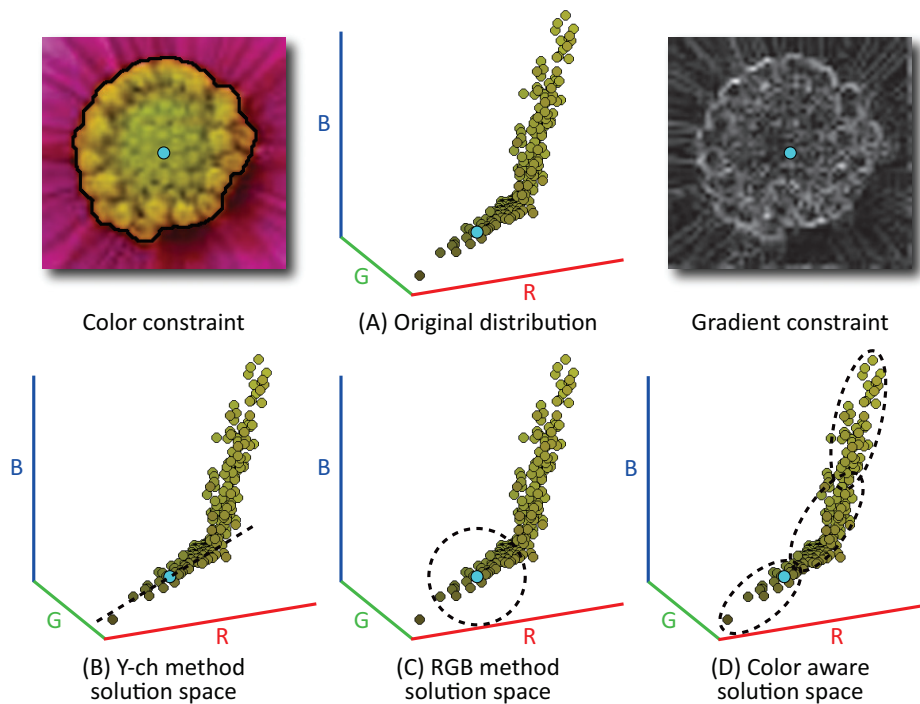
Figure 5.1: Solution spaces (denoted by the dotted line) of the marked pixel using different 0th domain regularization methods.

only the luminance channel (we refer to this manner as Y-ch method in the rest of this chapter); or by 2) applying the L_2 regularization to all three channels separately (we refer to this manner as RGB method). When only one channel is processed, the regularization effectively constrains the output solution so that each pixel is restricted to a 1-D space (Figure 5.1(B)). Although this approach does not shift the chromacity, it can produce outputs that appear flat and less vivid. This can be seen in Figure 5.2(B). When all three channels are processed, the per pixel solution space is constrained to lie within the sphere about its original value as shown in Figure 5.1(C). This conventional regularization is applied irrespective to how the scene colors are distributed in the original input. As a result, satisfying the regularization constraint may also introduce colors that are quite different than those

| Input image | Y-ch method | RGB method | Our color aware method |
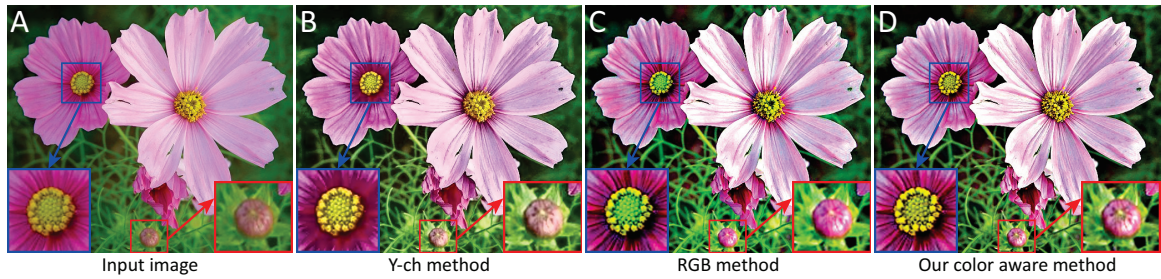
Figure 5.2: This figure compares conventional 0th domain regularization applied to an image that has had its gradient boosted. A) Input image. B) Result using L_2 regularization over the Y channel only. C) Result using L_2 regularization over all three channels of the RGB input. D) Our color-aware regularization result. Note the flat output colors exhibited by Y-ch method in B, and the subtle color-shifting exhibited by RGB method in C.

in the original image. This can be seen in Figure 5.2(C) where the solution of the gradient boosting has resulted in noticeable color shifts.

Our work is motivated by the observation that objects' RGB colors in natural images follow unique distributions. For example, in Figure 5.1(A), the pixel marked in cyan is plotted with all other pixels belonging to the same object. It is easy to see that the pixel belongs to a distinct color distribution in the RGB space. Such unique distributions have been observed by Omer and Werman [56], who showed that colors in natural images tend to form elongated clusters (referred to as lines) in the RGB space. Our color-aware approach constrains the solution space to more tightly follow the original distribution in order to avoid color shifting as shown in Figure 5.1(D) and Figure 5.2(D).

Contribution Our contribution is the introduction of a regularization term that more faithfully follows the distribution of colors in the input image. Our approach applies a simple segmentation to the input image to assign each pixel to a color distribution represented as a Gaussian mixture model (GMM). Based on these GMMs

we can formulate the color-aware regularization using anisotropic Mahalanobis distances [20], which allow the problem to be expressed as a linear system. This color-aware regularization constrains the output solution to better fit the original input color distributions thereby avoiding color shifts. Our approach can be easily incorporated into existing gradient-domain formulations. We demonstrate the effectiveness of this regularization on a variety of inputs for three selected applications, gradient transfer, gradient boosting and saliency sharpening. We compare our results with conventional L_2 regularization approaches (Y-ch method and RGB method) as used by [9, 85, 40, 82, 79, 67].

5.2 Color-aware Regularization Framework

5.2.1 Overview

An overview of our framework is shown in Figure 5.3. Each pixel is first assigned to a color distribution via segmentation. We found that a simple superpixel segmentation [64] followed by k-means clustering [20] is enough to find the underlying color distributions. These individual color distributions are then fit with a series of 3D Gaussian distributions in the RGB color space. The input to our algorithm is an image where each pixel is assigned to a single distribution represented by a series of Gaussians, i.e. $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m$.

A bi-objective function is then used to transfer the new gradients to the input while regularizing each output pixel to lie within a minimum distance from one of the Gaussian distributions used to model its associated color distribution. This regularization is formulated as an optimization problem similar to the work of [9, 85, 40, 82, 79, 67].
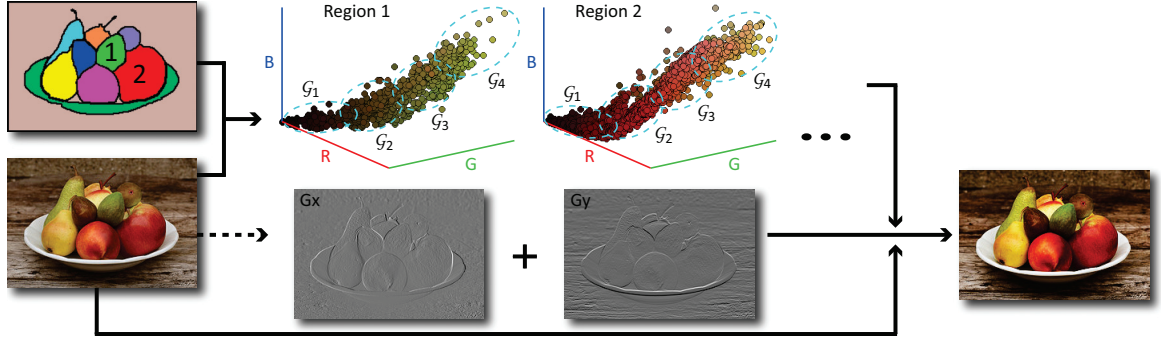
Figure 5.3: The overall workflow of our color-aware regularization framework.

