GRADIENT VARIATION: A KEY TO ENHANCING PHOTOGRAPHS ACROSS ILLUMINATION

ZHANG XIAOPENG

NATIONAL UNIVERSITY OF SINGAPORE

2009

GRADIENT VARIATION: A KEY TO ENHANCING PHOTOGRAPHS ACROSS ILLUMINATION

ZHANG XIAOPENG
(B.Sc., Beijing Institute of Technology, 2000)

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE THE DEGREE OF

Doctor of Philosophy

in

SCHOOL OF COMPUTING

NATIONAL UNIVERSITY OF SINGAPORE
SINGAPORE, 2009

Copyright © 2009 by Zhang Xiaopeng

This thesis is dedicated to my beloved wife,
Bian Huimin

Acknowledgements

First of all, I would like to express my deepest appreciation to my supervisor Dr. Terence Sim for his enthusiastic and patient guidance, constant assistance and encouragement on me throughout my PhD studies.

I would like to thank Dr. Kok-Lim Low and Dr. Michael S. Brown for being part of my committee and providing me valuable guidance and suggestions to improve this thesis.

I would also like to thank Assoc. Prof. Leow Wee Kheng for his valuable suggestions and encouragement in my research work. His expertise greatly honed my research abilities, and matured my work ethics.

My deep appreciation also goes to my dear colleagues in Computer Vision Lab: Zhang Sheng, Miao Xiaoping, Wang Ruixuan, Ding Feng, Janakiraman Rajkumar, Zhuo Shaojie, Guo Dong, Ye Ning, Li Hao, Qi Yingyi, Ha Mailan, Li Jianran, Lu Zheng, Yu-Wing Tai, and *etc*. I really enjoyed my collaborations and fruitful discussions with these brilliant people.

I would also like to thank my lovely friends in Singapore: Sheng Chang, Zhang Xia, Xiang Junlian, Wang Xianjun, Chen Su, and *etc*., for their warm friendship which made my life here easy and joyful.

Finally, I owe my families for their steadfast support, encouragement, faith and love throughout. I am indebted to my wife, Bian Huimin, for her eternal love and loyal support from the other end of the earth.

Abstract

Extreme lighting conditions, such as a high dynamic range (HDR) scene or a low light environment, may introduce many kinds of defects for digital photography. These artifacts, including pixel saturation, noise, blurring, *etc*., have been studied as individual problems in the area of image processing for decades. Recently, Computational Photography, an emerging area across computer vision and computer graphics, attempts to resolve the problems by "reinventing" digital cameras. In light of previous contributions in this area, this thesis presents two novel solutions to enhance photographs in HDR scene and low light environment respectively. The key ideas behind both solutions are based on one fundamental observation: the image gradient variation due to illumination change. However, such variation has been either ignored or over simplified in the existing works. To enhance photography in HDR scenes, we build a dual-camera system, being able to capture a normal (visible-light) photograph and its near infra-red counterpart with a single shot. By manipulating the gradient magnitude based on near infra-red information, the visible photograph can be enhanced in both contrast and texture. For low light environment photograph enhancement, we propose a selective re-flashing method which requires only two input flash photographs with different flash intensities. By decomposing the radiance gradient, our method can faithfully recover the ambient image and allows the user to re-flash the selected regions. Compared to the existing works, our approach is simple yet efficient and able to achieve higher visual quality. Besides, we hope that our work can help better understand the relationship between gradient and illumination, and provide more insights for digital photograph enhancement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *Light makes photography. Embrace light. Admire it. Love it. But above all, know light. Know it for all you are worth, and you will know the key to photography.*
>
> George Eastman
> (The Founder of *Kodak*)

Photography is all about capturing and recording light. This rule, since the birth of daguerreotype nearly 170 years ago [Coe 1976], remains the same even with the most advance of the photo sensing technology today. Light plays a critical role in photography, but it is not always easy to control. For example, in the scene where the brightness contrast is very high, *a.k.a.* high dynamic range, the bright regions are easily over saturated while the dark regions are still underexposed, as shown in Figure 1.1. Another example could be photography in low light environment. In some scenarios, such as outdoor night activities, dimly lit parties, or even simple indoor portraits, photographs are very likely corrupted by noise, motion blur, or flash artifacts. Figure 1.2 and 1.3 illustrate two kinds of common low light situations, in which many kinds of artifacts are exhibited.

The imperfections in conventional digital cameras stimulate the increasing popularity of Computational Photography which "re-invents" cameras and redefines ways of capturing and enhancing digital photography. Existing works can be classified into three categories:

Computational Optics Novel optic elements enable cameras to gain more information from a scene with more flexible exposure settings, such as refocusing using 4D lightfield [Ng et al. 2005; Lumsdaine and Georgiev 2009; Levin et al. 2009], depth recovering with coded radiance [Veeraraghavan et al. 2007; Levin et al. 2007; Bando et al. 2008], deblurring through obtained motion kernels [Ben-Ezra and K.Nayar 2004; Agrawal and Raskar 2007; Tai et al. 2008; Agrawal et al. 2009; Cho and Lee 2009], *etc*.

Computational Illumination High dynamic range illumination can be effectively captured using the approaches presented in [Debevec and Malik 1997; Nayar and Mitsunaga 2000], and represented within low dynamic range as proposed in [Reinhard et al. 2002; Fattal et al. 2002]. In low light environment, the ambient image can be enhanced by adding into it rich and sharp details which are extracted from its flash counterpart, as proposed in [Eisemann and Durand 2004; Petschnigg et al. 2004; Raskar et al. 2004a; Miao and Sim 2005; Agrawal et al. 2005b; Krishnan and Fergus 2009].

Computational Processing Recent works have enriched image processing with new techniques. For instance, Poisson Image Editing [Pérez et al. 2003] opens new era for gradient domain image processing, Bilateral Filtering [Tomasi and Manduchi 1998], as well as Non-local Means [Mahmoudi and Sapiro 2005], have become the de facto standard techniques for feature-preserved image smoothing. Besides, human intervention has been increasingly incorporated in image processing to provide high level prior knowledge, such as the scribble-based marking used in [Li et al. 2004; Agarwala et al. 2004].

1.1 Overview

This thesis deals with new techniques in the second and third category above, *i.e*. computational illumination and computational processing, specifically, on how to enhance photographs taken in the high dynamic scene or the low light environment. Our techniques are based on our major observation about the connection between the radiance gradient and illumination variation. We find that such connection can be a key to understanding and resolving the problems, however, it has been

neglected in previous studies. In this work, we first study radiance gradient variation based on the BRDF model, and then derive its mathematical relationship with illumination change. Based on this study, we propose: (1) a novel solution that enhances low dynamic range representation of HDR scenes by showing better contrast and recovering lost texture details, (2) a new computational re-flashing approach which is able to capture and present low light ambient radiance with high visual quality.

The subsequent sections first present the motivation of our work by summarizing the challenges faced by digital photography under two extreme lighting conditions, *i.e.* high dynamic range and low light environment, followed by our research objectives and contributions.

1.2 Challenges under Various Lighting Conditions

1.2.1 High Dynamic Range

High dynamic range (HDR) means the intensity range of the scene illumination is much higher than that of digital photographs. The usual contrast ratio of an outdoor scene in natural daylight can be up to 10^{10}, which is much greater than the dynamic range recorded by the current digital cameras (10^2), resulting in the problems of overexposure or underexposure. Figure 1.1 gives an example: under high brightness, both color and textures are washed out, while in the darkness, details are barely seen.

Given the limited dynamic range of camera sensor, it is very difficult for a common digital camera to capture HDR radiance with a single image. Professional photographers therefore choose to take photographs in RAW format, in which scene radiance is recorded with 12 or 14 bits per channel, much wider than the conventional 8 bit per channel. As a result, more details can be revealed in the image if the brightness contrast is properly adjusted. However, such adjustment usually needs to be done region by region, which requires significant manual effort as well as experience. Furthermore, the dynamic range of RAW format is still limited.

Other than using a single input image, researchers have proposed ways to

Figure 1.1: This figure shows two typical high dynamic range scenes. The left figure is taken inside a hall with glass ceiling, and the right one is taken in a forest toward the bright sky. As it can be seen, both images suffer from the similar problem: the bright regions (the glass ceiling and the sky) have been over saturated, but some regions are still in the darkness completely.

capture HDR radiance by taking a set of photographs with different exposure settings [Debevec and Malik 1997]. Then the recovered HDR radiance can be compressed to low dynamic range with tone-mapping methods [Reinhard et al. 2002; Fattal et al. 2002]. In theory, HDR based tone-mapping can preserve and reveal all the details of the original radiance. However, the demanding requirement of multiple inputs restricts the method from being widely used.

1.2.2 Low Light Photography

Low light environment, such as outdoor night, candle lighting, dim indoor lighting, is another challenge faced by photographers, where the ambient illumination is too weak to be sensed by today's digital cameras. Though various ways can be used to compensate the low lighting, they are not without their problems and limitations. For example, using larger aperture can gain more incident light but lose the depth of field; prolonging exposure time can also capture more ambiance but may cause blurring artifacts; increasing ISO can avoid blurring but boost image noise.

Flash photography has been widely used in low light environment. However, artificial flash is typically strong and harsh, which usually ruins the natural ambient radiance and causes uneven exposure. Figure 1.2 gives an example. The left image is taken without flash but with high ISO value and slow shutter speed. As we can see, due to the low light condition, the subject in the image suffers from motion blur and noise. In contrast, the image on the right, a typical photograph taken with strong flash and fast shutter speed, presents a sharp and clear image. However, the image fails to convey an crucial element in the scene - the gentle atmosphere rendered by the candle light. To mitigate such unpleasant flash effects, experienced photographers often choose *slow sync flash mode* (or *night portrait mode* in low-end cameras) as an alternative, which is essentially a long exposure ending with a quick flash burst. Figure 1.3 shows a comparison between images in this flash mode and a normal flash mode. The upper image is taken with a normal flash mode. Again, due to the strong flash and fast shutter speed, the ambient radiance looks pale and the flash is yet unable to reach the distant background which is eventually veiled in darkness. In contrast, when the slow sync flash mode is applied, as shown in the lower image, the ambiance can be well preserved. However, this solution is

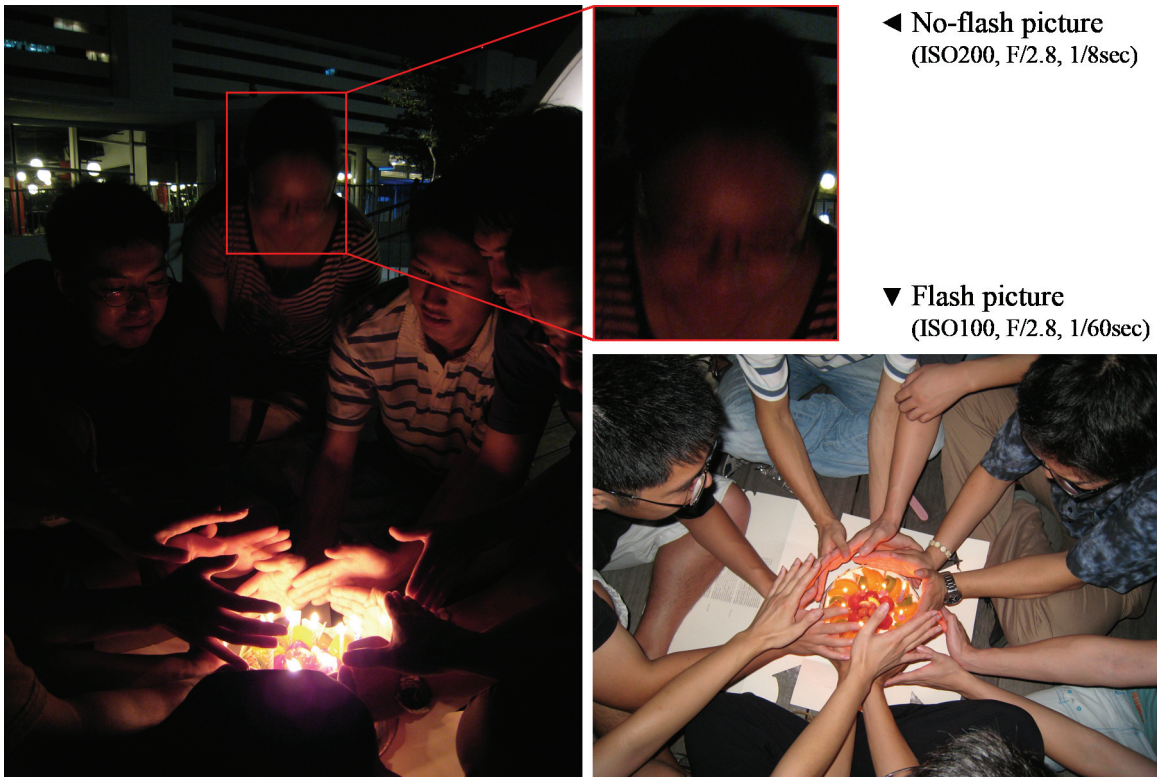


Figure 1.2: This set of images were taken in an outdoor birthday party at night. In the no-flash image on the left, the image quality is seriously degraded by many artifacts, such as noise, blur, and underexposure. On the contrary, the flash image on the right exhibits high image quality, but the ambient radiance has been completely ruined by the strong and harsh flash.

not perfect. As we can see in the image, the blurring and "ghost" effects seriously degrade the visual quality. This is because the exposure time of the slow sync flash image is as long as 1.3 seconds, which is too long for a casual hand-held shooting, causing motion blur; and the quick flash at the end creates a sharp image overlapping with the blur, causing the "ghost" effect.

In Computational Photography, researchers tended to take advantage of both flash and no-flash photographs [Eisemann and Durand 2004; Petschnigg et al. 2004; Agrawal et al. 2005b], or exploit multiple flash pictures [Raskar et al. 2004c; Miao and Sim 2005], to enhance the ambient image while eliminating the flash artifacts. A common issue of these methods is their strong dependence on the static scene assumption, since they usually require at least two images of the same scene as

Figure 1.3: These two images were taken inside a restaurant, where the small ceiling lights and the table candles are the only ambient illuminations. Under the normal flash mode shown in the upper image, similar to Figure 1.2, fast shutter speed and strong flash yield high image quality at the cost of losing the ambient radiance. Under the slow sync flash mode (or night portrait mode) shown in the lower picture, the longer exposure with the final quick flash can help preserve the ambient radiance but may cause motion blurring and ghost effects, as highlighted in the figure.

inputs. More details will be elaborated in the next chapter.

1.3 Motivation

While many contributions have been made in recent years to address the challenging problems of enhancing photographs in HDR scenes or low light environment, previous approaches are limited in the following two ways:

Dynamic Scenes Most methods require multiple input photographs (*e.g.* photos with different exposure settings or flash/no-flash pairs) to be aligned pixel by pixel. Such aligned photographs can only be easily obtained in static scenes. In dynamic scenes, however, those methods are hardly applicable since it is difficult to gather the required inputs during one exposure.

Dynamic Illumination While more inputs provide more information, new problems may be introduced especially when images are captured under different lighting conditions. For example, flash may cause cast shadows, specularities, and inter-reflection, while weak ambient image may introduce noise and blur. Removing such side effects is not easy and increases the overall complexity.

In the previous works, the impact of illumination variation is either painfully removed or simply treated as variation of the image intensity. In this thesis, we will show that the illumination variation, on the contrary, is important to understand and resolve the problems listed in Section 1.2. To be precise, we first derive a mathematical relationship between illumination change and radiance gradient. Then we exploit the relationship to enhance digital photography in different scenarios.