5.2.2 Conventional optimization framework

A detailed review of conventional optimization framework has been presented in Section 2.2.2 in Chapter 2. Here we just briefly present some important equations again. Assume we transform the gradients of image g to the target image f while still preserving the original look-and-feel of the target image. That is, we look for a new image u that minimizes the following bi-objective function:

$$E(u) = \sum_{p \in u} (\lambda E_d(p) + E_s(p)), \tag{5.1}$$

$$E_d(p) = (u_p - f_p)^2, \tag{5.2}$$

$$E_s(p) = \left(\frac{\partial u}{\partial x} - c \cdot \frac{\partial g}{\partial x}\right)^2_p + \left(\frac{\partial u}{\partial y} - c \cdot \frac{\partial g}{\partial y}\right)^2_p, \tag{5.3}$$

where p is the pixel index; λ is used for the balance between E_d and E_s; $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ denote the partial derivative operators in x- and y-direction; c is a scaling factor to control the strength of the target gradient field.

5.2.3   Color-aware regularization term

As shown in Figure 5.1, the solution space of each pixel in the resulting image u is constrained either to lie on a 1-D solution space if only a single channel is processed (Y-ch method), or to lie within a sphere centered at each pixel if applied to all three channels (RGB method). Since the Euclidean distance is blind to the inherent correlation among variables, neither of these methods is able to take into account the color distribution information of the input image f. This can lead to flattened colors or noticeable color shifts in the output image. To solve this problem, we change the conventional L_2 regularization to an anisotropic Mahalanobis distance that more tightly fits the original color distribution. By using the Mahalanobis distance, 0th domain solutions along the shorter axis of each pixel's associated Gaussian model are penalized. This forces the solution to move along the longer axis, thus constraining the solution to lie closer to the original color distribution.

Single Gaussian Model   We first consider the case where we can model a color distribution using a single Gaussian distribution. We define our color-aware 0th order regularization term as:

$$E_{mdd}(p) = (\vec{u}_p - \vec{f}_p)^T S_p^{-1} (\vec{u}_p - \vec{f}_p), \tag{5.4}$$

where p is the pixel index; both $\vec{u}_p$ and $\vec{f}_p$ are the RGB pixel values represented by 3D column vectors; S_p is a 3×3 covariance matrix of the Gaussian that pixel p is assigned to. The term E_{mdd} is the squared Mahalanobis distance, which is a dissimilarity measure between the two vectors $\vec{u}_p$ and $\vec{f}_p$. The benefit of the Mahalanobis distance is that, unlike the Euclidean distance, it considers the correlation of data elements in the vector, in our case the RGB values of the pixels.

Combining Eq. 5.3 and Eq. 5.4 using matrix notation we can write our quadratic form bi-objective cost function as

$$
\begin{aligned}
E(u) \;=\;& \lambda E_{mdd}(u) + E_s(u) \\
=\;& \lambda(u - f)^T \Sigma (u - f) \\
& + (G_x u - c \cdot G_x g)^T (G_x u - c \cdot G_x g) \\
& + (G_y u - c \cdot G_y g)^T (G_y u - c \cdot G_y g),
\end{aligned}
\tag{5.5}
$$

where u, f and g are RGB images reshaped into the column vector form (e.g. $[R_1 G_1 B_1 \dots R_N G_N B_N]^T$); Σ is a 3N×3N (N is the number of pixels) block-diagonal matrix containing the 3×3 inverse covariance matrices of Gaussian models that each pixel is assigned to; the matrices G_x and G_y are discrete forward differentiation operators. Note that gradient constraint g does not necessarily form a 3-channel image since we may transfer gradients of a grayscale image to a color image (see Section 5.3). In that case, image g is extended to an RGB image by copying itself three times. Minimizing Eq. 5.5 amounts to taking its derivative, setting it to zero, and solving for vector u that is uniquely defined as the solution of the linear system:

$$
(\lambda \Sigma + G_x^T G_x + G_y^T G_y)u = \lambda \Sigma f + c \cdot (G_x^T G_x g + G_y^T G_y g).
\tag{5.6}
$$

To solve this linear system, we use the conjugate gradient (CG) method [6] that is also used by [43] and [9]. Note that further improvement can be made to the 1st order term $E_s(u)$ in Eq. 5.5 since the L_2 norm is known to be sensitive to noise and may result in haloing artifacts in the output image. To solve this problem, we add two spatial-varying weights to $E_s(u)$ using the same weighting scheme presented

in [9]:

$$w^x(p) = \left(\left| \frac{\partial f}{\partial x} - \frac{\partial g}{\partial x} \right|_p + 1 \right)^{-\alpha} \tag{5.7}$$

$$w^y(p) = \left(\left| \frac{\partial f}{\partial y} - \frac{\partial g}{\partial y} \right|_p + 1 \right)^{-\alpha} \tag{5.8}$$

where the parameter α (typically $1.2 \leq \alpha \leq 3$) determines the sensitivity of $E_s(u)$ to noise. By using this per-pixel weighting scheme halo artifacts are effectively reduced.

Multiple Gaussian Models Instead of using a single Gaussian model per color distribution, we use several Gaussian models to represent each color distribution more precisely. As shown in Figure 5.3, each pixel is first assigned to a color distribution (region) via segmentation. Each color distribution is represented by a series of 3D Gaussian models $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m$ and each pixel is initially assigned to its nearest Gaussian model $\mathcal{G}_i$ via Eq. 5.4. All pixels in the same region (color distribution) share the same set of Gaussian models and can be reassigned to any Gaussian models within this set when iteratively solving the output image. We integrate this reassignment scheme with the conjugate gradient algorithm and show that it can further decrease the objective cost function (see Figure 5.4).

Assume that we divide the input image into k color distributions and each distribution is represented by m 3D Gaussian models, resulting in $k \times m$ Gaussian models in total; $\mathcal{G}_{i,j}$ is the jth Gaussian model of the ith color distribution ($1 \leq i \leq k; 1 \leq j \leq m$) and $S_{i,j}$ is the covariance matrix of $\mathcal{G}_{i,j}$. The expression $\|A\|_F$ denotes the Frobenius norm of matrix A. We now outline the overall algorithm of our reassignment approach as shown in Algorithm 1.

Convergence Analysis Without using the spatially-varying weights on the 1st

---

Algorithm 1 Gaussian model reassignment

Require: input image f and g, initial assignment map of all pixels ASG (a matrix), maximum number of reassignment operations T, number of CG solver iterations t, small tolerance $\epsilon > 0$

1: $u = f$
2: **for** *reselect* = 1 to T **do**
3:     $u = \text{conjugate\_gradient\_solver}(f, g, u, ASG, t); ASG\_old = ASG;$
4:     **for** $i = 1$ to k **do**
5:         **for all** $p \in Region(i)$ **do**
6:             $j_0 = \underset{j \in [1,m]}{\text{argmin}} \ (\vec{u}_p - \vec{f}_p)^T S_{i,j}^{-1} (\vec{u}_p - \vec{f}_p)$
7:             reassign p to the Gaussian model $\mathcal{G}_{i,j_0}$ (one element of ASG is updated)
8:         **end for**
9:     **end for**
10:     **if** $\|ASG - ASG\_old\|_F < \epsilon$ **then**
11:         break
12:     **end if**
13: **end for**
14: **return** the output image u

---

order constraint term, minimizing the conventional bi-objective cost function reviewed in Section 5.2.2 is known to be a convex problem. Our color-aware optimization framework (using single Gaussian model) differs from the conventional formulation by only introducing a block-diagonal matrix Σ on both sides of the linear system $Au = b$ (see Eq. 5.6). We know that the covariance matrix Σ is positive-semidefinite. As a result, introducing the matrix Σ does not violate the convex property of this optimization problem and a global optimal solution can be achieved.

When using multiple Gaussian models and the reassignment scheme, the convex property remains intact. As shown in Algorithm 1, the reassignment scheme is actually a combination of several independent conjugate gradient solving procedures. After each reassignment step is done, the cost value is guaranteed to be
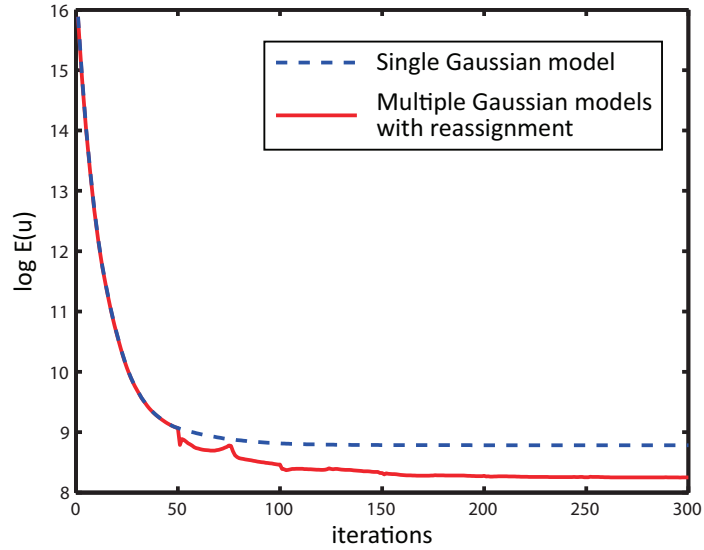
Figure 5.4: Comparison of cost values (with spatial-varying weights applied) when using single Gaussian model (blue dashed line) and multiple Gaussian models (red solid line). For multiple Gaussian models, the reassignment operation is carried out every 50 iterations ($t = 50$ in CG solver).

decreased (or at least remain unchanged) by reassigning each pixel to the Gaussian model whose covariance matrix can minimize the 0th order term E_{mdd} while keeping the 1st order term E_s unchanged.

However, the optimization problem is no longer convex once the spatial-varying weights are used. In this case, the global optimum solution may not be attainable, but we can still use conjugate gradient method to find an approximate solution. In practice, we find our framework works well to minimize Eq. 5.5 within 250 iterations (see Section 5.3.1). Two plots of the cost values during conjugate gradient iterations are shown in Figure 5.4. As we can see, with the help of multiple Gaussian models and the reassignment scheme, the cost value has been further decreased compared to the result achieved by the single Gaussian model. Note that the cost values are shown in *log* scale.

69

5.3 Experiments

We compare the results obtained by our color-aware regularization against those obtained using a conventional optimization framework [9, 85, 40, 82, 79, 67] based on L_2 0th order regularization in the two manners previously discussed (i.e Y-ch method and RGB method). The fast deconvolution algorithm presented by [41] is used to perform the conventional optimization. Comparisons are conducted on three selected tasks including gradient transfer, gradient boosting and saliency sharpening. Before carrying out experiments we briefly explain the parameters we used for these tasks.

5.3.1 Experiment setups

For all the three methods, the gradient scaling factor c is set to 1.0 in gradient transfer task and 2.0 or 3.0 in gradient boosting/saliency sharpening tasks. To keep the comparisons fair, we adjust the balancing factor λ for each method to make sure that a comparable amount of gradient has been transferred or boosted for each example (see quantitative comparison in Section 5.3.3).

For our color-aware regularization method, we use over segmentation algorithm followed by k-means clustering to detect underlying color distributions of an image and k is chosen between 10 and 15 range (see Section 5.3.3 for detail explanation). The number of Gaussian models used to represent each color distribution is fixed to $m = 5$. We restrict the number of Gaussian reassignment operations within 5 times ($T = 5$) and set 50 iterations for the CG solver ($t = 50$). With the above settings, the running time for an 800×600 image is around 3 minutes (Matlab implementation on an Intel Core 2 Duo 2.8GHz computer). We note that using

more than 5 Gaussian models does not significantly improve the results.

5.3.2   Image gradient manipulation tasks

Gradient transfer The first two examples demonstrate gradient transfer of the gradients from a near-infrared (NIR) image to an ordinary RGB image. Such gradient transfer has been demonstrated to improve some forms of photography [40, 88] since NIR often contains more details that cannot be seen in the visible spectrum. In the first example, we show an example of an outdoor scene of a castle where the clouds and other textures are notably stronger in the NIR image. Two input images (NIR and RGB) are shown in Figure 5.5(A-a) and Figure 5.5(A-b). Figure 5.5(A-c) shows the result generated by the Y-ch method. While the desired gradients (clouds) are transferred, the color of the green plants below the castle change to cyan. Figure 5.5(A-d) shows the result produced by the RGB method. Note that the red color of the plants and rocks changes to green. Our result is shown in Figure 5.5(A-e). The colors of both the red rocks and green plants are preserved well. Another example is shown in the second row of Figure 5.5. Note that the color of the nebula (highlighted by a green dashed box) changes significantly in Figure 5.5(B-c) and the color of the stars (highlighted by a red dashed box) is washed out in Figure 5.5(B-d). Our method achieves a better result in Figure 5.5(B-e) with colors that are more similar to the input RGB image.

Gradient boosting The second example targets gradient boosting that is aimed to enhance image contrast. In Figure 5.6, column (a) shows original input images; column (b) shows the scaled gradient magnitudes after boosting (rendered as a *hot map* for better visualization); column (c), (d) and (e) are the results generated by the three different methods. We can see that when using the RGB method (column (d)

(a) NIR image   (b) RGB image   (c) Y-ch method result   (d) RGB method   (e) Our color aware method
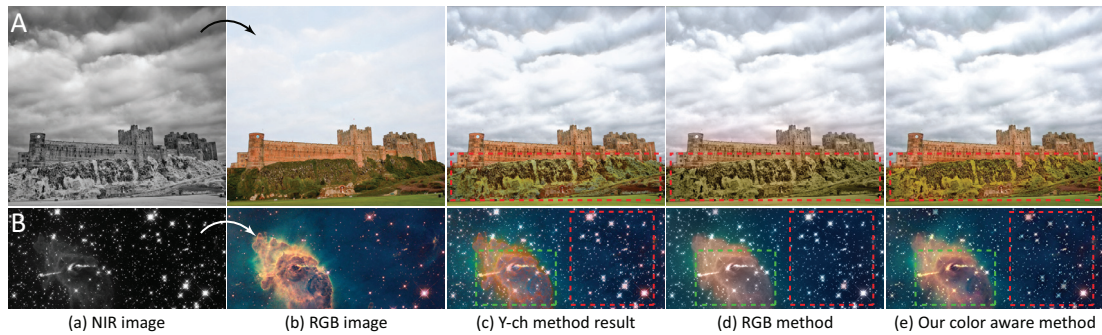
Figure 5.5: Examples of gradient transfer: (a) input NIR image; (b) input RGB image; (c) result using L_2 regularization over the Y channel only; (d) result using L_2 regularization over R/G/B channels; (e) our color-aware regularization result. Regions with color-shifting problem have been highlighted in red and green dashed boxes.