1.4 Objectives

As discussed in the previous sections, the problems of enhancing digital photography in two different scenarios: high dynamic range scenes and low light environment, are challenging and worthy of study. Although the photographs in

these two scenarios exhibit completely different appearance and artifacts, in this thesis the problems are bridged together by the same key factor: gradient variation due to illumination change. The objective of our work is to solve the two problems by exploiting this key factor. More specifically:

- For scenes in high dynamic range illumination, we aim at acquiring a low dynamic representation with better brightness contrast and richer texture details than conventional photographs.
- For scenes in low lighting condition, our goal is to provide an efficient solution such that the weak ambient radiance can be recorded and presented with high visual quality.
- Being an approach requiring multiple photographs as inputs, our method should minimize its dependence on the static scene assumption, thus avoiding artifacts such as motion blurring or thermal noise.
- Our computational techniques will also incorporate with computer vision and/or human interaction. With computer vision, our approach should be as automatic as possible. Being aided by human, the intervention must be simple and intuitive enough.

1.5 Contributions

In terms of original contributions to the field of computational photography, this thesis makes the following two:

1. Derive and analyze a mathematical relationship between illumination variation and radiance gradient.
2. Demonstrate such relationship is a key to understanding and enhancing digital photography under different illumination, by successfully applying it in two complicated scenarios: high dynamic range and low light environment.

1.6 Road Map

The rest of this thesis is organized as follows: the next chapter provides the background and recent related work on high dynamic range imaging, computational flash photography, gradient based image editing, and a brief review of interactive computer vision. Chapter 3 presents our theoretical study on the relationship between radiance gradient variation and illumination change. Chapter 4 and 5 demonstrate how our theory can be applied to enhance photographs under high dynamic range illumination or low light environment. Finally, Chapter 6 concludes the thesis along with the discussion of future research directions.

Chapter 2

Literature Survey

In this thesis, our research focuses on gradient variation across different illuminations, and in particular, we will tackle photography enhancement in high dynamic range and low light environment. The digital photographs captured in such situations could be degraded by many artifacts, such as the blooming effect by overexposure, blurring and noise due to slow shutter speed, shadows or specularities produced by flash. Some of these artifacts, as classical problems in the domain of image processing, have been studied individually for decades. Instead of resolving these artifacts one by one through post processing, in computational photography, we aim to invent new capturing and representing solutions which can easily remove or avoid these artifacts as a whole. In [2009], Raskar and Tumblin already gave an intensive introduction of most recent research in Computational Photography. In this chapter, we will briefly review some of the most related works of its two branches: computational illumination and processing.

2.1 Computational Processing

Broadly speaking, many classical problems can be categorized in the area of computational processing, such as image segmentation, matting, colorization, *etc*. These problems have attracted great attention recently, and been attempted by a large amount of new methods. In this section, we will review two basic techniques, namely, Bilateral Filtering and Poisson Image Editing, which are widely used in

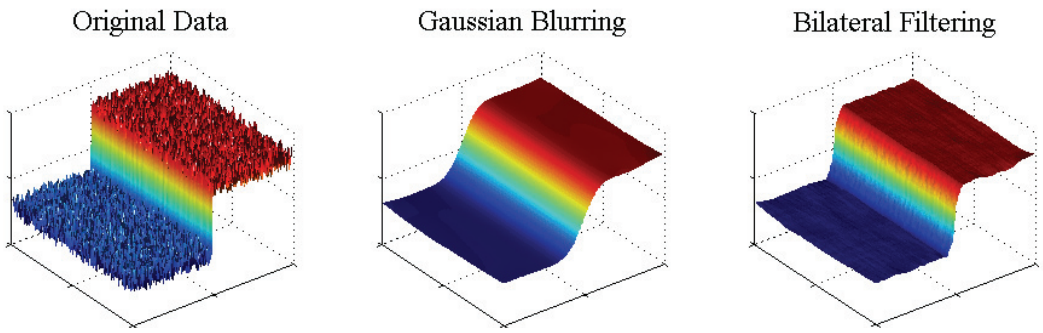


Figure 2.1: Compared with Gaussian blurring, bilateral filter can smooth the small noise while preserving the strong feature.

many of these methods (including ours). Since Poisson image editing is processing based on gradient domain, we will also review some latest techniques on image reconstruction from gradients.

2.1.1 Bilateral Filtering

Bilateral filter was first invented by Tomasi and Manduchi in [1998]. It is, in fact, a blurring filter consists of two Gaussian weighting functions, as shown in Equation (2.1). The usual Gaussian blurring of a pixel p is a weighted average of the pixel and its neighbors ($G_{\sigma_S}(\|p-q\|)$). On top of it, bilateral filter introduced another weighting factor: intensity difference, to preserve strong edges ($G_{\sigma_r}(|I_p - I_q|)$). Pixels, which are much different from their neighbors, will be assigned low weight. And pixels having similar intensities are given higher weight for blurring. Thus, bilateral filter is an edge-preserving filters. Figure 2.1 illustrates the difference between a usual Gaussian filter and bilateral filter.

$$BF[I]_p = \frac{1}{W_p} \sum_{q \in S} G_{\sigma_S}(\|p-q\|) G_{\sigma_r}(|I_p - I_q|) I_q \tag{2.1}$$

In Computational Photography, bilateral filter is widely used to decompose an image to a large scale layer and a detail layer [Oh et al. 2001; Durand and Dorsey 2002; Tumblin and Turk 1999; Bae et al. 2006], or separate the two layers across different images (*a.k.a.* joint bilateral filter) [Petschnigg et al. 2004; Eisemann and Durand 2004]. In our work described in Chapter 4, we used bilateral filter to

separate large scale layer of brightness for contrast transferring.

2.1.2 Poisson Image Editing

Poisson image editing (PIE), brought by Pérez *et al*. in [2003], is a significant contribution on gradient-based image processing. Pérez *et al*. were the first to formulate image compositing problem as a Poisson equation with a guided vector field. The goal of image compositing is to blend the region of interest (ROI) from the source onto the target image, as illustrated in Figure 2.2. In the final composition image c, let Ω denote the unknown closed region with boundary $\partial\Omega$, and let r denote the pixel value over Ω. Then PIE introduces another vector field $\mathbf{v}$, and defines r as the solution of the following minimization problem,

$$\min_r = \iint_\Omega |\nabla r - \mathbf{v}|^2 \quad with \quad r|_{\partial\Omega} = t|_{\partial\Omega}, \tag{2.2}$$

where ∇ is the gradient operator. The solution to Equation (2.2) is given by the following Poisson equation with Dirichlet boundary conditions,

$$\Delta r = div(\mathbf{v}) \quad with \quad r|_{\partial\Omega} = t|_{\partial\Omega}, \tag{2.3}$$

where $div(\mathbf{v})$ is divergence of $\mathbf{v}$.

By defining different $\mathbf{v}$, solving Equation (2.2) can accomplish different goals. For instance, to achieve seamless compositing, $\mathbf{v}$ is usually defined as the gradient of s, *e.g*. $\mathbf{v} = \nabla s$. In this way, Equation (2.2) forces the pixel values inside Ω to be as close as s, and the pixel values on the boundary $\partial\Omega$ to be as close as t, resulting in a seamless blending without visible discontinuities around the boundary. Similarly, by changing $\mathbf{v}$, PIE can successfully obtain many different compositing effects, such as concealment, exchange texture, local illumination and color change, *etc*.

Undoubtedly, this idea inspires many later works on image composition, such as Interactive Photomontage [Agarwala et al. 2004], image completion [Sun et al. 2005], Drag-and-drop Pasting [Jia et al. 2006], scene completion [Hays and Efros 2007], *etc*. More importantly, such a Poisson problem formulation is also extended to many other applications, such as Poisson Matting [Sun et al. 2004], Poisson

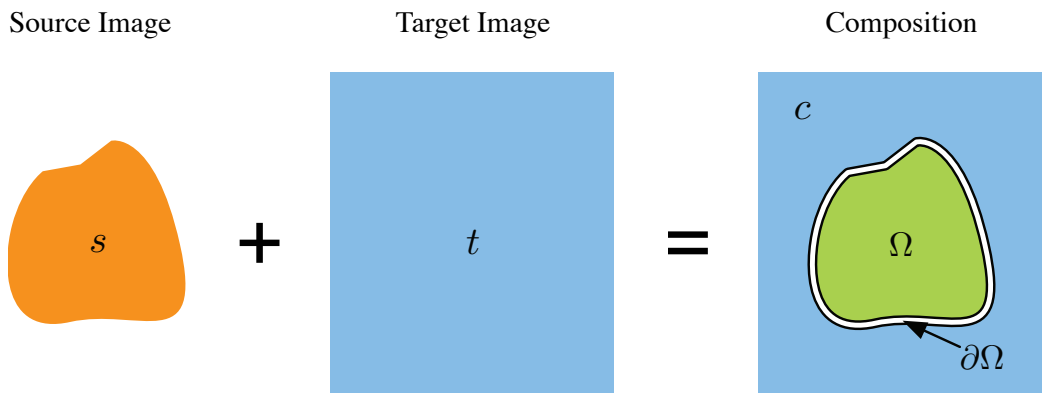


Figure 2.2: An illustration of image compositing problem. s denotes the source image, t denotes the target image, c denotes the final composition, and Ω denotes the unknown closed region with boundary $\partial\Omega$.

Surface Reconstruction [Kazhdan et al. 2006], face makeup [Guo and Sim 2009], *etc*.

However, most of these works seldom studied how the guided vector field is affected by the illumination. In PIE, Pérez *et al*. simply changed the magnitude of the vector field to change local brightness, which was similar to [Fattal et al. 2002]. Such gradient technique is based on an assumption that image gradients, especially their orientation, are usually invariant to illumination change. But, as shown in Chapter 3, this assumption does not always hold: for surfaces with arbitrary BRDFs, in low-light conditions, both the magnitude and orientation of image gradients *are* indeed sensitive to illumination changes.

2.1.3 Image Reconstruction from Gradients

Reconstructing image Z from its gradient (p, q) means integrating the gradient field under integrable constraint

$$\frac{\partial^2 Z}{\partial x \partial y} = \frac{\partial^2 Z}{\partial y \partial x} \tag{2.4}$$

and under certain boundary condition (e.g. Dirichlet condition). As shown in poisson image editing, the problem can be formulate as solving a Poisson equation:

$$\Delta Z = \frac{\partial p}{\partial x} + \frac{\partial q}{\partial y} = div(p, q), \tag{2.5}$$

where $div()$ is divergence operator. And this equals to solving a sparse linear system $Ax = b$, where A is a sparse matrix standing for Laplacian operation, vector x contains image intensity of every pixel, and b is divergence value. Usually (p, q) are not integrable, and least square solution given by standard Poisson solver is not good enough.

In [2005a], Agrawal *et al*. tried to force integrability by correcting curl value, since integrable gradients should have zero curl. In their following work [2006b], they compared several gradient reconstruction algorithms and proposed a better solution using anisotropic diffusion. The idea is simple: they applied an edge-preserving diffusion tensor (similar to affine transformation) on both sides of the Poisson equation, and formed a new sparse system $A'x = b'$, which is proved to be very robust to noise. In our work, we applied their method reconstruct radiance from manipulated gradient field.

2.2 Computational Illumination

2.2.1 High Dynamic Range Imaging

Essentially, high dynamic range imaging (HDRI) includes two parts: capturing and representing. For a common digital camera, it is difficult to record the HDR radiance with a single shot, because the photodides cannot be "excited" if the incident light is too strong [Reinhard et al. 2006]. Therefore, capturing HDR radiance using multiple photographs is still the most popular way so far. Representing high dynamic range radiance on low dynamic range medium, such as printing on paper or displaying on monitors, is called HDR tone mapping problem. The challenge for tone mapping is how to reproduce visual appearance, meaning the mapped radiance must appear as same as the HDR scene in human eyes. A common pipeline of high dynamic range imaging is illustrated in Figure 2.3.

Capturing high dynamic range radiance

Under assumption of linear camera response, the HDR radiance can be achieved by properly combining radiance value of multiple low dynamic range (LDR) pho-

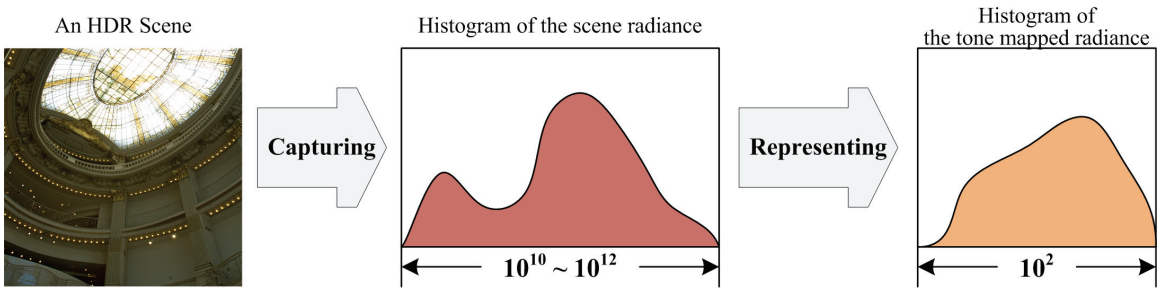


Figure 2.3: An illustration of high dynamic range imaging procedure. The contrast of radiance intensity in an HDR scene could be as large as $10^{10} \sim 10^{12}$. After tone mapping, the radiance can be properly displayed on the low dynamic range medium.

tographs. The photographs are captured in a static scene using different exposure time. Because of linear response, the image intensity can be converted to radiance domain by simply dividing the exposure duration. Then for each pixel, the radiance value should be similar excluding the over/under exposed recordings. By averaging corresponding pixel radiance across proper exposures, the result is HDR radiance.

However, the response of most digital cameras are not linear. Debevec and Malik [1997] demonstrated a simple and robust way to recover the nonlinear camera response function from a series of aligned exposures. They sampled a set of points which consisted of a segment of camera response curve in every exposure. A smoothed response curve was produced by fitting theses segments together with linear optimization. Similarly, Mitsunaga and Nayar [1999] modeled the response function with a polynomial equation, which was determined using the linear least square fitting.

Tone mapping

Tone mapping methods can be roughly divided into two categories:

Global operator: compressing the dynamic range with an identical function for each pixel

Local operator: reducing the dynamic range with a nonlinear spatially varying function.

Table 2.1: A comparison between HDR + Tone mapping and our method

	HDR + Tone mapping	Our Method
Input	A series of aligned photographs with different exposure	A visible photograph and its corresponding near infra-red picture, obtained with a single shot
Applicability in dynamic scene	No	Yes
Automation	No. Tone mapping method usually have thresholds for global or local operators.	Yes. Fully automatic.
Details preserving	All details are preserved	Most details can be preserved (benefiting from texture transfer)
Visual quality	Depending on tone mapping methods and parameter tuning	High

To match the visual appearance globally, researchers derived many different nonlinear curve mapping radiance to image, based on some psychological model of human vision, such as [Tumblin and Rushmeier 1999; Reinhard and Devlin 2005]. However, such operators can hardly preserve fine details in high dynamic range data. In [2002], Reinhard *et al*. scaled the radiance globally based on a "key" luminance, and then they invented a dodge-and-burn operator to manipulate contrast for each pixel differently. In gradient domain, Fattal *et al*. [2002] proposed a method to adaptively attenuate gradients with large magnitude. Thus, the fine details with small gradient magnitude were well preserved after dynamic range reduction.