(a) Original image   (b) Scaled gradient map   (c) Y-ch method   (d) RGB method   (e) Our color aware method
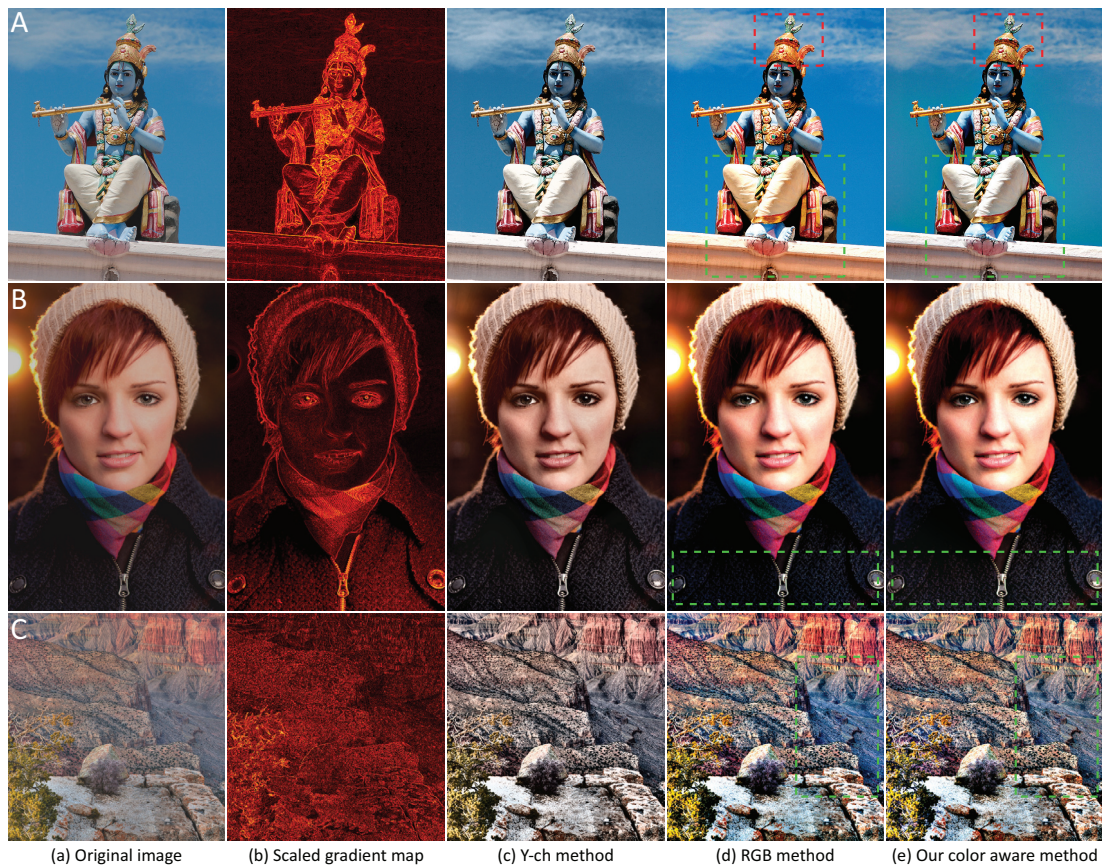
Figure 5.6: Examples of gradient boosting: (a) input RGB image; (b) scaled gradient map providing target gradients; (c) result using L_2 regularization over the Y channel only; (d) result using L_2 regularization over R/G/B channels; (e) our color-aware regularization result. Regions with color-shifting problem have been highlighted in green dashed boxes.

(a) Original image	(b) Saliency map	(c) Y-ch method	(d) RGB method	(e) Our color aware method
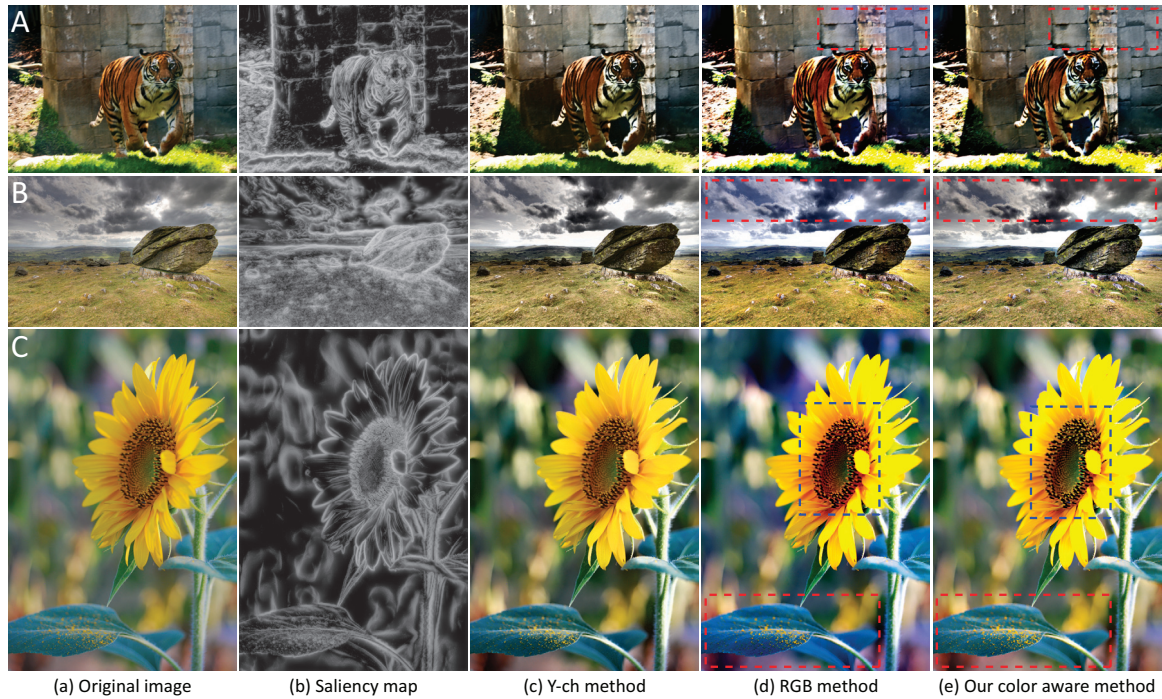
Figure 5.7: Examples of saliency sharpening: (a) input RGB image; (b) saliency map of the input image; (c) result using L_2 regularization over the Y channel only; (d) result using L_2 regularization over R/G/B channels; (e) our color-aware regularization result. Regions with color-shifting problem have been highlighted in red and blue dashed boxes.

in Figure 5.6), the results suffer from noticeable color-shifting in some regions. For instance, the color of the wall and Krishna's legs in example A becomes yellowish; the color of the woman's clothing in example B changes from brown to blue; the color of the valley in example C also shifts to blue. Although color shifts less noticeable when using the Y-ch method (column (c) in Figure 5.6), the overall color of these images seems to be flattened and less vivid. Our results (column (e) in Figure 5.6) show the images with boosted contrast and without color shifts. In addition, our results are more vivid and colorful compared to the Y-ch method.

Saliency sharpening Saliency sharpening is similar to gradient boosting applica-

tion. The only difference is that the gradient boosting globally enhances gradients by a factor c, while saliency sharpening strengthens gradients in a spatially varying manner based on the image saliency map. We adopted the gradient attenuation function proposed in [24] to generate a grayscale saliency map M (brighter regions indicate larger scale factors and stronger boosting). In this case, the global scale factor c in Eq. 5.5 will be replaced by $(1 + c \cdot M)$. As shown in Figure 5.7, our method produces results visually more appealing compared to the other two methods. Note the visible color-shifting on the wall behind the tiger (Figure 5.7(A-d)), the cloud above the rock (Figure 5.7(B-d)) and the sunflower (Figure 5.7(C-d)). Again, results from the Y-ch method (column (c) in Figure 5.7) appear flat, similar to the examples in gradient boosting application. However, our results (column (e) in Figure 5.7) successfully preserve the original color of input images after saliency sharpening process.

5.3.3 Evaluation and analysis

In order to show how our color-aware regularization method preserves the original color distribution more faithfully than the other two methods, we plot the original color distribution of a selected region in the input image and compare it with color distributions of the same region in three output images. In Figure 5.8, column (a) shows the input image and its color-coded segmentation map; column (b) plots the color distribution (data points are randomly sub-sampled for better visualization) of the selected region in the RGB space; column (c), (d) and (e) plot different results generated by the Y-ch method, the RGB method and our method respectively. The plots show that the color distribution of our output image (selected regions) preserves the original distribution much more faithfully than the other two in terms

(a) Input RGB image and its segmentation map   (b) Original distribution   (c) Y-ch method   (d) RGB method   (e) Our color aware mehtod
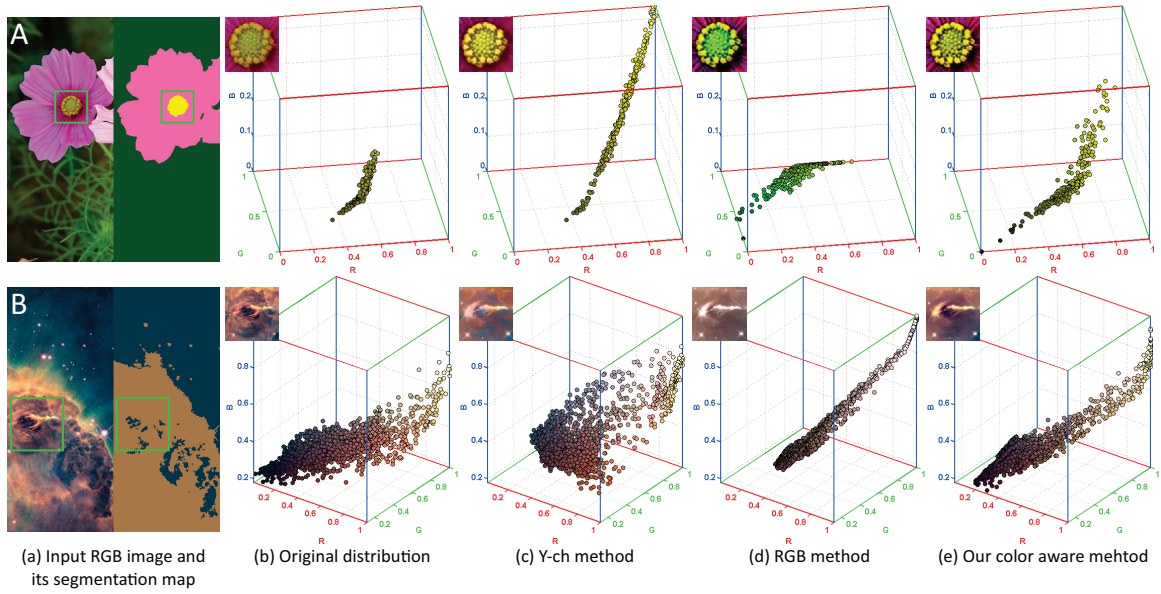
Figure 5.8: Distributions of the solutions using different 0th domain regularization methods: (a) input RGB image and its segmentation map; (b) original color distribution of the selected region (highlighted in green solid boxes); (c) resulting distribution using L_2 regularization over the Y channel only; (d) resulting distribution using L_2 regularization over R/G/B channels; (e) our color-aware regularization distribution. Note that our distribution better maintains the shape and trend of the original.

of shape and trend.

Other than YUV and RGB spaces, we also compared our method with traditional L_2 regularization applied on other commonly-used color spaces. Similar to the Y-ch method, we convert the input image into YIQ/HSV color space and regularize the luminance/brightness channel only. As shown in Figure 5.9(b, c), the results are similar to that of the Y-ch method and also suffer from flattened colors due to the limitation that the output pixels are restricted to a 1-D space (refer to Figure 5.1(B)). Similar to the RGB method, we convert the input image into LAB color space and regularize three channels separately. Using LAB color space we

(a) Original image   (b) YIQ space   (c) HSV space   (d) LAB space   (e) Our method
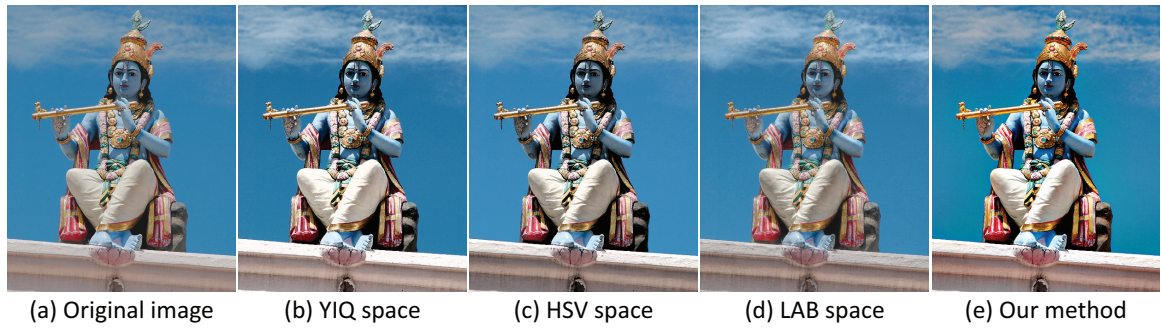
Figure 5.9: Comparison of other color spaces: (a) input RGB images; (b), (c) gradient boosted results using L_2 regularization over the luminance or brightness channel of YIQ/HSV color spaces; (d) result of L_2 regularization over all channels of LAB color space; (e) our color-aware regularization result.

get the result (Figure 5.9(d)) that also appears flat and less colorful compared to our result (Figure 5.9(e)).

We also want to examine the amount of gradient effectively transferred by each method. To do so, we compare the average per-pixel Euclidean distance of the gradient maps of three output images with the constrained gradient map. Table 5.1 lists the amount of gradient transferred for each example. Note that all methods transfer a comparable amount of gradient. This verifies that 1) our approach is able to transfer gradient as effective as the other methods; and 2) the results shown are fairly compared because they have each transferred approximately the same amount of gradient.

Our color distributions are determined by over-segmentation followed by k-means clustering, resulting in k distributions, each of which is further decomposed into GMMs. To determine the sensitivity of our results to the choice of k, we performed experiments with k ranging from 5 to 40. We found the results do not vary too much for values of k greater than 15 (see accompanying supplemental

Table 5.1: This table shows the overall amount of gradient transferred by each method (average L_2 difference between output and input gradients) is similar for all example images shown in Figure 5.5(A, B), Figure 5.6(A, B, C) and Figure 5.7(A, B, C).

Methods	Figure 5.5		Figure 5.6			Figure 5.7		
	A	B	A	B	C	A	B	C
Y-ch method	0.0040	0.0047	0.0369	0.0148	0.0899	0.0849	0.1344	0.0518
RGB method	0.0041	0.0036	0.0372	0.0123	0.0591	0.0757	0.1244	0.0467
Our method	0.0044	0.0046	0.0340	0.0113	0.0533	0.0747	0.1182	0.0453

materials). As a result, we advocate using the range between 10 and 15.

Lastly, since our approach is subjective in nature, we performed a simple user study on user's preference of the results on 14 examples (3 for gradient transfer, 7 for gradient boosting and 4 for saliency sharpening). Twenty participants (average age around 25) were asked to choose their preferred results out of the outputs of the three different methods. Participants were not trained before the experiment, but over half of them had experience with image editing software such as Photoshop. Our user study showed that 18 participants preferred our results for the gradient transfer application, and 15 participants preferred our results for the gradient boosting application. For saliency sharpening application, 16 participants preferred the results produced by our color-aware regularization method. Figure 5.10 shows a graph of these results.