Comparison with our method

In Chapter 4, we describe a novel method to enhance visible photographs captured in HDR scene with its corresponding near infra-red pictures. A comparison between HDR + Tone mapping and our method is summarized in Table 2.1

2.2.2 Flash Photography Enhancing

Photographs taken in low light environment often exhibit many kinds of artifacts, such as noise, blurring, underexposure, *etc*. Research on removing such artifacts has been carried on for decades. So far, firing strong flash to compensate weak ambiance is still the most popular way to obtain high quality photos in low light situation. But flash will cause other problems, such as uneven exposure, shadows and specularities. This leads in an interesting research topic in computational illumination: how to use flash to enhance low light photography?

Conventional Flash Exposure Bracketing

Flash Exposure Compensation (FEC) and Flash Exposure Bracketing (FEB) are common techniques being used by photographers for years [Lefkowitz 1981; Canon]. When subjects appear too dark compared with background, photographers need to flash them to compensate such low light condition. In such case, flash output is usually set at low level to illuminate subjects only, and it is set according to photographers' experience or camera's evaluation system (such as E-TTL of Canon). Some latest models of flash allow photographers to take three photos continuously with different flash intensity, and photographers subjectively pick the best one from the 3 pictures. Usually flash intensity is changed within a range of 1-3 stops below and above the normal output, which is called flash exposure bracketing. Since FEC and FEB will calculate and adjust flash output based on subject's lighting condition, they can help photographers take properly-exposed photos by minimizing flash artifacts. However, FEC and FEB usually require long exposure time because of low flash intensity, which may still cause motion blurring. Besides, such conventional tricks are highly dependent on hardware and photographers' experience.

Conventional flash photography cannot avoid flash artifacts completely, and does not work well when ambient illumination is very low. Many computational approaches have been proposed to increase visual quality of low-light photographs by combining strength of flash/no-flash image pairs, multiple illuminating, image based re-lighting, *etc*.

Flash and No-flash Image Pairs

Among the earliest research in the area of enhancing low-light photography using flash must be the contemporaneous work of [Petschnigg et al. 2004] and [Eisemann and Durand 2004]. Petschnigg *et al*. used bilateral filter to extract detail-layer of flash image and base-layer of ambient image, and blended them together based on the mask of shadow and specularity. The Bilateral filter can catch slow variation of an image while preserving strong features. By utilizing this characteristic, their results successfully combined sharp details from flash image with overall ambient illumination. But their techniques on shadows and specularities removal require thresholds and are not fully automatic. Eisemann and Durand proposed a similar method incorporating with bilateral filter either: they not only transferred details, but also transferred color information from flash image to final output. They treated harsh shadows and soft shadows caused by flash in a manual way. The major contribution of these approaches is that they apply bilateral filter to extract and fuse desired features from multi-modal images, which is proven to be both simple and efficient. However, bilateral filtering is a pixelwise operation so the result is very sensitive to noise. Meanwhile, shadows and specularities can not be identified automatically.

Agrawal *et al*. [2005b] presented a gradient projection method that can automatically remove highlights from flash image and undesired reflection from ambient image. They assume that the orientation of image gradients does not change in flash/no-flash image pairs. Based on this assumption, they corrected gradients orientation for pixels with artifacts in both flash and ambient image. In [2006a], Agrawal *et al*. proposed an edge suppression operator which can suppress edges by projecting edge gradient vector onto its own orthogonal direction, *i.e*. reducing the gradient magnitude to zero. Therefore, if the ambient shadow edges do not show in the corresponding flash image, their method can suppress such edges and thus remove the shadow. Similarly, the reflection in flash image can be removed as well. Again, their method still does not consider the variation of gradient orientation due to illumination change.

Almost all approaches based on flash/no-flash image pair strictly require the static scene assumption, meaning both camera and the captured scene are static

during exposure. The reason is that these methods require flash and ambient images to be taken in the same scene or at lease geometrically aligned for pixelwise processing. Obviously, such static assumption has greatly limited application of these methods. Also, the non-flash image is captured either using a high ISO, which results in grainy image noise, or using a long exposure, which requires both the camera and scene to be stationary. This greatly limits the method's practicality.

Multiple Flash Imaging

Multiple flash imaging refers to taking photographs from the same viewpoint but under different flash conditions, such as with various flash intensities or with various flash positions. A flash and no-flash image pair can be regarded as a special case of multi-flash images with just two flash intensities (flash on and flash off).

Raskar *et al*. built a camera prototype with 4 flashes located in different positions [2004c], as shown in Figure 2.4. These flashes will cast shadows in different directions, which makes it easy to identify perceptual edges in flash images for non-photorealistic rendering. Their method is the first computational approach to utilize multiple flashes, and is built on a simple observation that shadows' variation due to flash from different directions. Unfortunately, their study ignored flash effects on other aspects such as specularity, intensity, and gradient, which could be important clues to enhance low-light photography.

Miao and Sim [2005] proposed a linear model to recover ambient image from multiple flash images taken the same scene with different flash intensities. They then re-render the ambient image by simulating a longer exposure time, resulting in brighter images that still preserves the original ambient illumination. Their method is simple and computationally efficient, yet the linear model is very sensitive to noise and their method does not suppress noise effectively.

Comparison with our method

In Chapter 5, we present a new approach to enhance low light photography using only two flash photographs. Given known flash intensity ratio, our method can recover the ambient and flash-only radiance, thus allowing the user to selectively re-flash the low light ambient image. Here, we compare our method against methods

Figure 2.4: A prototype of multi-flash camera invented by Raskar *et al*. in [2004c]

using flash and no-flash pair [Petschnigg et al. 2004; Eisemann and Durand 2004; Agrawal et al. 2005b] and multiple flash images [Miao and Sim 2005], and the result is summarized in Table 2.2.

2.3 Interactive Computer Vision

Recently, more and more approaches begin to seek user's interaction to gain high-level visual information which could be difficult to derive automatically. Such interaction can be implemented as selecting desired regions by simple scribbling or rough painting. The selected pixels mainly serve two purpose: (1) indicating the regions for processing (*e.g*. inpainting, matting, cloning, *etc*.) in such as [Pérez et al. 2003; Sun et al. 2004; Agarwala et al. 2004; Jia et al. 2006; Wang et al. 2007; Wu et al. 2007], (2) labeling the regions for data training, in such as [Li et al. 2004; Lu et al. 2009].

In our computational re-flashing method described in Chapter 5, we also incorporate with user interactive scribbling, marking the regions in which to re-flash or keep the ambiance.

Table 2.2: A comparison of methods: flash & no-flash, multiple flash, and our selective re-flashing approach.

	Flash and No-flash	Multiple Flash	Our Method
Input	A flash image and an ambient image	Multiple flash images with known flash intensity	Two flash images, only the ratio of two flash intensity is required
Static scene assumption	Strongly dependent	Strongly dependent	Weakly dependent
Method	• *Joint bilateral filter*: transferring texture or color from flash to ambient • *Gradient projection*: remove reflection or hot spots	• *Linear model*: recovering and re-rendering the ambient image	• *Gradient decomposition*: recovering the ambient and flash-only image, selective re-flashing
Trade-off	Parameter setting	Calibration	Calibration
Artifacts handling	• Shadows, specularities, and reflections are detected and removed using different ad hoc methods.	• Shadows and interreflection can be well separated. • Specularities are removed using two flashes.	• Shadows and interreflection are naturally separated and selectively suppressed. • Specularities can be effectively detected and removed based on visual cues from two flash images.
Visual quality	• Enhancing the image quality by fusing ambient and flash images, or removing flash artifacts. • The final result is largely dependent on the visual quality of captured ambient image.	• Re-rendering various effects. • But recovering is sensitive to noise.	• Recovering the ambient and flash-only images with high visual quality. • Allowing user to selectively re-flash or keep the ambiance of desired regions.

Chapter 3

Theory

3.1 Overview

It has been widely known that image gradients are affected by strong features, such as object shapes, textures, and depth edges (object boundaries). For smooth surfaces or flat texture with grainy, small and low-contrast features, image gradients usually have small magnitudes and change smoothly. Therefore, image gradients, particularly for their magnitudes, are often assumed to be invariant to illumination in a number of recent works, *e.g.* [Agrawal et al. 2006a; Agrawal et al. 2006b; Agrawal et al. 2005b; Raskar et al. 2004b]. Our study shows that this assumption does not hold for all situations. It is acceptable in situations in which there is sufficient ambient light. However, in low-light environments where ambient lighting is weak, image gradients are indeed sensitive to illumination changes.

In this chapter, we firstly map the image intensity to scene radiance with camera response function, and then derive a mathematical relationship between radiance gradients and illumination change based on BRDF model. We show that the illumination variation affects both gradient magnitude and orientation. We also discuss the conditions under which such affection on gradient orientation can be ignored and which not.

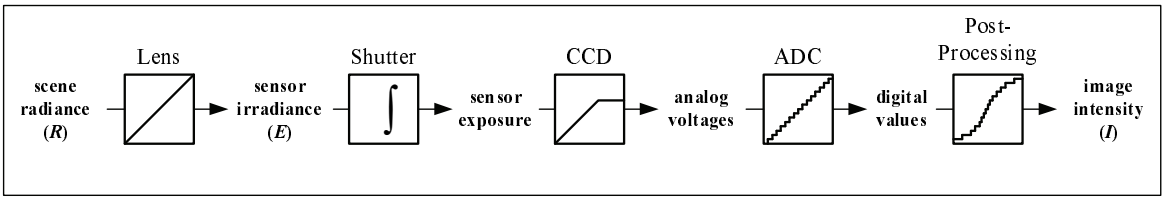


Figure 3.1: Image acquisition is a nonlinear mapping from scene radiance to image intensity. Such non-linearity can occur during exposure, digitization, and post processing. This figure is modified from the one in [Debevec and Malik 1997].

3.2 Scene Radiance and Image Intensity

The intensities of pixels, captured using a common digital camera, are usually nonlinear mappings of the scene radiance. Such non-linearity is often characterized by photo sensor response and post digital processing in camera, as illustrated in Figure 3.1. Mathematically, the relationship between scene radiance and image intensity can be defined using nonlinear and monotonically increasing function f, which is named camera response function (CRF). Therefore, we have

$$I = f(R), \tag{3.1}$$

where R is scene radiance and I is image intensity. Debevec and Malik have presented a method [1997] to recover f using a set of photographs taken with varying, known exposure time. Apparently, its inverse f^{-1} is also well defined. With obtained f, we can easily recover scene radiance R [1] from

$$R = f^{-1}(I). \tag{3.2}$$

Figure 3.2 shows the camera response function of our Canon 350D DSLR camera, which is used in our further experiments.

There are no two cameras having the same camera response function, even if they are of the same model. This implies that the nonlinear mapping from illumination intensity to pixel intensity is different from one camera to another.

[1] In fact, it should be sensor irradiance (E) but not scene radiance (R) here, as shown in Figure 3.1. But in modern cameras, since R is linear proportional to E constantly across all pixels [Debevec and Malik 1997; Kolb et al. 1995], it is safe to say that R is recovered up to an unknown scale factor.

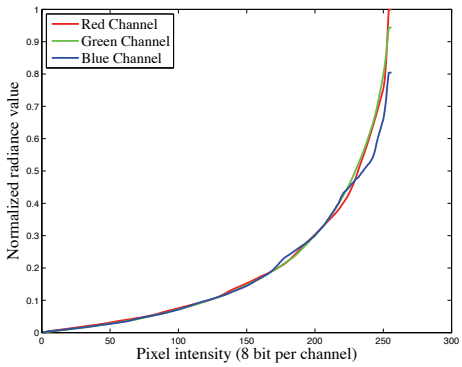


Figure 3.2: The camera response function of our Canon 350D DSLR camera. We recovered it using Debevec and Malik's method in [1997].

Although in some cameras, the mapping is close to linear in the middle range of pixel intensity, such a partial linearity is not sufficient in our work which studies the illumination variation over a wide range.

In contrast, radiance is a linear measurement of illumination intensity and is independent of cameras. And the gradient operation is linear as well. Therefore, when we investigate image gradients under varying lighting conditions, we exploit gradients of radiance value instead of gradients of pixel intensity.

3.3 Radiance Gradient

3.3.1 Radiance gradient based on BRDF model

To find out the relationship between radiance gradient and illumination, we first need to examine how radiance is reflected from illumination. As illustrated in Figure 3.3, the incoming irradiance incident on the surface point P from the direction ω_i, and the reflected radiance exit along ω_o. The direction ω can be parametrized by azimuth angle θ and elevation angle ϕ w.r.t. surface normal n. In computer vision and computer graphics, the ratio of outgoing radiance to the incident irradiance is usually modelled with the bidirectional reflectance distribution function (BRDF)

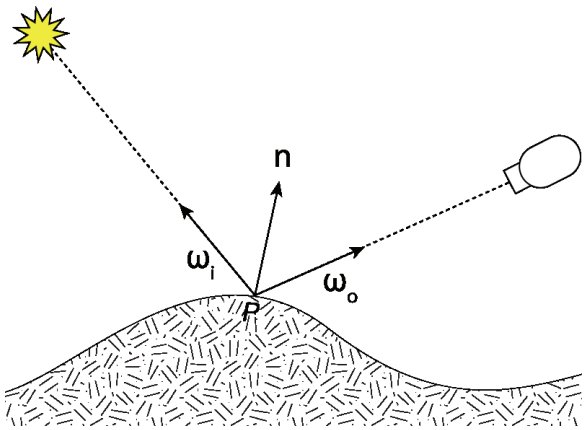


Figure 3.3: Diagram showing vectors used to define the BRDF, taken from [Wikipedia]. ω_i and ω_o are unit vectors, and n is the surface normal.

[Forsyth and Ponce 2002], as shown in Equation (3.3)

$$\rho_{bd}(\omega_o, \omega_i) = \frac{radiance}{irradiance} = \frac{R_o(P, \omega_o)}{R_i(P, \omega_i) \cos\theta_i d\omega_i}, \tag{3.3}$$

where

$$d\omega_i = sin(\theta_i) d\theta_i d\phi_i \tag{3.4}$$

is the differential solid angle.

Therefore, the radiance leaving point $P_{x,y}$ on the 2D radiance map in a particular direction (ω_o) can be expressed as

$$R_o(P_{x,y}, \omega_o) = \rho_{x,y}(\omega_i, \omega_o) R_i(P_{x,y}, \omega_i) \cos\theta_i d\omega_i. \tag{3.5}$$

By image gradient[2], we mean the pixel forward difference, a commonly used gradient approximation. That is to say, in a radiance map R, the gradient field

[2]Strictly speaking, it should be *radiance gradient* in this thesis. We will use both terms synonymously, justified by the fact that the camera response function establishes a one-to-one mapping between the two.