5.4 Summary

We have presented a straight-forward approach to perform 0th domain regularization in a manner that more faithfully follows the original input color distribution. This results in gradient transfer that avoids color shifting while still producing
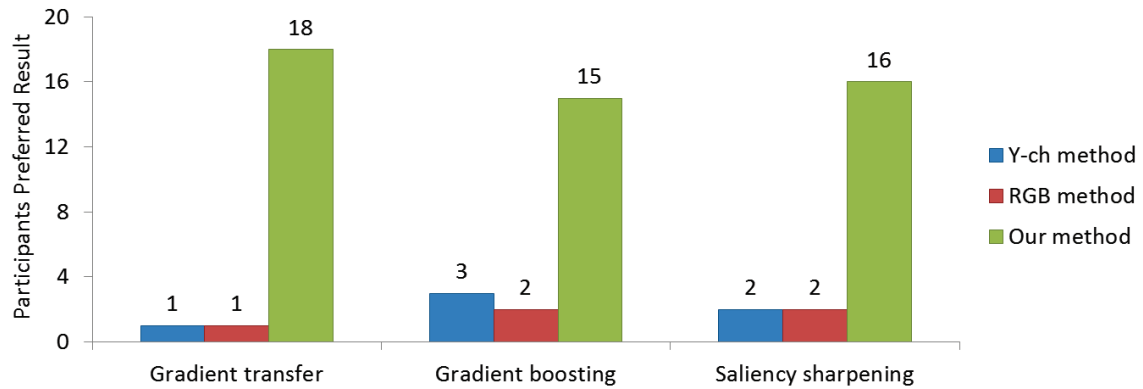
Figure 5.10: Participants preferred results of three different methods.

vivid results. While our approach requires an initial segmentation to determine the distinct color distributions in the image, we found that the segmentation stage is not a crucial issue and any basic over segmentation algorithm (e.g. watershed [75] or superpixel [64]) gave good results. More sophisticated segmentation algorithm like Ridge-based Distribution Analysis [73] were tried but generated similar results. We also note that our approach is not significantly slower than conventional techniques and can be easily incorporated into existing image gradient manipulation methods.

Chapter 6

Conclusion

This chapter concludes the dissertation by giving a short assessment of the works presented in previous chapters, including the visual enhancement framework for historical documents, the compression artifact reduction method tailored for tone-adjustment operation, and the color-aware regularization approach for use with gradient domain image manipulation. Limitations and possible future research directions for each work are also discussed.

6.1   Assessment

This dissertation explored gradient domain solutions for two image processing tasks. The general idea behind these works is to firstly manipulate the gradient field of the input image for the sake of enhancing visual appearance or reducing artifacts, and then reconstruct the final image from the modified gradient field. These works are respectively summarized as follows.

We presented a visual enhancement framework for historical documents based on gradient domain fusion technique in Chapter 3. The goal of this work was

to enhance the legibility of drawing-based documents and the visual quality of text-based documents corrupted with ink-bleed/corrosion/foxing artifacts. The enhancement was done in the gradient domain by selecting desired gradients (with more details or less artifacts) from different NIR spectral images and compositing the enhanced gradient field, from which the final image was reconstructed. The experimental results showed that our enhancement framework can significantly improve the visual quality of degraded old documents. The feedback from our collaborators at the Nationaal Archief of the Netherlands (NAN) was highly encouraging. In addition, our framework was integrated as a part of a comprehensive hyperspectral image visualization tool used by NAN.

In Chapter 4, we introduced a new compression artifact reduction method by combining the strengths of several state-of-the-art techniques. We built a gradient dictionary from a small set of uncompressed training images that had been compressed and tone-mapped using the same compression quality and tone-mapping function as those of the input image. With the help of the error mask that indicates corrupted regions in the input image, we used a learning-based method to replace the gradients of those regions by artifact-free gradients retrieved from the gradient dictionary. Finally, we estimated the new image using gradient domain reconstruction with the new composited gradient map. Experimental results and the user study showed that our method can significantly suppress blocking artifacts and provide more user-pleasing results in comparison with other existing methods.

In Chapter 5, we proposed a straight-forward color-aware regularization method to avoid the color shift problem that usually occurs in the reconstruction phase of many gradient domain image manipulation methods. Motivated by the observation that objects' RGB colors in natural images often follow unique distributions

in the RGB space, our approach was designed to perform the regularization in a manner that more faithfully follows the original input color distribution when reconstructing the final output image. This was achieved by using an anisotropic Mahalanobis distance as the regularization term in the objective function. Our color-aware regularization is simple, easy to implement, and does not introduce significant computational overhead compared to conventional regularization method. The effectiveness of our method was shown by various input images tested on three gradient domain tasks: gradient transfer, gradient boosting, and saliency sharpening.

6.2   Limitations

In this section, we provide a discussion on limitations for each work presented in this dissertation. They are summarized as follows:

Visual enhancement of old documents One limitation of this work is the accuracy of segmenting the document artifacts from the foreground ink that is less sensitive to thresholding. In some cases such as in Figure 3.8, spectrums of strong ink-bleed and corrosion are very similar to the foreground spectrum, which makes the segmentation results rely greatly on the thresholding. Since choosing an optimal threshold is often challenging, some of the foreground texts may be removed while removing the document artifacts as can be seen in Figure 3.8(c).

Compression artifact reduction In our experimental analysis, we found that the blocking artifact appears not only in the luminance channel, but also in the chrominance components. Thus, we have to build six dictionaries, one for each channel and in each gradient direction. This leads to a large computational overhead when

synthesizing the new gradient field (e.g searching for 20 candidate patches when computing data-costs for the MRF). Another issue is the selection of training images, since the quality of training set may greatly affect the final result. In our experience, a good training set should consist of images that have homogeneous regions (e.g. sky, walls) and rich texture regions. Currently, we used 33 natural images (landscape/portrait, indoor/outdoor images) and two images of Macbeth color chart (with the purpose of providing more color variations), which generated sufficiently good results in our experiment.

Color-aware regularization This work is motivated by the observation that colors of objects in natural images typically follow distinct distributions in the RGB space. As such, we assume that the images we are processing contain reasonable color distributions that could be represented by several 3D Gaussian models (i.e. the elongated clusters). For input images with rather flat colors (or grayscale images), our method may not produce satisfying results, since we may not find enough color distributions to formulate our color-aware regularization term and our method is reduced to the conventional L_2 regularization (using identity matrix as the covariance matrix Σ).

6.3 Future Work

In this section, we discuss several future research directions for the three contributions presented in this dissertation. They are summarized as follows:

Visual enhancement of old documents As discussed in the previous section, the segmentation may affect the enhancement result for some difficult cases. To further improve the enhancement result, we can employ a more sophisticated

user assistance approach as in [52] in the future work. Additionally, we can also consider extracting several spectral bands that are more powerful in distinguishing the foreground rather than using the entire HSI spectrums. This allows us to use only a few bands for similarity analysis, as prior research in the archival domain has established that certain bands are more suitable for various tasks and materials being observed. This selective band strategy can also be used to amplify desired artifacts, such as tears and rips, and for managing future data collection in which only the useful bands may need to be captured.

Compression artifact reduction Given the fact that this work has a large computational overhead, one future task is to speed up the whole processing procedure by using more effective searching and MRF labeling algorithms. In addition, in our current MRF labeling algorithm, we select only gradient patches that match well with neighbors in an overlap region (2-pixel overlapping) to compose smooth boundaries. However, this approach still has boundary artifacts in some cases. To further improve the composed boundary, we can refine the MRF labeling algorithm by selecting the best seam (refer to seam carving [5]) through the boundary region of neighboring patches to remove most artifacts. We can also consider using image matching techniques to select good training images in order to further improve the performance as well as reducing the size of the dictionary needed to perform this task.