$\nabla R = (\frac{\partial R}{\partial x}, \frac{\partial R}{\partial y})$ is computed as

$$\begin{aligned}\frac{\partial R}{\partial x} &= R_o(P_{x+1,y}, \omega_{o_{x+1,y}}) - R_o(P_{x,y}, \omega_{o_{x,y}}) \\ &= \rho_{x+1,y} R_i(P_{x+1,y}, \omega_{i_{x+1,y}}) \cos(\theta_{i_{x+1,y}}) d\omega_{i_{x+1,y}} - \rho_{x,y} R_i(P_{x,y}, \omega_{i_{x,y}}) \cos(\theta_{i_{x,y}}) d\omega_{i_{x,y}}, \end{aligned} \tag{3.6}$$

and similarly for $\frac{\partial R}{\partial y}$. In Equation (3.6), since $P_{x+1,y}$ and $P_{x,y}$ are adjacent points, according to the smoothness of illumination, we can assume the radiance arriving these two points are equal, *i.e.*

$$R_i(P_{x,y}, \omega_{i_{x,y}}) = R_i(P_{x+1,y}, \omega_{i_{x+1,y}}). \tag{3.7}$$

Thus, from Equation (3.6), we can derive

$$\frac{\partial R}{\partial x} = R_i(P_{x,y}, \omega_{i_{x,y}})(\rho_{x+1,y} \cos(\theta_{i_{x+1,y}}) d\omega_{i_{x+1,y}} - \rho_{x,y} \cos(\theta_{i_{x,y}}) d\omega_{i_{x,y}}), \tag{3.8}$$

and similarly

$$\frac{\partial R}{\partial y} = R_i(P_{x,y}, \omega_{i_{x,y}})(\rho_{x,y+1} \cos(\theta_{i_{x,y+1}}) d\omega_{i_{x,y+1}} - \rho_{x,y} \cos(\theta_{i_{x,y}}) d\omega_{i_{x,y}}). \tag{3.9}$$

And the magnitude and orientation of ∇R are, respectively,

$$|\nabla R| = \sqrt{\left(\frac{\partial R}{\partial x}\right)^2 + \left(\frac{\partial R}{\partial y}\right)^2}, \quad \text{and} \tag{3.10}$$

$$\begin{aligned}\tan\alpha &= \frac{\partial R}{\partial y} \Big/ \frac{\partial R}{\partial x} \\ &= \frac{\rho_{x,y+1} \cos(\theta_{i_{x,y+1}}) d\omega_{i_{x,y+1}} - \rho_{x,y} \cos(\theta_{i_{x,y}}) d\omega_{i_{x,y}}}{\rho_{x+1,y} \cos(\theta_{i_{x+1,y}}) d\omega_{i_{x+1,y}} - \rho_{x,y} \cos(\theta_{i_{x,y}}) d\omega_{i_{x,y}}}, \end{aligned} \tag{3.11}$$

where α is the orientation of radiance gradient.

From these equations, for a given scene, of which the BRDF at each point is fixed, the gradient of reflected radiance in a particular direction has two properties: (a) the gradient magnitude, $|\nabla R|$, depends on both intensity and direction of incoming radiance, R_i, as shown in Equation (3.10); and (b) the gradient orientation, α, ONLY

depends on the direction of R_i, but not its intensity, as shown in Equation (3.11). In particular, the magnitude of radiance gradient varies linearly with magnitude of R_i, and non-linearly with θ_i.

Although the above conclusion is drawn w.r.t. the outgoing radiance in a particular direction (ω_o), it is easy to conclude that the gradient of total radiance leaving the point $P_{x,y}$ also has the same properties. This is because the radiance leaving the surface is the sum over contributions from all incoming directions.

3.3.2 Discussions

There are two special cases where radiance gradient may not change according to illumination.

Homogeneous surface When diffused light strikes on a homogeneous surface, such as a textureless wall or flat table surface, the radiance reflected from the adjacent surface patches will be almost the same. The reasons are: (a) the incident radiance of diffused light are almost the same everywhere; and (b) the BRDF and surface normal of each point are the same for homogeneous material. In this case, no matter how the intensity or direction of incident light are changed, the radiance gradient is always very small and close to zero.

Specular surface When illuminants arrive specular surface in one direction, they are reflected in a small lobe of directions around the specular direction, as described in Figure 3.4. As the result, the radiance leaving along the specular direction is much stronger than those leaving along other directions. For example, in Figure 3.4, the adjacent points P_1 and P_2 receive the same irradiance, but the radiance leaving the points have large difference due to different reflection directions. In such case, the radiance gradients are largely dependent on the shape of specular lobe, and the gradients magnitude are not linear in the intensity of incoming illumination.

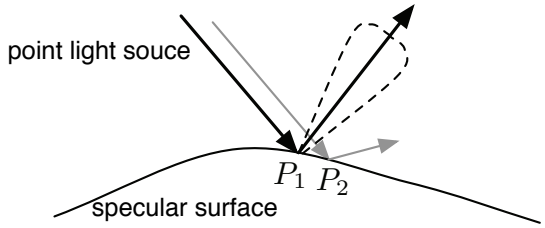


Figure 3.4: The radiance reflection from specular surface can be modeled as a small lobe of directions around the specular direction.

3.4 Gradient Variation across Illumination

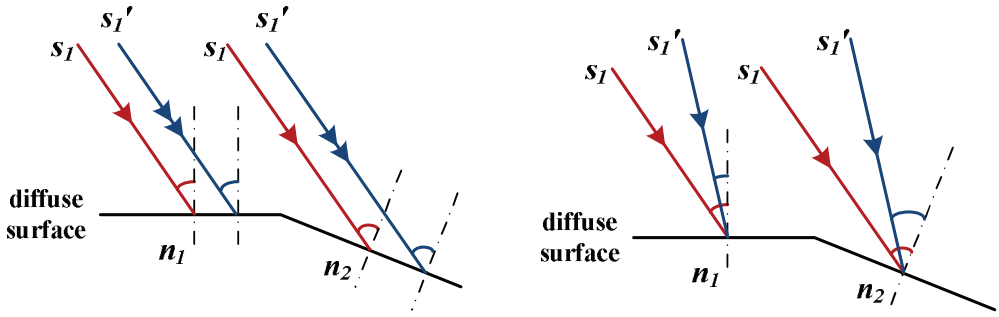
3.4.1 Gradient variation under a single illuminant

Single illuminant can change in both its intensity and direction, thus affecting the incident irradiance, as illustrated in Figure 3.5. From Equation (3.10), it can be seen that the gradient magnitude $|\nabla R|$ is affected by both illumination intensity (*i.e.* the radiance R_i) and incident direction ω_i. And Equation (3.11) shows that the gradient orientation α is dependent only on ω_i but not on incoming radiance.

Equation (3.10) and (3.11) also explain the reason why researchers, in many recent works, assume that image gradients, particularly their orientation, are invariant to illumination changes. In situations in which only the illumination intensity changes, *e.g.* a scene illuminated evenly by diffused lighting, the assumption performs very well because such change only scales the gradient magnitude $|\nabla R|$ without altering the gradient orientation α. In other words, the overall gradient map vary only by a scaling factor.

However, the assumption does not hold when illumination direction also changes. An example of this is direct sunlight, which has both direction and intensity. As illustrated in Figure 3.5(b), when illumination changes from s_1 to s_1', the angles between the illuminant and surface normals of adjacent non-parallel patches changes as well. And this will nonlinearly change the radiance reflected from the patches. As a result, both the magnitude and orientation of the radiance gradient will be affected.

When two varying light sources illuminate a scene, we may easily analyze the



(a) Single illuminant intensity change (b) Single illuminant direction change

Figure 3.5: When illumination changes from s_1 to s_1', the change in the angles between the illuminant and surface normals at adjacent patches will affect radiance gradients.

situation as well. As shown in the next section, and in Figure 3.6 it is the dominant light source that will dictate how the radiance gradients change.

3.4.2 Gradient variation under two illuminants

Photographic scenes that contain two light sources are common enough to warrant closer study. One such situation is when flash is used in a low ambient light environment. In Chapter 5, we will demonstrate how such gradient variation due to flash can be applied to low light photography enhancement. While in this section, without loss of general, by illuminants, we mean general light sources.

Based on the physics of photography, the overall radiance map captured by a camera can be modeled as a sum of radiance maps of different illuminants [Agrawal et al. 2005b]. Suppose we have two radiance maps R_1 and R_2, corresponding two illuminants s_1 and s_2, and let R denote the overall scene radiance captured by the camera. In a static scene with the position and direction of the illuminants fixed, the captured radiance R is:

$$R = R_1 + R_2. \tag{3.12}$$

Since the gradient is a linear operator, the radiance gradient field can be expressed as:

$$\nabla R = \nabla R_1 + \nabla R_2. \tag{3.13}$$

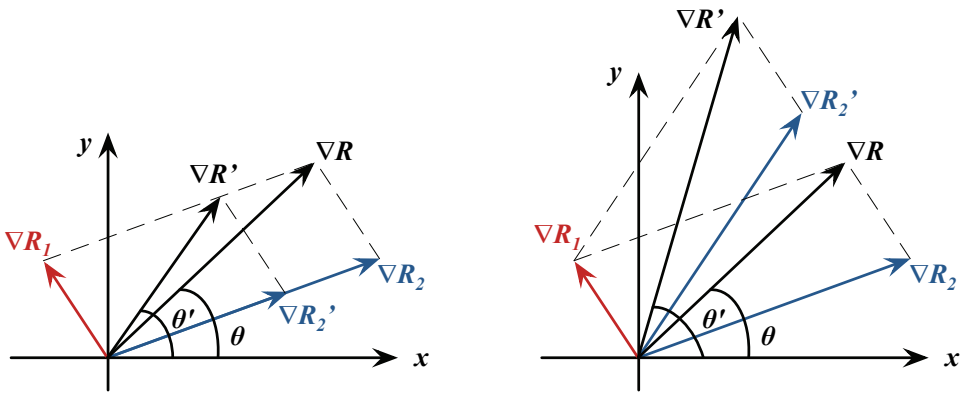
Equation (3.13) shows that the captured radiance gradient is composed of the gradients of R_1 and R_2. Since ∇R, ∇R_1, and ∇R_2 are all vectors, we may easily visualize them with a diagram such as Figure 3.6, where ∇R is the vector sum of ∇R_1 and ∇R_2.

With two illuminants, we can change both of them simultaneously, or only change one of them. In fact, the conclusion of study on changing one of illuminants can be easily propagated to the case of two illuminants variation. Therefore, let us focus on the change of one of the illuminants, and analyze how the captured radiance gradient ∇R will be affected in this situation.

Changing intensity If we change only intensity of s_2 and not its direction, *e.g.* changing from s_2 to s_2', Equation (3.10) and (3.11) say that the gradients ∇R_2 and $\nabla R_2'$ are parallel and differ only in their magnitudes. As for another illuminant (whether directional or not), there is no change. Thus, the gradient ∇R_1 remains unchanged also. The vector diagram for this situation is depicted in Figure 3.6(a). It is clear that the captured radiance gradient will change from ∇R to $\nabla R'$, due solely to the intensity change in the s_2. In fact, both its magnitude and orientation will change, as shown in the figure.

Changing direction In another way, we can change only the direction of s_2 such that its radiance gradient changes from ∇R_2 to $\nabla R_2'$, as shown in Figure 3.6(b). Similarly, both magnitude and orientation of ∇R will change.

To demonstrate the gradient variation under two illuminants, we set up an example, as follows: we use two light sources to illuminate a figurine from two different directions, and we fix the intensity of one light and adjust that of the other one. The photos under two different illumination intensity and corresponding radiance gradient are shown in Figure 3.7. It is very clear that as the intensity of the light on right side increases, the magnitudes of radiance gradient are enlarged and the gradient orientation are shifted. This experiment again, proves that the radiance gradient varies when illumination changes.

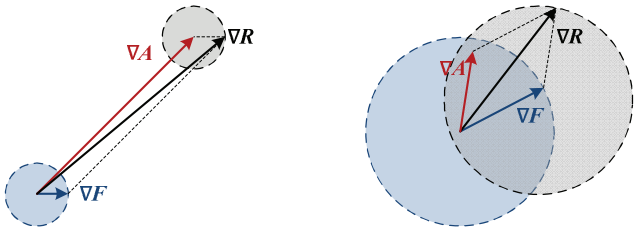


(a) Gradient magnitude variation (b) Gradient orientation variation

Figure 3.6: Illustration of how the scene radiance gradient ∇R is the vector sum of the gradient ∇R_1 and ∇R_2. If the radiance intensity changes from R_2 to R_2', the radiance gradient ∇R will change in both magnitude and orientation.

(a) Weak lighting (b) Strong lightng (c) Gradient map (zoomed in)

Figure 3.7: Two light sources illuminate the figurine from left and right sides. The left light is fixed, while the intensity of the right lights is varied. In the zoomed-in gradient map, blue arrows denote gradient of radiance in (a), and red arrows denote gradient of radiance in (b). When the right light becomes brighter, both the magnitude and orientation of radiance gradients are changed accordingly.



(a) $|\nabla A| \gg |\nabla F|$ (b) $|\nabla A|$ and $|\nabla F|$ are comparable

Figure 3.8: When $|\nabla A| \gg |\nabla F|$, *i.e.* the scene radiance gradient is dominated by the ambient component, flash variation has little impact on the captured radiance gradient. Otherwise, such impact is very large and cannot be ignored.

3.5 Discussions on Gradient Variation

Figure 3.6 and 3.7 also shed light on when it is safe to assume that radiance gradients are invariant to illumination change. Here, we use flash photography with ambient lighting condition as an example to analyze the assumption. Let A denote ambient radiance and F denote flash radiance. Usually, A can be very large (*e.g.* day light) or extremely low (*e.g.* outdoor night), and cannot be changed by photographer. Thus, in flash photography, we normally adjust flash intensity to change the overall lighting condition.

When the ambient light is sufficiently strong, the magnitude of the ambient gradient can be much larger than that of flash gradient, *i.e.* $|\nabla A| \gg |\nabla F|$, it can be seen that the radiance gradient is hardly affected by the change in flash gradient. This situation occurs when changing the flash intensity contributes little to the strong ambient light. As illustrated in Figure 3.8(a), the blue circled region illustrates the variation range of flash gradient, and the gray region stands for the variation range of radiance gradient. It can be seen that when $|\nabla A|$ is dominant the composed ∇R can be assumed to be the same with ∇A as they are very close to each other. However, in low-light conditions when $|\nabla A|$ is comparable to, or much less than, $|\nabla F|$, we cannot ignore the change in radiance gradients. Such situation is dictated in Figure 3.8(b): the variation range of ∇R in fact remains the same due to the same ∇F, but compared with ∇A, the variation of ∇R cannot be ignored any more.

In the next two chapters, we demonstrate how we can apply these mathematical

insights to enhance photographs in different scenarios. In Chapter 4, by matching histogram of gradient magnitude from visible image to near infrared image, we successfully enhanced photographs captured in high dynamic scene. In Chapter 5, we present a novel method that can cleanly decompose the ambient and flash gradients, and finally enhance low-light photography for the whole or selected parts of an image.

Chapter 4

Enhancing Photographs with Near Infrared Images

4.1 Overview

By studying the image gradient variation under single illuminant, we know that the change of illumination intensity will affect the gradient magnitude only. In other words, by manipulating image gradient magnitude, we may adjust illumination intensity thus improving the brightness contrast of the captured scene radiance. Inspired by this idea, in this chapter, we introduce a novel method to enhance a photograph by using the contrast and texture information of its corresponding near infrared image.

Near Infra-Red (NIR) images of natural scenes usually have better contrast and contain rich texture details that may not be perceived in visible light photographs (VIS). To exploit these good characteristics of NIR, we first decompose the NIR/VIS pair into average and detail wavelet subbands. We then transfer the contrast in the average subband and transfer texture in the detail subbands. Transferring contrast, one of the most important steps in our method, is done by matching distribution of gradient magnitude in VIS to that in NIR. We built a special camera mount that optically aligns two consumer-grade digital cameras, one of which was modified to capture NIR. Compared with tone-mapped HDR images, our results exhibit higher visual quality. This work has been published in CVPR'08 [Zhang et al. 2008].

4.2 Background and Related Work

The radiance from natural scenes usually spans a very wide dynamic range, far exceeding what a digital camera can capture. For instance, in a sunny outdoor environment, the dynamic range could reach as high as 10^9. In contrast, a professional-grade digital camera that uses 14 bits per channel can capture a range of only 10^4. Consumer-grade cameras are even worse. One common technique around this problem is to first compute an high dynamic range (HDR) image, usually from multiple shots of varying exposures, and then to map this into a lower dynamic range (LDR) image suitable for display devices. However, such a tone-mapping procedure does not usually produce a perceptually pleasing result. Usually, pixels end up becoming too bright or too dark, and rich scene information such as color and texture are almost completely lost. Figure 4.1(a) shows a typical photo taken under an HDR environment, where the footpath is very bright but the region inside the building can barely be seen.

(a) Visible Image (b) Near Infrared Image (c) Our Enhanced Result

Figure 4.1: We proposed a novel image enhancement method by transferring contrast and texture from near infrared image to visible image. Figure 4.1(a) and Figure 4.1(b) are an improper exposed photo taken under high dynamic range environment and its corresponding near infrared photo. With only these two input images, our approach can adaptively and automatically adjust contrast and enrich visible details in over- or under-exposed areas, as shown in Figure 4.1(c).

We have reviewed many recent methods of recovering the HDR image and tone-mapping techniques in Chapter 2. In contrast, our method uses Near Infrared (NIR) light. This lies between visible red light and Long Infra-Red (LIR) light in the electromagnetic spectrum. NIR light has wavelength in the range 750 – 1400 nm, which is longer than visible light (380 – 750 nm). Human eyes can not see NIR

light but most digital cameras can sense it very well. For example, some models of SONY digital cameras or camcorders have a *Night Shot* mode which increases cameras visual range by letting the sensor acquire more NIR light. However, most manufacturers insert an IR cutoff filter over the camera sensor to filter out NIR light, to avoid some unwanted artifacts. In fact, NIR images usually have better brightness contrast and provide rich texture details, as seen in Figure 4.1(a) and 4.1(b). The details of trees and leaves are barely seen in the visible image, but look clear and sharp in the NIR image. We exploit this fact in our work.

Inspired by the camera's ability to record NIR light and by recent works on tonal transfer [Bae et al. 2006; Neumann and Neumann 2005; Fredembach and S usstrunk 2008], we propose a novel method that can adjust a photograph's contrast adaptively and enrich texture details fully automatically with just one shot (*i.e*. one VIS/NIR image pair). NIR photography is not new; it is commonly appreciated for its artistic value [Maher and Berman], but has not been fully exploited in computational photography. Morris *et al*. observed that the wavelet coefficients of long infrared (LIR) natural images closely follows the Laplacian distribution [2007]. Encouraged by their work, we build a dual-camera system that can capture visible photo and NIR photo of the same scene simultaneously, and find that NIR images also have similar statistical properties. Moreover, we notice that NIR images of natural scenes usually exhibit lower dynamic range and contain rich texture details. In terms of transfer techniques, the Neumann brothers showed how to transfer color style from a source image to an arbitrary target image by applying histogram matching [2005]. Similarly, Bae *et al*. also presented a method to transfer tonal quality from one image to another, using histogram matching and bilateral filter in [2006]. In the light of their work, we propose an original way of using NIR information to enhance visible photographs. Given as input one VIS image and its corresponding NIR image, our approach can adaptively detect unsatisfactory pixels in the VIS image, and transfer contrast information and high frequency texture from its NIR counterpart. We use histogram matching in the gradient domain to transfer contrast and use wavelet coefficients to transfer texture information. We were able to achieve very pleasing results (see Fig.4.1).

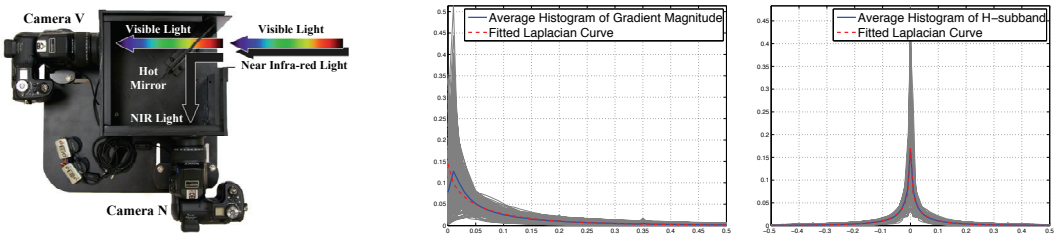
4.3 Near Infrared Imaging

4.3.1 Dual-camera system

NIR light lies adjacent to visible red light in the electromagnetic spectrum, and has longer wavelength than visible light. NIR is not visible to human eyes, but can be recorded by CCD or CMOS sensors. However, most manufacturers of digital cameras install an IR cutoff filter over the sensor to suppress infrared light and avoid unwanted artifacts.

To capture both visible and NIR pictures for the same scene simultaneously, we built a dual-camera system which comprises two Sony F828 digital cameras and one hot mirror. A hot mirror is a specialized dielectric mirror which can reflect NIR light when incident light arrives at a certain angle. We used a 45° hot mirror, meaning it can reflect NIR light with angle of incidence of 45° but does not block visible light. Figure 4.2(a) illustrates how our system works. Although the Sony F828 has built-in *Night Shot* mode which can temporarily move the IR cutoff filter away to allow NIR imaging, Sony has intentionally limited such NIR imaging to only allow long exposure times. Our modified camera does not suffer from this limitation. We also modified the remote control of the camera so that it can trigger two cameras at the same time. We have carefully setup two cameras to ensure that they are optically aligned. They also share the same camera settings, such as focal length and aperture size, to guarantee the geometric alignment of the image pair. Currently, we do not force the two cameras to use the same shutter speed, because digital cameras are designed to be less sensitive to NIR thus requiring a slightly longer exposure.

The NIR picture captured in this way is actually an RGB color image and looks reddish since NIR light is just adjacent to red light. However, because of the filters we use, the NIR light we capture is almost monochromatic and should not contain any color information. So we use only intensity information by converting to *HSV* color space and using V channel. Fig.4.1(a) and Figure 4.1(b) show an example image pair captured by our dual-camera system. Our prototype hardware may look bulky, but this can be miniaturized. Our goal is to show the usefulness of NIR images.



(a) Dual-camera System (b) Gradient Magnitude (c) Wavelet Coefficient H

Figure 4.2: (a): Our VIS-NIR dual-camera prototype. Camera V and N are optically aligned and connected to the same remote control, allowing a VIS/NIR image pair of the same scene to be captured with a single shot. (b-c): Statistical properties of NIR images. (b) shows distribution of gradient magnitude, similar to statistics of visible images [Huang and Mumford 1999]. (c) shows distribution of H wavelet subband of Haar transform, similar to statistics of IR images [Morris et al. 2007]. Subbands V and D have similar distribution.

4.3.2 Statistics of NIR images

Huang and Mumford [1999] have shown that the gradient histograms of natural images follow a generalized Laplace distribution which can be expressed as Equation 4.1:

$$P(x) = k \cdot e^{-|x/s|^{\alpha}}. \tag{4.1}$$

Recently, Morris *et al*. [2007] found wavelet coefficients of LIR (wavelength lies in 4000 – 120000 nm) images of natural scenes can also be well fitted with a Laplacian curve. In this paper, we show that NIR natural images share similar statistical properties, as illustrated in Figure 4.2. We collected a total of 220 NIR photos for statistical analysis. Some of them are collected from the web and others are captured by ourselves, mostly covering subjects of natural scenes and people. Similar to [Huang and Mumford 1999; Morris et al. 2007], we use gradient magnitude and the Haar wavelet coefficients. We calculate the histograms of all images for both gradient magnitude and wavelet coefficients in horizontal (H), vertical (V) and diagonal (D) directions. All these histograms are calculated on logarithm of the actual values and normalized based on image pixels. In Figure 4.2, all gray lines denote the actual histograms, the blue lines show the average histogram distribution, and the red dash lines show the fitted Laplacian curve (Equation 4.1). We can see that the fit is good, meaning that NIR images have similar statistical

properties as visible and LIR images (Please refer details to [Huang and Mumford 1999; Morris et al. 2007]). In Sec.4.4.3 we will show how we can use these statistical properties to guide the enhancement process.

4.4 Visible Image Enhancement

4.4.1 Workflow

The workflow of our approach is illustrated in Figure 4.3. There are three main steps:

Computing the weighted region mask We calculate a weighted mask based on saturation and brightness of the input visible image. The weight value assigned for each pixel will be incorporated in enhancement of next two steps. The details of this step are introduced in Section 4.4.2.

Transferring contrast To transfer contrast from NIR to VIS, we manipulate gradient magnitude of visible image so that it has similar brightness distribution to that of NIR. We achieve this via matching histograms of gradient magnitude. More details are introduced in Section 4.4.3.

Transferring texture The NIR image may reveal many rich details that lack in its visible counterpart. We enhance those detail textures in VIS by combining high frequency subbands after applying Haar wavelet transformation. The details are introduced in Section 4.4.4.

Note that all inputs are the logarithm of original image values as we mentioned in the previous section.

4.4.2 Computing the weighted region mask

Intuitively, regions that suffer a loss of details are typically too bright or too dark, and have low saturation. From this observation, a weighted mask can be calculated based on saturation and brightness value. Let W_s and W_v denote weighted mask

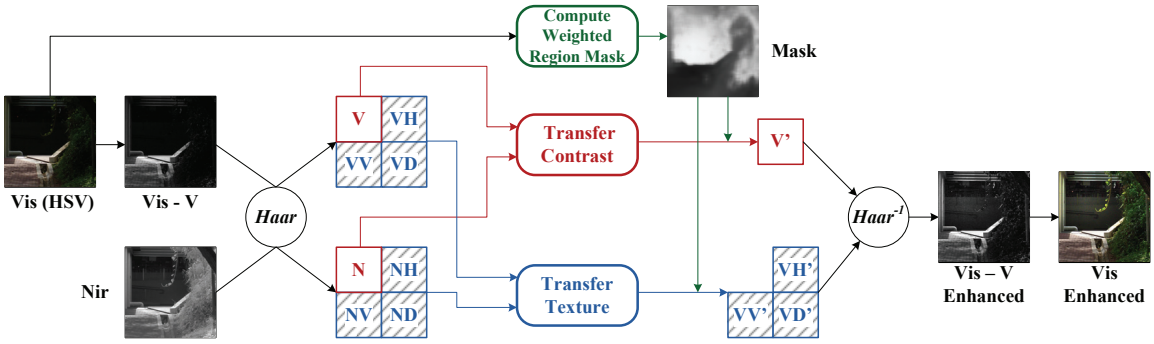


Figure 4.3: The workflow of our approach. The enhancement process uses the Haar wavelet decomposition, and comprises three major steps: computing the weighted region mask, transferring contrast, and transferring texture. See text for details.

of saturation and brightness, W denote the final weighted region mask indicating areas to be enhanced. Then W can be obtained using following equations:

$$W_s = 1 - e^{-p_s|s-1|}, \quad p_s \in [0,1],\ s \in [0,1] \tag{4.2}$$

$$W_v = 1 - e^{-p_v|v-0.5|}, \quad p_v \in [0,1],\ v \in [0,1] \tag{4.3}$$

$$W = W_s \cdot W_v\,, \tag{4.4}$$

where s and v are the saturation and brightness intensity, and p_s and p_v denote the probability that s and v appear in visible image respectively. p_s and p_v can be easily obtained from the normalized histograms of channels S and V. The meaning of p_s and p_v is that the pixels to be enhanced should distribute over large areas, rather than in small regions. Enhancing large areas while ignoring small regions usually achieves better perceptual quality.

Note that W is calculated adaptively and fully automatically, not requiring any thresholds. The weighted region mask is used as a mask for brightness and texture transfer later. Figure 4.4 shows what our weighted region mask looks like and how it is generated. A higher value in W means more information will be transferred from NIR image, and vice versa. To reduce noise, a Gaussian blurring is first applied on W.

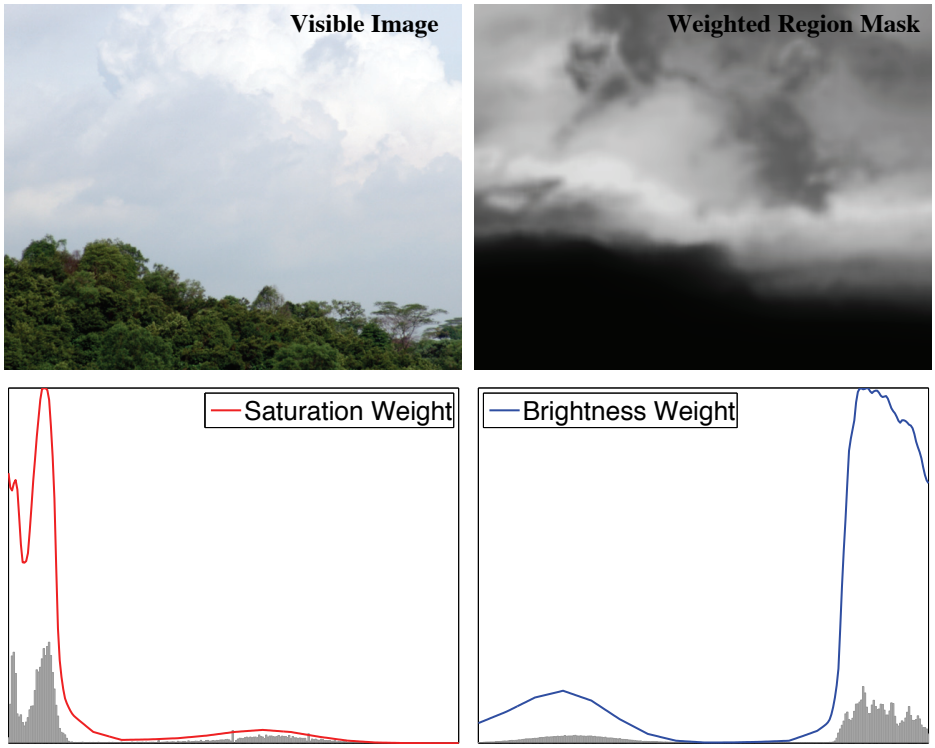


Figure 4.4: Weighted region mask computation. The second row shows the histograms of saturation and brightness channel (gray regions). The red and blue lines illustrate the weights computed according to Equation 4.2. The curves are smoothed for reducing noise. The sky and cloud have high brightness and relatively low saturation, and thus have higher weights; vice versa for the tree tops.

4.4.3 Transferring contrast

Our contrast transfer method is based on histogram matching and gradient techniques. Bae *et al*. [2006] also applied histogram matching to transfer photograph tonality. Instead of matching histogram in the intensity domain as they did, we show that histogram matching in the gradient magnitude can achieve better and reliable results.

Histogram matching

The histogram matching problem can be simply defined as: given an image I and a target histogram (PDF) $h(z)$, the problem is to find a new image J by transforming I, so as to make histogram of J be as same as h. The problem can be solved by using the cumulative distribution function (CDF), f. Define $f(x) = \int_0^x h(z)dz$, where x is image intensity. Let I_{ij} and J_{ij} denote each pixel intensity in I and J. Then the desired image J can be obtained using Equation 4.5, and the detailed proof can be found in [Gonzalez and Woods 2002]. In this work, we applied histogram matching on gradient magnitude, instead of image intensity.

$$J_{ij} = f_J^{-1}(f_I(I_{ij})) \tag{4.5}$$

Large-Scale contrast transfer

The brightness contrast of a visible image is affected by environment illumination, as well as object shape and texture in the scene. Therefore, the brightness map of an image should change smoothly while preserving major features such as strong edges. To achieve a smooth brightness map of visible image V and NIR image N (V and N are actually the average subbands in the Haar decomposition, as shown in Figure 4.3), we apply bilateral filtering [Tomasi and Manduchi 1998] to decompose images to large-scale layer and detail layer, and use the larger-scale layer as brightness map, as in Equation 4.6:

$$\begin{aligned} V_L &= bf(V),\ V_D = V - V_L \\ N_L &= bf(N),\ N_D = N - N_L. \end{aligned} \tag{4.6}$$

V_L and N_L are large-scale layers, and V_D, N_D are corresponding detail layers (after taking the logarithm). We use a similar definition for the bilateral filter function bf and parameter selection as in Bae *et al.*'s work.