Color-aware regularization Currently we demonstrate our color-aware regularization on three gradient domain tasks. A natural extension of this work is to explore more applications of our method and build a comprehensive optimization framework for exploring gradient domain solutions for image and even video processing problems. In addition, the running time of our method is quite limited by

the iterative conjugate gradient solver. A fast GPU implementation may greatly reduce the running time and even make it possible for real-time video processing. At last, in some cases, the color distribution of the input image may be very complicated and hard to be modelled by simple 3D Gaussian models. We can consider to find a better way to model the color distributions instead of just Gaussian clusters.

Bibliography

[1] A. Agrawal, R. Raskar, S.K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005. 1.1, 3.2

[2] F. Alter, S. Durand, and J. Froment. Adapted total variation for artifact free decompression of jpeg images. *Journal of Mathematical Imaging and Vision*, 23(2):199–211, 2005. 1.2

[3] F. Alter, S.Y. Durand, and J. Froment. Deblocking dct-based compressed images with weighted total variation. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 221–224, 2004. 1.2

[4] F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183, 1954. 1.1

[5] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, 26(3):10, 2007. 6.3

[6] M. Avriel. *Nonlinear programming: Analysis and methods*. Dover Publishing, 2003. 2.2.2, 5.2.3

[7] P.G.J. Barten. Contrast sensitivity of the human eye and its effects on image quality. In *International Society for Optical Engineering*, 1999. 1.1

[8] E. Bennett, J.L. Mason, and L. McMillan. Multispectral bilateral video fusion. *IEEE Transactions on Image Processing*, 16(5):1185–1194, 2007. 3.2

[9] P. Bhat, C.L. Zitnick, M. Cohen, and B. Curless. Gradientshop: a gradient-domain optimization framework for image and video filtering. *ACM Transactions on Graphics*, 29(2):1–14, 2010. 1.1, 2.2.2, 5.1, 5.1, 5.2.1, 5.2.3, 5.3

[10] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004. 3.3.2

[11] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001. 4.3.2

[12] C.H. Chen and P.P. Ho. Statistical pattern recognition in remote sensing. *Pattern Recognition*, 41(9):2731–2741, 2008. 3.2

[13] C. Collet, M. Louys, A. Oberto, and C. Bot. Markov model for multispectral image analysis: application to small magellanic cloud segmentation. In *Proceedings of International Conference on Image Processing*, pages 953–956, 2003. 3.1

[14] P. Cotte and D. Dupraz. Spectral imaging of leonardo da vinci's mona lisa:an authentic smile at 1523 dpi with additional infrared dat. In *Proceedings of IS&T Archiving Conference*, pages 228–235, 2006. 3.1

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 4.2

[16] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 886–893, 2005. 4.3.3

[17] F.B. Deng, S.J. Kim, Y.W. Tai, and M.S. Brown. Color-aware regularization for gradient domain image manipulation. In *Asian Conference on Computer Vision*, 2012. 1.3

[18] W. Dong, L. Zhang, and G. Shi. Centralized sparse representation for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1259–1266, 2011. 4.2

[19] Q. Du and C.I. Chang. A linear constrained distance-based discriminant analysis for hyperspectral image classification. *Pattern Recognition*, 34(2):361–373, 2001. 3.1

[20] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*, volume 2. Wiley New York, 2001. 1.3, 5.1, 5.2.1

[21] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*, 23(3):673–678, 2004. 3.2

[22] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, 2006. 4.2

[23] G. Farinella and S. Battiato. On the application of structured sparse model selection to jpeg compressed images. *Computational Color Imaging*, pages 137–151, 2011. 1.2

[24] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, 2002. 1.1, 2.1.1, 2.2.1, 3.1, 5.3.2

[25] G. Finlayson, S. Hordley, and M. Drew. Removing shadows from images. In *Proceedings of European Conference on Computer Vision*, pages 129–132. Springer, 2006. 1.1

[26] A. Foi, V. Katkovnik, and K. Egiazarian. Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images. *IEEE Transactions on Image Processing*, 16(5):1395–1411, 2007. (document), 1.2, 4.1, 4.2, 4.4, 4.1

[27] W.T. Freeman and C. Liu. *Markov Random Fields for Super-resolution and Texture Synthesis*, chapter 10, pages 155–166. MIT Pres, 2011. 4.3.1, 4.3.2

[28] O. Gallo, N. Gelfandz, W.C. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *IEEE International Conference on Computational Photography*, pages 1–7, 2009. 4.2

[29] A. Gothandaraman, R.T. Whitaker, and J. Gregor. Total variation for the removal of blocking effects in dct based encoding. In *Proceedings of International Conference on Image Processing*, volume 2, pages 455–458, 2001. 1.2

[30] Y. Huang, M.S. Brown, and D. Xu. A framework for reducing ink-bleed in old documents. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2008. 3.4

[31] J. Jia, J. Sun, C.K. Tang, and H.Y. Shum. Drag-and-drop pasting. *ACM Transactions on Graphics*, 25(3):631–637, 2006. 1.1, 2.2.1

[32] N. Joshi and M.F. Cohen. Seeing mt. rainier: Lucky imaging for multi-image denoising, sharpening, and haze removal. In *IEEE International Conference on Computational Photography*, pages 1–8, 2010. 4.2

[33] C. Jung, L. Jiao, H. Qi, and T. Sun. Image deblocking via sparse representation. *Signal Processing: Image Communication*, 27(6):663–677, 2012. 4.2

[34] V. Katkovnik, A. Foi, K. Egiazarian, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *International Journal of Computer Vision*, 86(1):1–32, 2010. 4.2

[35] S.J. Kim, F.B. Deng, and M.S. Brown. Visual enhancement of old documents with hyperspectral imaging. *Pattern Recognition*, 44(7):1461–1469, 2011. 1.3

[36] S.J. Kim, S.J. Zhuo, F.B. Deng, C.W. Fu, and M.S. Brown. Interactive visualization of hyperspectral images of historical documents. *IEEE Transactions on Visulization and Computer Graphics*, 16(6):1441–1448, 2010. 1.3

[37] Y. Kim, C.S. Park, and S.J. Ko. Fast pocs based post-processing technique for hdtv. *IEEE Transactions on Consumer Electronics*, 49(4):1438–1447, 2003. 1.2

[38] M.E. Klein, B.J. Aalderink, R. Padoan, G. de Bruin, and Th.A.G. Steemers. Quantitative hyperspectral reflectance imaging. *Sensors*, 8(9):5576–5618, 2008. 3.1

[39] S. Kockanat, N. Karaboga, and T. Koza. Image denoising with 2-d fir filter by using artificial bee colony algorithm. In *IEEE International Symposium on Innovations in Intelligent Systems and Applications*, pages 1–4, 2012. 1.1

[40] D. Krishnan and R. Fergus. Dark flash photography. *ACM Transactions on Graphics*, 28(3), 2009. 2.2.2, 3.2, 3.3.3, 3.3.3, 5.1, 5.1, 5.2.1, 5.3, 5.3.2

[41] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. *Advances in Neural Information Processing Systems*, 22:1–9, 2009. 3.3.3, 4.3.4, 5.3

[42] Y.L. Lee, H.C. Kim, and H.W. Park. Blocking effect reduction of jpeg images by signal adaptive filtering. *IEEE Transactions on Image Processing*, 7(2):229–234, 1998. 4.2

[43] A. Levin, R. Fergus, F. Durand, and W.T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, 26(3), 2007. 5.2.3

[44] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(9):1647–1654, 2007. 3.3.3

[45] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image stitching in the gradient domain. In *Proceedings of European Conference on Computer Vision*, pages 377–389, 2004. 1.1, 3.1

[46] H. Li, C.W. Fu, and A.J. Hanson. Visualizing multiwavelength astrophysical data. *IEEE Transactions on Visulization and Computer Graphics*, 14(6):1555–1562, 2008. 3.1