We implement three different methods to transfer contrast from the NIR image to the VIS image. A comparison of their results can found in Figure 4.5 and 4.6.

Method 1: Histogram Matching Inspired by Bae *et al.*'s method [2006], we can simply match intensity histogram of V_L with N_L to transfer intensity distribution. This method is easy and efficient, but histogram matching blindly alters pixel values and thus very possibly destroy illumination consistency. From Figure 4.5, we see that histogram matching does improve the contrast significantly. However, we also see that pixels in the tree bark are over brightened and inconsistent with the illumination in the original image. After applying the gradient constraint, the result looks more natural.

Method 2: Histogram Matching with Gradient Constraint Based on our analysis in Chapter 3, if we blindly adjust image intensity without being aware of gradient variation as Method 1, we may easily falsely change the gradient orientation, thus breaking the illumination consistency. Therefore, to maintain such consistency, we can check the gradient direction of the altered brightness map pixel by pixel. Once we find the gradient direction that is reversed or changed too much from the original brightness map, we force them to be zero. After applying the gradient constraint, the enhanced result looks more natural compared with method 1 (see Figure 4.5). But in some cases, where gradients change abruptly along their original directions due to the histogram matching step, this constraint will fail, as shown in Figure 4.6. The gradient constraint cannot remove the banding-effect on the pillar and wall, because the gradients in those areas are not actually reversed.

Method 3: Gradient Magnitude Matching Method 1 fails to maintain the illumination consistency, while Method 2 cannot warrant smoothness of radiance variation. The reason why these two methods do not work well is that the image intensity manipulation in both of them ignores the relationship between gradient variation and illumination change. In this case, the ambient illumination can be treated as a single illuminant. Thus adjusting image brightness

equals to changing illumination intensity, which will only affect the gradient magnitude. Therefore, a better solution for this case, as shown in Method 3, is matching the histogram of brightness gradient magnitude instead of brightness intensity. We define V_G and N_G as the gradient magnitude of V_L and N_L:

$$\begin{aligned} V_G &= \sqrt{V_{G_x}^2 + V_{G_y}^2} = \sqrt{(\frac{\partial V_L}{\partial x})^2 + (\frac{\partial V_L}{\partial y})^2} \\ N_G &= \sqrt{N_{G_x}^2 + N_{G_y}^2} = \sqrt{(\frac{\partial N_L}{\partial x})^2 + (\frac{\partial N_L}{\partial y})^2}. \end{aligned} \tag{4.7}$$

In Section 4.3.2 we have shown that gradient magnitude histogram of NIR image can be well fitted with a generalized Laplacian curve. Because N_L is a smoothed version of the NIR image, its gradient magnitude N_G also has same statistical property. Let l denote the Laplacian curve that can fit histogram of N_G. Instead of matching histogram of V_G with histogram of N_G directly, we use l as the target histogram to produce a smoother and noise-free distribution transfer. In this case, the functions f_I and f_J in Equation 4.5 are the CDFs of l. Let $V_{G'}$ denote the histogram matching result, we can easily compute new gradients by scaling V_{G_x} and V_{G_y} along their original directions respectively:

$$\begin{aligned} V_{G'_x} &= \frac{V_{G'}}{V_G} \cdot V_{G_x} \\ V_{G'_y} &= \frac{V_{G'}}{V_G} \cdot V_{G_y}. \end{aligned} \tag{4.8}$$

From $V_{G'_x}$ and $V_{G'_y}$, we reconstruct new large-scale brightness map $V_{L'}$ by using Agrawal *et al.*'s improved Poisson solver [2006b]. The final contrast transferred V' is obtained by blending enhanced brightness map and its original version V together using alpha-blending

$$V' = W \cdot (V_{L'} + V_D) + (1 - W) \cdot V, \tag{4.9}$$

where the weighted map W is used as the alpha channel and $|\cdot|$ denotes pixel-wise multiplication. This method naturally maintains illumination consis-

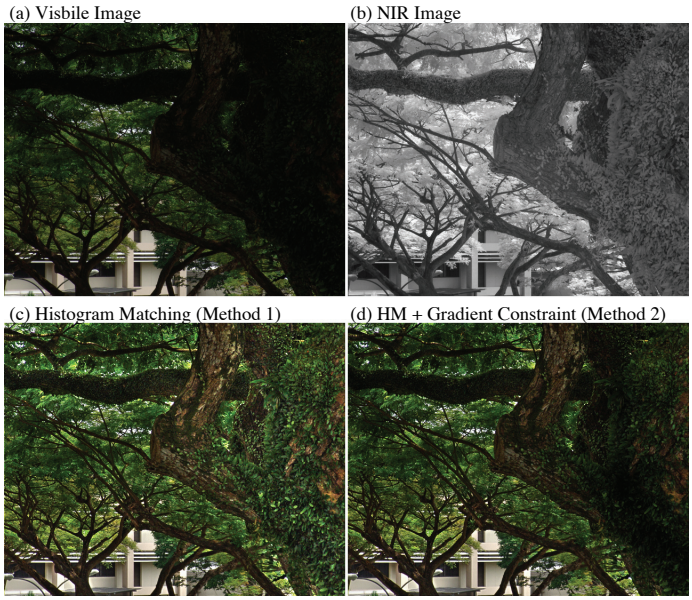


Figure 4.5: Comparison of histogram matching (method 1) with histogram matching with gradient constraint (method 2). The result of histogram matching in (c) looks artificial because it breaks the consistency of overall brightness distribution. After applying gradient constraints, (d) looks more natural.

tency and achieves the best result among these three methods. See Fig.4.6 for comparison. Note that the banding-effects of methods 1 and 2 are completely suppressed, while overall contrast has been improved.

Since the final brightness is the alpha-blending of the original brightness and the histogram-matched brightness using mask W as alpha channel, the contrast transferring is in fact achieved by local histogram matching. Our final result therefore, will not exhibit global brightness shifting.

4.4.4 Transferring Texture

As we state in our workflow (Fig.4.3), after applying Haar wavelet transformation, the wavelet subbands in horizontal, vertical, and diagonal directions actually contain rich texture information. To transfer those details, we use alpha blending

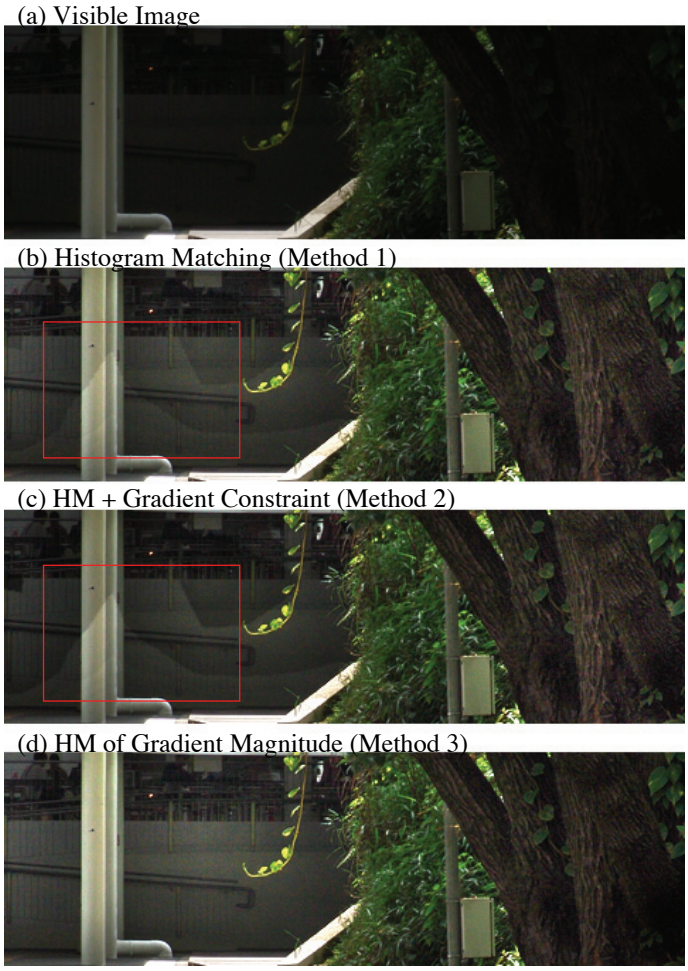


Figure 4.6: Comparison of Methods 1, 2, and 3. Note that the band-effects (regions in red box) due to blind histogram matching have been successfully suppressed by our guided histogram matching of gradient magnitude. The result of Method 3 achieves the least artifacts and best perceptual quality.

again to combine corresponding subbands together:

$$VH' = W \cdot NH + (1 - W) \cdot VH. \tag{4.10}$$

VV' and VD' are obtained similarly. The new subbands VH', VV', and VD' not only inherit texture details from the VIS image, but are also enhanced by rich high frequency details from NIR image. Figure 4.7(g) show the result with high frequency details transferred. The textures on the roof in the original image are almost lost completely. By transferring high frequency details from NIR to visible image, those lost textures are successfully recovered, and those weak textures are also reinforced greatly. Finally, we apply inverse Haar wavelet transform to enhance the V channel.

Please note that blending such high frequency components may introduce local gradient reverse, since brightness gradients of NIR can be very different from the one of VIS. In this case, our texture transferring approach should not be applied.

4.5 Experiments and Results

A common HDR scene is the natural outdoor environment under bright sunlight. To demonstrate the strength of our techniques, we test our approach with pictures taken in such HDR situations. All input visible and NIR image pairs have been geometrically aligned. In outdoor daylight, tree leaves and some objects, such as cloth and skin, reflect NIR light strongly, so they look bright and have much details even in shaded areas (see Figure 4.1, 4.9, 4.10, and4.8). Such features in NIR images are useful for enhancing visible images.

We also find that contrast transfer and texture transfer are both equally important for enhancement. As shown in Figure 4.7: Figure 4.7(e) is the enhanced result with only texture transferred, where most of roof details are successfully recovered but the picture still looks over-exposed; Figure 4.7(f) is the result with only contrast transferred, which has lower contrast but roof details are still lost. Obviously, after transferring the contrast and texture, Figure 4.7(g) exhibits better visual quality.

To show that histogram matching of gradient magnitude (Method 3) can preserve overall illumination map and achieve higher perceptual quality, we compare

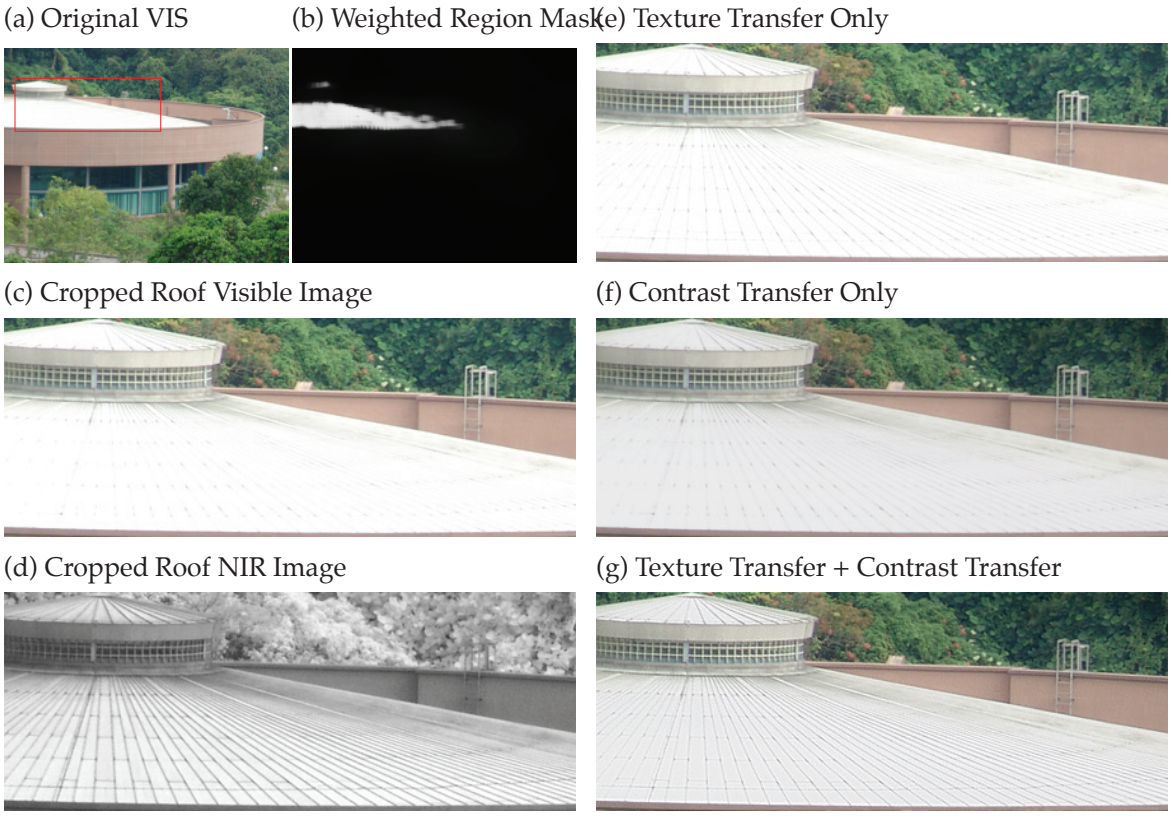


Figure 4.7: Comparison of results with either contrast transfer or texture transfer. c-g show the zoomed-in roof details.

of our results with a naively blended output (see Figure 4.9(d) and Figure 4.10(d)). This trivial result is obtained using alpha-blending of V and N based on the weighted region mask W, *i.e.* using each pixel value in W as alpha value.

We also compare our results with tone-mapped HDR images of same scenes, as shown in Figure 4.9(h) and Figure 4.10(e). To get the HDR image, we take multiple images with different exposure (usually 5-7 pictures with exposure difference of 1 stop), and assemble them using the HDR Shop software developed by Debevec [Debevec]. We recovered the camera response curve of our camera to generate HDR image more precisely, and we applied Reinhard *et al.*'s algorithm for tone mapping [Reinhard et al. 2002]. Tone-mapped HDR images are supposed to produce a range-compressed image with rich details and high visible quality. However, tone mapping algorithms usually have a strict assumption that the scene must be static, *i.e.* no moving objects throughout the whole image sequence. Such an assumption is easily broken in outdoor scenes, as shown in Figure 4.9(h). The walking pedestrian and leaves waving in the wind cause serious "ghosting effect" (shown in red boxes) in the tone mapped results. Because the inputs of our approach are captured in a single shot, our results are free of such artifacts. Besides, our approach preserves consistency of overall illumination distribution while recovering scene details, therefore our results gain better perceptual quality on brightness contrast than tone-mapped results, as shown in Figure 4.9(h) and Figure 4.10(e).

4.6 Chapter Summary

In this chapter, we presented an approach of enhancing visible photograph using NIR information based on a dual-camera prototype. Without manual segmentation or interaction, our method can calculate the enhancing weight for each pixel automatically and transfer brightness and texture details from the NIR image to the visible image. We show that our histogram matching of gradient magnitude can well maintain large-scale illumination and achieve better perceptual quality. Also, by combining the wavelet H, V, and D subbands of NIR and visible image high frequency details are effectively enhanced.

The highlights of our contribution include:

(a) Visible Image

(b) Near Infrared Image

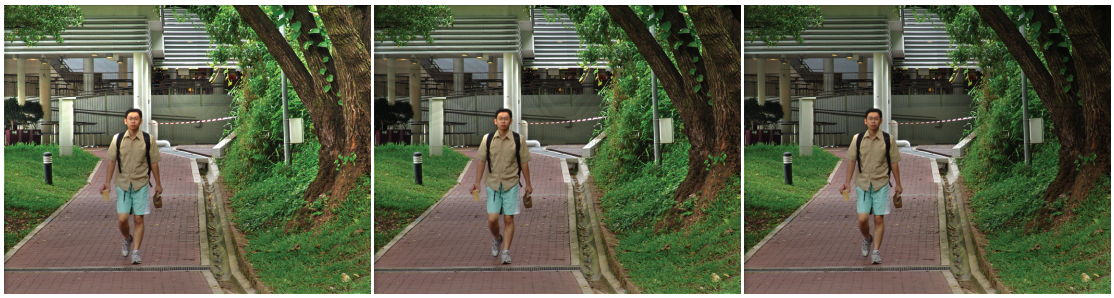
(c) The Weighted Mask

(d) Our Enhanced Result

Figure 4.8: 4.8(a) shows a typical landscape photograph captured in HDR scenario. The forest region is obviously too dark and details of those trees are barely seen. 4.8(b) is the NIR counterpart, showing better contrast. 4.8(c) and 4.8(d) are computed weighted mask and our enhanced result. The overall visible quality becomes much better: the contrast is enhanced and the rich details in Forest region are revealed.

- Our method can adaptively enhance the VIS image by transferring the contrast and details from its NIR counterpart. Thus, our approach can be widely used in various scenarios where near infrared light is sufficient.
- The three major steps in our approach — computing weighted mask, transferring contrast and transferring details — do not require any thresholds and proceed fully automatically without any human interaction.

(a) Visible Image (b) NIR Image (c) Weighted Region Map



(d) Method 1 (e) Method 2 (f) Method 3

(g) Alpha-blending (h) HDR Tone Mapping

Figure 4.9: Comparison of our approach with alpha-blending and HDR tone mapping. The naive alpha-blending result appears bad, since simple pixel-wise blending cannot transfer overall contrast. As for HDR tone mapping, the result exhibits "ghosting effect" (shown in red boxes) because objects have moved during the capture of multiple exposures.

(a) Visible Image (b) NIR Image (c) Weighted Region Map

(d) Alpha-blending (e) HDR Tone Mapping (f) Our Result (Method 3)

Figure 4.10: Another comparison of our approach with alpha-blending and HDR tone mapping. Our method successfully enhances brightness and texture by using NIR information.

- Our method does not incorporate with any machine learning routines, thus is time efficient.

Since common digital cameras are capable of recording NIR light, ours is a practical method to enhance LDR images. As far as we can tell from searching the published literature, we are the first to successfully make use of NIR information for photograph enhancement. Compared with tone mapping an HDR image, our method requires only one shot (*i.e.* one VIS/NIR image pair). Benefiting from the better contrast in the NIR images and a spatially varying weighted mask, our method can produce results that are aesthetically more pleasing than tone-mapped methods.

The comparison of Method 1, 2, and 3 (see Figure 4.5 and 4.6) also testify our theory in Chapter 3. Our histogram matching on gradient magnitude successfully enhances the brightness contrast while maintaining illumination consistency. We intentionally avoid changing gradient orientation in this application because we assume the ambient illumination is a single illuminant and we only need to change its intensity. In next chapter, we will apply our theory in a more complicated problem: enhancing flash photography in low light condition, in which two illuminants are involved: ambient lighting and flash.

Chapter 5

Selective Re-flashing

5.1 Overview

In the previous chapter, we have demonstrated how we can manipulate gradient magnitude to change image brightness contrast when the ambient illuminant is dominant and stable. In such lighting conditions, researchers in many recent works assume that image gradient, particularly its orientation, is invariant across illumination. Unfortunately, this assumption is not true when the ambient illuminant is comparable to, or much less than other illuminants. A common scenario of such a case is flash photography in low light conditions. Based on our analysis in Chapter 3, the captured radiance gradient will change greatly, in both magnitude and orientation, as flash output varies. Can we make use of such gradient variation to enhance the flash photography? The answer is: Yes.

In this chapter, we present a novel method, namely Selective Re-flashing, that allows users to selectively re-light parts of the image with any amount of flash illumination. In order to do this, our method first separates the ambient and flash-only scene components from just two input flash photographs, and then re-renders an image using different contribution of the two components. We achieve this by exploiting the finding that both magnitudes and orientations of image gradients vary with illumination. Compared to existing works, our method does not require any long-exposure ambient image, and yet successfully suppresses various kinds of flash artifacts, such as shadows, specularities and inter-reflections, in the result.

5.2 Low light photography

Photographs captured in low-light scenarios, such as outdoor night scenes, romantic candlelight dinners, or even simple indoor portraits, are very likely degraded by noise or motion blur. The reason is that the sensors in most digital cameras are not sensitive enough to handle such low-light conditions. A popular way to compensate for weak ambient lighting is flash photography. Unfortunately, flash light usually ruins the ambient illumination, creates harsh shadows and specularities in the scene, and causes uneven exposure. Experienced photographers try to overcome this by taking a series of pictures (usually 3 shots) using different flash intensities, and then selecting the best one. The flash intensity is determined either automatically by the camera, or based on the photographer's experience. This approach is called flash exposure bracketing (FEB) [Canon].

In this paper, we re-invent FEB in a computational way: by taking two flash pictures with different flash intensities, we can recover the ambient and flash-only components of the scene, and then re-render selected parts of the scene with different amounts of ambient and flash lighting. In other words, users can, with minimal input, freely adjust the contribution of both the flash and ambient illumination to selected parts of the image based on their preferences. For example, in Figure 5.1, the human subject is re-lit to be more visible while preserving the ambient shadow region of the hand. Our goal is to provide a powerful, yet simple, low-light photography solution allowing photographers to capture pictures with high visual quality using fast shutter speed.

In Chapter 2, we have reviewed many computational approaches, which enhance low-light photography based on multi-modal imaging [Agrawal et al. 2005b; Ben-Ezra and K.Nayar 2004; Eisemann and Durand 2004; Miao and Sim 2005; Petschnigg et al. 2004; Yuan et al. 2007]. These methods either take a flash/non-flash image pair or multiple images as inputs, and require both the scene and camera to be static. Furthermore, a non-flash photo taken by a common digital camera in low-light conditions needs a long exposure time or a high ISO setting, which in turn leads to motion blur or a low signal-to-noise ration (SNR). In contrast, our approach requires only two flash pictures captured successively with fast shutter speed. Thus the input images are not degraded by noise and our method is

less reliant on the static scene assumption. Moreover, note that none of the existing work permit selective re-flashing of parts of the image. The entire image is re-lit as a whole. In contrast, our method permits the user, with a few simple sketches, to re-flash selected parts of the scene. This is useful because the user may wish to preserve the ambient illumination in certain parts of the scene (*e.g.* the region near a candle in a romantic dinner scene) while brightening the rest.

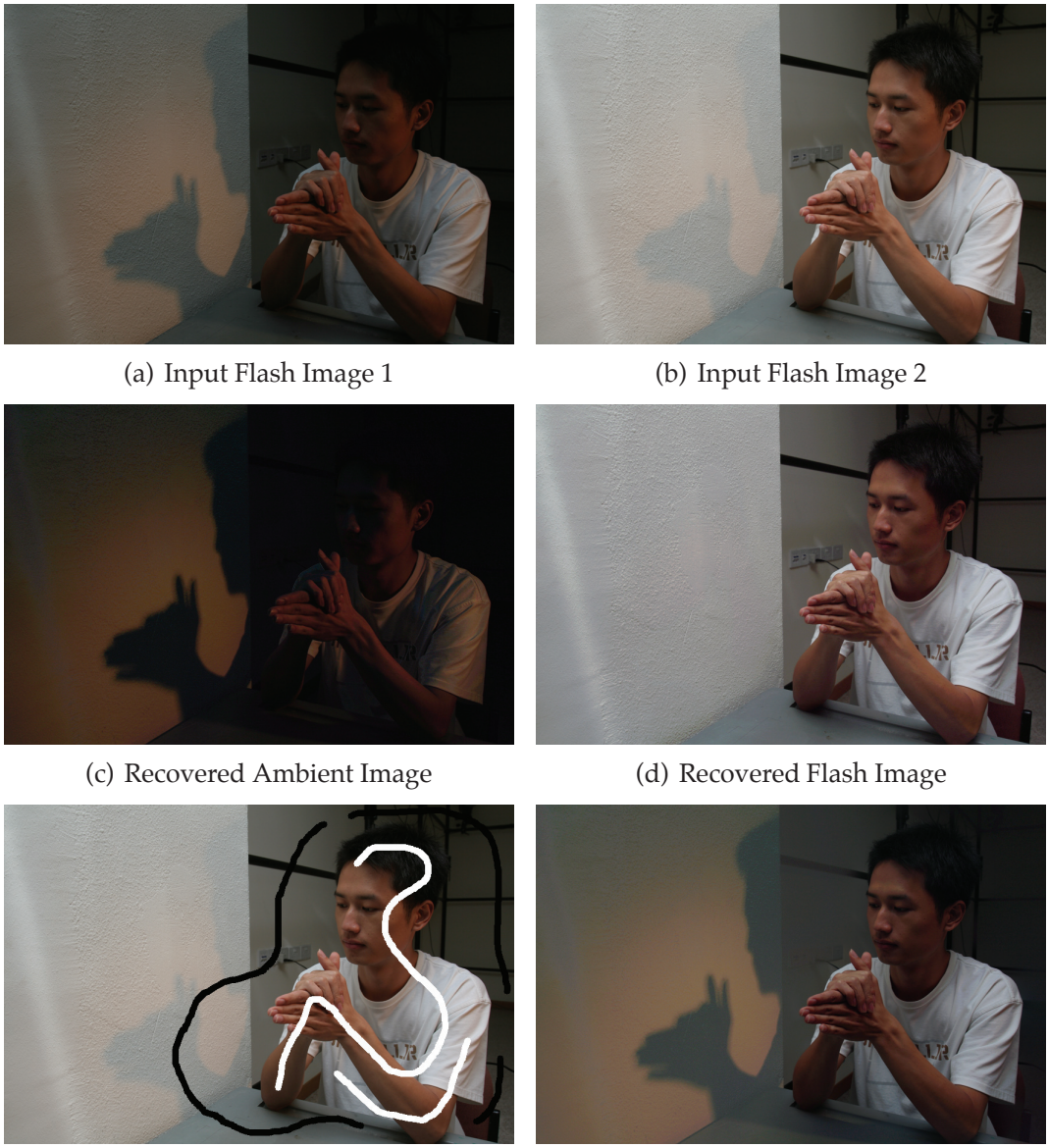
5.3 Computational Re-flashing

For flash photography in a low-light environment, it is always difficult to balance between ambient light and flash light. The purpose of the flash is to provide an artificial illuminant so that the scene becomes bright enough to be captured by the camera. Yet flash light ruins the ambient lighting and the mood associated with that ambiance. So while the user wants the scene to be well lit for a good photo, the user also wants to preserve the ambient lighting as much as possible. Based on our analysis in the Chapter 3, we now propose a novel gradient decomposition method which allows the user to selectively re-flash parts of a scene, while preserving the ambient lighting in other parts. Our method requires just two flash images, taken with the same exposure setting but with different flash intensities. No long exposure or high ISO ambient image is required.

Our major assumptions are:

- That ambient illumination is constant between the two shots.
- Both the camera and scene objects are static.
- The position and orientation of the flash are fixed; only the flash intensity is changed.

These assumptions are widely made in many multi-modal imaging approaches. And although we require the scene to be static, this is not a severe limitation because both images can be captured quickly within a short interval of each other.



(a) Input Flash Image 1 (b) Input Flash Image 2

(c) Recovered Ambient Image (d) Recovered Flash Image

(e) User Marking (f) Re-flashed Result

Figure 5.1: We propose a novel Selective Re-flashing method for low-light photography. From just two input photos captured using different flash intensities, our method first recovers the ambient and flash-only scene components, and then allows the user to select, with a few quick sketches, which parts of the image to re-light, and which to preserve the ambient illumination. The user also decides how much to re-light. Finally, our method renders the output image, according to the user's specifications, via a fast algorithm. In our re-flashed result above, note that the hand shadows produced by the ambient illumination is well preserved, while the human subject is re-lit to be more prominent.

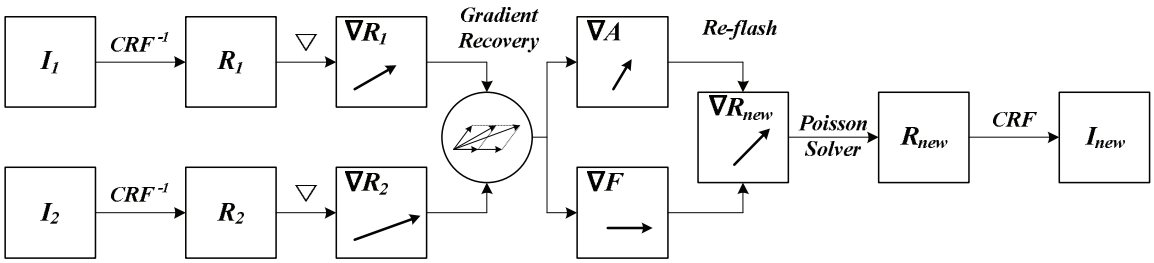


Figure 5.2: Our workflow of radiance gradient recovery and image re-flashing.

5.3.1 Workflow

The entire workflow of our radiance recovery and image re-flashing method is summarized in Figure 5.2. The main steps are summarized as follow:

1. Capturing two flash images (spatially aligned) with different flash intensity
2. Mapping the intensity of two images to radiance with inverse camera response function and computing radiance gradients (pixel forward difference)
3. Decomposing radiance gradients
4. Selectively composing radiance gradients
5. Reconstructing radiance with Poisson solver and mapping radiance to re-flashed image.

5.3.2 Decomposing the radiance gradient

A typical flash discharge is a rapid burst of light energy followed by a slower decay, and the total discharge time is significantly less than the camera exposure time. Thus F is the radiance map produced by total amount of flash output. Denote the ratio of the two flash intensities by r. Since the radiance map F is linearly scaled by flash intensity, we have $F_2 = rF_1$, as shown in Figure 5.3.

Now it is easy to derive:

$$\begin{aligned} \nabla F_2 &= r\nabla F_1 \\ \nabla R_2 - \nabla R_1 &= \nabla F_2 - \nabla F_1 = (r-1)\nabla F_1. \end{aligned} \tag{5.1}$$

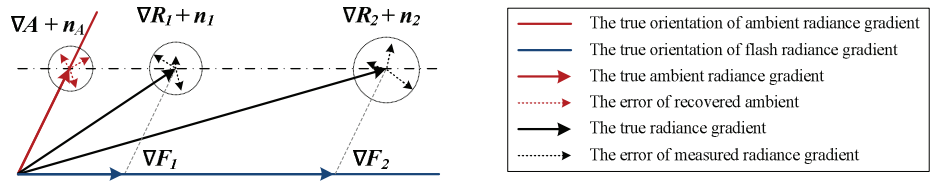


Figure 5.3: Decomposition of the radiance gradient by changing the flash intensity by a ratio $r = \nabla F_2 / \nabla F_1$. Additive Gaussian noise is used to model the uncertainty in our measurements.