[47] Z. Li and E.J. Delp. Block artifact reduction using a transform-domain markov random field model. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(12):1583–1593, 2005. 1.2

[48] A.W.C. Liew and H. Yan. Blocking artifacts suppression in block-coded images using overcomplete wavelet representation. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(4):450–461, 2004. 4.2

[49] P. List, A. Joch, J. Lainema, G. Bjontegaard, and M. Karczewicz. Adaptive deblocking filter. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(7):614–619, 2003. 4.2

[50] S. Liu and A.C. Bovik. Efficient dct-domain blind measurement and reduction of blocking artifacts. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12):1139–1149, 2002. 4.2

[51] M.A. Loghmari, M.S. Naceur, and M.R. Boussema. A spectral and spatial source separation of multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12):3659–3673, 2006. 3.1

[52] Z. Lu, Z. Wu, and M.S. Brown. Directed assistance for ink-bleed reduction in old documents. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2009. 3.4, 6.3

[53] Y. Luo and R.K. Ward. Removing the blocking artifacts of block-based dct compressed images. *IEEE Transactions on Image Processing*, 12(7):838–842, 2003. 4.2

[54] J. McCann and N.S. Pollard. Real-time gradient-domain painting. *ACM Transactions on Graphics*, 27(3):1–7, 2008. 1.1, 2.2.1

[55] T. Meier, K.N. Ngan, and G. Crebbin. Reduction of blocking artifacts in image and video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(3):490–500, 1999. 1.2

[56] I. Omer and M. Werman. Color lines: Image specific color representation. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2004. 1.2, 5.1

[57] R. Padoan, T.A.G. Steemers, M.E. Klein, B.J. Aalderink, and G. de Bruin. Quantitative hyperspectral imaging of historical documents : Technique and applications. In *International Conference on NDT of Art*, 2008. 3.1

[58] Z. Pan, G. Healey, M. Prasad, and B. Tromber. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, 2003. 3.1

[59] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on Graphics*, 22(3):313–318, 2003. 1.1, 2.1.2, 2.2.1, 3.1, 5.1

[60] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3):664–672, 2004. 3.2

[61] C. Pohl and J.L. Van Genderen. Multisensor image fusion in remote sensing : concepts, methods and applications. *International Journal on Remote Sensing*, 19(5):823–854, 1998. 3.2

[62] M.T. Qadri, K.T. Tan, and M. Ghanbari. Frequency domain blockiness and blurriness meter for image quality assessment. *International Journal of Image Processing*, 5(3):352, 2011. 1.1

[63] R. Raskar, A. Ilie, and J. Yu. Image fusion for context enhancement and video surrealism. *International Symposium on Non-Photorealistic Animation and Rendering*, pages 85–93, 2004. 2.1.2, 2.2.1, 3.2

[64] X. Ren and J. Malik. Learning a classification model for segmentation. In *Proceedings of IEEE International Conference on Computer Vision*, 2003. 5.2.1, 5.4

[65] M. Sambongi, M. Igarashi, T. Obi, M. Yamaguchi, N. Ohyama, M. Kobayashi, Y. Sano, S. Yoshida, and K. Gono. Analysis of spectral reflectance using normalization method from endoscopic spectroscopy system. *Optical Review*, 9(6):238–243, 2002. 3.1

[66] Z. Shi and V. Govindaraju. Historical document image enhancement using background light intensity normalization. In *Proceedings of International Conference on Pattern Recognition*, 2004. 3.4

[67] T. Shibata, A. Iketani, and S. Senda. Image inpainting based on probabilistic structure estimation. In *Asian Conference on Computer Vision*, pages 109–120, 2010. 2.2.2, 5.1, 5.1, 5.2.1, 5.3

[68] D.A. Socolinsky and L.B. Wolff. Multispectral image visualization through first-order fusion. *IEEE Transactions on Image Processing*, 11(8):923–931, 2002. 3.1

[69] D. Sun and W.K. Cham. Postprocessing of low bit-rate block dct coded images based on a fields of experts prior. *IEEE Transactions on Image Processing*, 16(11):2743–2751, 2007. (document), 1.2, 4.2, 4.4, 4.1

[70] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, and V. Kolmogorov. A comparative study of energy minimization methods for markov random fields. In *Proceedings of European Conference on Computer Vision*, 2006. 3.3.2

[71] C.L. Tan, R. Cao, and P. Shen. Restoration of archival documents using a wavelet technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(10):1399–1404, 2002. 3.4

[72] C.C. Tseng and S.L. Lee. Digital image sharpening using fractional derivative and mach band effect. In *IEEE International Symposium on Circuits and Systems*, pages 2765–2768, 2012. 1.1

[73] E. Vazquez, R. Baldrich, J. van de Weijer, and M. Vanrell. Describing reflectances for color segmentation robust to shadows, highlights, and textures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):917–930, 2011. 5.4

[74] B.V.K. Vijaya Kumar, M. Savvides, K. Venkataramani, and C. Xie. Spatial frequency domain image processing for biometric recognition. In *Proceedings of International Conference on Image Processing*, 2002. 1.1

[75] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991. 5.4

[76] C.Y. Wang and L.W. Lee, S.M. Chang. Designing jpeg quantization tables based on human visual system. *Image Communication*, 16(5):501–506, 2001. 4.1

[77] A.B. Watson. Dct quantization matrices visually optimized for individual images. In *Proceedings of the International Society for Optics and Photonics*, volume 1913, pages 202–216, 1993. 4.1

[78] Y. Weiss. Deriving intrinsic images from image sequences. In *Proceedings of IEEE International Conference on Computer Vision*, pages 68–75, 2001. 3.1

[79] X. Xiao and L. Ma. Gradient-preserving color transfer. In *Computer Graphics Forum*, volume 28, pages 1879–1886, 2009. 2.2.2, 5.1, 5.1, 5.2.1, 5.3

[80] Z.F. Xie, R.W.H. Lau, Y. Gui, M.G. Chen, and L.Z. Ma. A gradient-domain-based edge-preserving sharpen filter. *The Visual Computer*, pages 1–13, 2012. 1.1

[81] Z. Xiong, M.T. Orchard, and Y.Q. Zhang. A deblocking algorithm for jpeg compressed images using overcomplete wavelet representations. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(2):433–437, 1997. 1.2

[82] W. Yang, J. Zheng, J. Cai, S. Rahardja, and C.W. Chen. Natural and seamless image composition with color control. *IEEE Transactions on Image Processing*, 18(11), 2009. 2.2.2, 5.1, 5.1, 5.2.1, 5.3

[83] C. Yim and A.C. Bovik. Quality assessment of deblocked images. *IEEE Transactions on Image Processing*, 20(1):88–98, 2011. 4.2

[84] A. Zakhor. Iterative procedures for reduction of blocking effects in transform image coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 2(1):91–95, 1992. 1.2, 4.2

[85] Y. Zeng, W. Chen, and Q.S. Peng. A novel variational image model: Towards a unified approach to image editing. *Journal of Computer Science and Technology*, 21(2), 2006. 2.2.2, 5.1, 5.1, 5.2.1, 5.3

[86] G. Zhai, W. Zhang, X. Yang, W. Lin, and Y. Xu. Efficient image deblocking based on postfiltering in shifted windows. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1):122–126, 2008. 1.2

[87] P. Zhang, S. Wang, and R. Wang. Reducing frequency-domain artifacts of binary image due to coarse sampling by repeated interpolation and smoothing of radon projections. *Journal of Visual Communication and Image Representation*, 2012. 1.1

[88] X.P. Zhang, T. Sim, and X.P. Miao. Enhancing photographs with near infra-red images. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 2008. 3.2, 5.3.2