Therefore,

$$\nabla F_1 = \frac{1}{r-1}(\nabla R_2 - \nabla R_1). \tag{5.2}$$

Given ∇F_1 and ∇F_2, ∇A can be recovered using

$$\nabla A = \nabla R_1 - \nabla F_1 = \frac{r\nabla R_1 - \nabla R_2}{r-1}, \tag{5.3}$$

In our method, the flash need not be calibrated, and the absolute flash intensity (in lumens) is also not required. We only need to know r, the ratio between two flash intensities. Once ∇A, ∇F_1, and ∇F_2 are solved, we can then reconstruct the ambient and flash radiance by integrating these gradient fields subject to the usual potential field constraint:

$$\frac{\partial^2 R}{\partial x \partial y} = \frac{\partial^2 R}{\partial y \partial x} \quad . \tag{5.4}$$

The standard way to solve this integration is via the Poisson Solver, and in this paper we use the method described in [Agrawal et al. 2006b].

Note our method can be easily extended to three or more input flash images. In this case, Equation (5.3) will need to be modified to compute ∇A as a least-squares solution in terms of the radiance gradients ∇R_i and the flash intensity ratios r_i. As before, these multiple input images can, in principle and with the right hardware, be acquired in a short period of time, so that object motion is not really an issue. Using three or more inputs is likely to reduce overall noise, but at the expense of more computation.

5.3.3 Determining the ratio of flash intensities

In practice, as shown in Figure 5.3, the measured ∇R_1 and ∇R_2 are affected by noise which we characterize by a zero mean Gaussian distribution:

$$\widetilde{\nabla R_1} = \nabla R_1 + n_1 \tag{5.5}$$
$$\widetilde{\nabla R_2} = \nabla R_2 + n_2. \tag{5.6}$$

Here, n_1 and n_2 are independent 2D random vectors each following a Gaussian probability density function (pdf). We assume both pdfs have zero means, but different covariance matrices. We further assume that the noise in both dimensions are independent, leading to diagonal covariance matrices. Thus we may write: $n_1 \sim N(\mathbf{0}, \mathbf{C_1})$ and $n_2 \sim N(\mathbf{0}, \mathbf{C_2})$, where

$$\mathbf{C_1} = \begin{bmatrix} \sigma_{1x}^2 & 0 \\ 0 & \sigma_{1y}^2 \end{bmatrix}, \ \mathbf{C_2} = \begin{bmatrix} \sigma_{2x}^2 & 0 \\ 0 & \sigma_{2y}^2 \end{bmatrix}, \tag{5.7}$$

and $\sigma_{1x}^2, \sigma_{1y}^2, \sigma_{2x}^2, \sigma_{2y}^2$ are unknown constants. As we will soon show, there is no need to estimate the values of these constants.

From Equation (5.3), the recovered gradients of ambient radiance can be expressed as:

$$\begin{aligned} \widetilde{\nabla A} &= \frac{r\widetilde{\nabla R_1} - \widetilde{\nabla R_2}}{r-1} \\ &= \frac{r\nabla R_1 - \nabla R_2}{r-1} + \frac{rn_1 - n_2}{r-1} \\ &= \nabla A + n_A, \end{aligned} \tag{5.8}$$

Therefore, to accurately recover ∇A, we need to find the optimal r that can minimize the magnitude of the noise term n_A. Because r is the ratio of two different flash intensities, we may add the constraint $r > 1$, *i.e.* the second flash intensity is always higher than the first. From our assumptions, it is straightforward to show that n_A follows a Gaussian pdf, *i.e.* $n_A \sim N(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = (\frac{r}{r-1})^2\mathbf{C_1} + \frac{1}{(r-1)^2}\mathbf{C_2}. \tag{5.9}$$

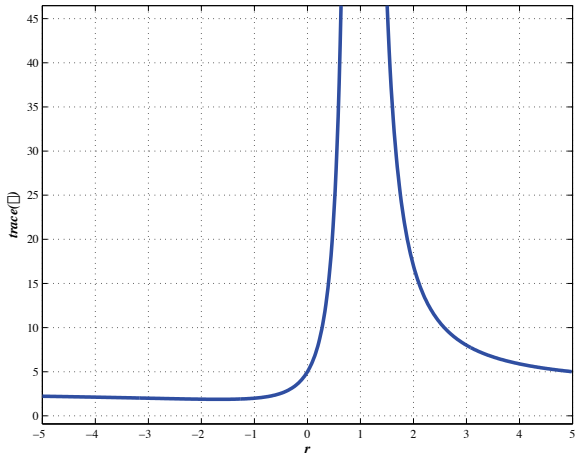


Figure 5.4: Plot of $trace(\mathbf{\Sigma})$ versus r. To minimize the curve, r should be as large as possible.

Because the mean is zero, we may equivalently minimize $trace(\mathbf{\Sigma})$. As illustrated in Figure 5.3, we want to find an optimal r to minimize the "radius of uncertainty" of n_A. Thus:

$$r = \arg\min_r \{trace(\mathbf{\Sigma})\}, \ r > 1 \tag{5.10}$$

$$= \arg\min_r \left\{ \left(\frac{r}{r-1} \right)^2 S + \frac{1}{(r-1)^2} T \right\}, \ r > 1 \tag{5.11}$$

where $S = \sigma_{1x}^2 + \sigma_{1y}^2$ and $T = \sigma_{2x}^2 + \sigma_{2y}^2$.

The plot of $trace(\mathbf{\Sigma})$ versus r is shown in Figure 5.4. Different values of σ_{1x}^2, *etc.* do not affect the general shape of the curve. Note that the curve goes to $+\infty$ at $r = 1$, and asymptotically approaches S as $r \to +\infty$. It can also be proved that the curve achieves its minimum when $r = -T/S$, but this is a negative value and thus not valid (the flash intensity ratio cannot be negative). Thus, we want r to be as large as possible, *i.e.* the second flash intensity to be much higher than the first. This is true whatever the actual values of the unknown variances σ_{1x}^2 *etc.* are.

A simple way to maximize the ratio of flash intensity is to take the first photo without flash, and the second one with maximum flash output. However, the flash intensity is also limited by practical concerns. First, if the flash intensity is too low and ambient illumination is also very weak, then the radiance will be dominated

by CCD and read out noise which leads to a low SNR [Holst 2006]. Second, if the flash is too strong, many parts of the scene will be over-exposed (pixel values will saturate), leading to an inaccurate radiance map R.

In our experiments, we choose the flash intensity such that the SNR of the captured photo is high enough, and the whole image is properly exposed. By "good SNR", we mean the SNR value should be at least 5, according to Bushberg *et al*. in [Bushberg et al. 2002].

5.3.4 Lower bound of ambient recovery

In theory, our method can recover the ambient component no matter how weak it is, but in practice the amount of recovered ambient radiance is limited by the camera's dynamic range and image quantization noise. As shown in Figure 5.5 [1], the camera response curve is relatively flat when radiance is low (= negative log radiance). This region is also affected by quantization noise. Therefore, when the ambient radiance is too weak, recovery is easily overwhelmed by quantization noise. Ideally, we want a lower bound below which our method will not work.

However, it is not easy to find such a lower bound in general, since the dynamic range and quantization noise is hardware dependent. For example, in our experiment, when the average of ambient image intensity is as low as 20 (the intensity range is $0 \sim 255$), our approach can still achieve reasonably good recovery. From Figure 5.5, the log radiance value corresponding to intensity of 20 is -4. And usually the average intensity of flash images [2] is 170, corresponding to 1 approximately in radiance logarithm. The ratio of ambient and flash radiance is about 1:32, which also means that our method can achieve high quality ambient image with a shutter speed that is nearly 5 stops faster.

[1] The camera response curve shown in this figure is in fact the same with the one in Figure 3.2. We plot log of radiance instead of normalized radiance just for better illustration in later.

[2] We always adjust flash exposure settings so that all pixels intensities lie in the middle to upper brightness range.

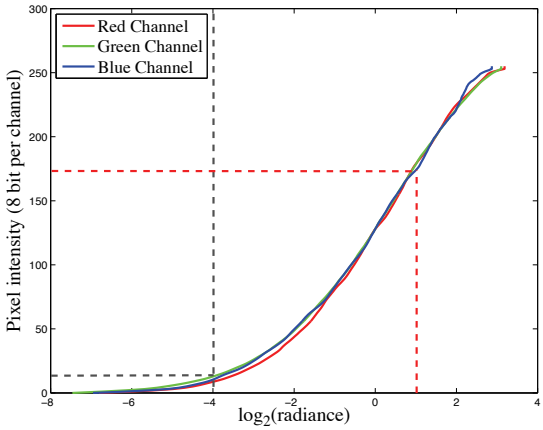


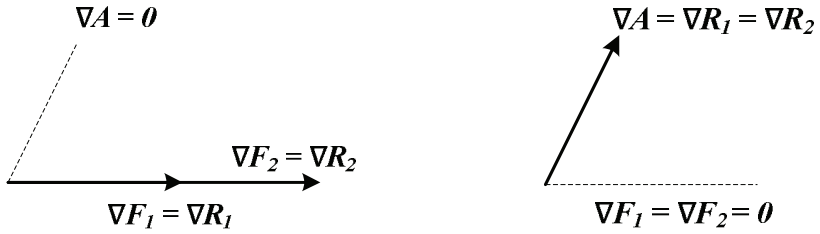
Figure 5.5: The color lines describe the camera response curves of our Canon 350D DSLR camera, and the gray dashed line illustrates the lower bound for ambient recovery. This means if the ambient log radiance is below -4, the recovered ambient component will be degraded by quantization noise. Note that the log radiance values (horizontal axis) are not absolute, but depend on the camera exposure settings. For example, enlarging the aperture will shift the curves to the right.

5.3.5 Recovering shadow regions

Typically, there are two kinds of shadows in flash photographs: those caused by the ambient illumination, and those caused by the flash. Unless the flash and ambient illuminants are co-linear, these shadow regions are unlikely to be one and the same. Moreover, an ambient shadow will be completely filled with flash light, especially when the flash intensity is high. Thus, in such regions the ambient radiance is zero and the captured radiance R equals flash radiance F. This in turn forces their gradients to be equal, as shown in Figure 5.6(a). On the other hand, a flash shadow receives no light from the flash, but only from the ambient illuminant. Thus, the captured radiance is solely due to the ambient radiance, which in turn forces their gradients to be equal. This is shown in Figure 5.6(b).

We may use the above insights to easily identify and recover the shadow regions. In Figure 5.6(a), ∇R_1 and ∇R_2 have the same orientation and their magnitude ratio is also r. From Equation (5.1), (5.2), and (5.3), we can easily derive that $\nabla R_1 = \nabla F_1$, $\nabla R_2 = \nabla F_2$ and $\nabla A = 0$, meaning that flash dominates the overall radiance. In Figure 5.6(b), we have $\nabla R_1 - \nabla R_2 = \mathbf{0}$, thus we know that $\nabla F_1 = \nabla F_2 = 0$, and

simply get $\nabla A = \nabla R_1 = \nabla R_2$.



(a) Ambient shadow (b) Flash shadow

Figure 5.6: In both ambient and flash shadow regions, ∇A and ∇F can still be recovered successfully.

5.3.6 Correcting specularities

Like shadows, specularities may be caused by either the ambient lighting or the flash. It is unlikely that both types of specularities will coincide, since specularities are highly sensitive to the exact placement of the camera, object, and light source. Ambient specularities appear as equally bright regions in both input images, since they are unaffected by the flash. These regions are also equal in size. In contrast, flash specularity regions grow in size with increasing flash intensity, because pixels surrounding a specular spot also receive more light and become saturated. This distinction allows us to quickly detect and distinguish between the two types of specularities.

Once detected, flash specularities should be corrected because these were not present in the low-light scene. To do so, we employ standard image inpainting algorithms. This approach is similar to existing methods which replace specular pixels with ambient information from their neighbourhood. We use neighboring pixels because the ambient radiance of specular pixels has been completely corrupted by over-exposure. Figure 5.8 shows an example of flash specularity detection and correction in an ambient image. In this paper, we use A. Telea's inpainting algorithm [Telea 2004] to remove repair specularities.

(a) User Marking (b) Weighted Mask

Figure 5.7: Users easily select their desired regions for re-flashing by drawing broad lines as shown in 5.7(a). White lines denote regions to be enhanced using flash, while black lines denote regions where ambient radiance should be preserved. After marking, a mask is automatically generated using a graphcut algorithm. Note the mask is intentionally blurred by a simple Gaussian for a smooth blending of the gradients inside and outside the mask.

5.3.7 Selective image re-flashing

Once the gradients of ambient component and flash component are recovered, we can combine them together by a simple weighted sum. The weights, which control the contribution of each component, may be chosen arbitrarily by the user. The combined gradient field in then integrated via the Poisson Solver to obtain the radiance map. Finally, we map the radiance values to pixel values by means of the camera response function (CRF). The entire workflow of our radiance recovery and image re-flashing method is summarized in Figure 5.2.

Instead of re-lighting the entire image, we can go a step further to allow the user to choose which parts of the image to re-flash, and which to retain the ambient illumination. To do this, the user quickly sketches (or scribbles), which regions to re-flash (marked in white), and which to preserve the ambiance (marked in black). Our scribble-based marking is similar to that in Photomontage by Agarwala *et al*. [2004]. From these simple markings, a mask is automatically generated using a graph-cut algorithm [Kwatra et al. 2003]. The boundary of the mask is intentionally blurred to smoothly blend the gradients inside and outside the mask. Figure 5.7 shows this process.

5.4 Experiments and Results

In our experiments, we use the Canon 350D DSLR camera and Canon Speedlite 580EX flash as to capture all the pictures. The Speedlite 580EX has a flash bracketing exposure function which allows us to take up to 3 photos successively with different flash output. Unless otherwise stated, all pictures are captured with exposure settings of ISO 100, F8.0, and a white balance for flash. Such settings are usually desired for taking photos with low-noise and a wide depth of field, and are very suitable for low-light photography. The ratio of flash intensities in the two input images is $r = 4$. Finally, our camera response function is pre-calculated using HDR Shop [Debevec].

5.4.1 Recovering and re-flashing an ambient image

One of the most interesting applications of our method is to recover and re-flash ambient images. Figure 5.1 already shows that in low-light conditions, traditional flash photography is not able to illuminate the scene while preserving ambient radiance. Note that the hand shadows are washed out by the flash in both the inputs. However, they are well preserved in the recovered ambient image, and completely suppressed in the flash-only image. With a few quick scribbles, the user indicates that the human subject should be re-lit with flash, while the hand shadow should be preserved under ambient lighting. The result of our method is a visually pleasing enhanced image that preserves the shadows while brightening the human subject.

Another example is shown in Figure 5.8, which demonstrates the effectiveness of our method at handling complex scenes that contain light sources, ambient and flash specularities. By comparing the Recovered Ambient Image with the Flash Specularity Correction image, it is clear that our method successfully detects and suppresses flash specularities. Moreover, in the Recovered Flash Image (which contains the flash-only radiance), the ambient specularities caused by the desk lamp and ceiling lights on the chrome sphere do not appear, as to be expected.

We also compared our ambient recovery with the ground truth and results of Miao and Sim's method [2005], demonstrating our method can faithfully recover

(a) Input Flash Image 1

(b) Input Flash Image 2

(c) Recovered Flash-only Image

(d) Recovered Ambient Image

(e) Detected Flash Specularities

(f) Recovered Ambient Image with Specularity Correction

Figure 5.8: These images show the effectiveness of our method at handling complex scenes that contain light sources, ambient and flash specularities. Our method successfully detects flash specularities and corrects for them in the Recovered Ambient Image. Moreover, in the Recovered Flash-only Image, the ambient specularities caused by the desk lamp and ceiling lights on the chrome sphere are absent, as to be expected.

ambient image and is robust to noise. In the experiment shown in Figure 5.9, the two input flash images are taken with exposure time of 1/30 sec, and the ground truth of ambient image is taken in 1/10 sec. To match the exposure of the ground truth, we re-rendered our ambient radiance using 3 times of recovered intensity. For Miao and Sim's result, we re-rendered the ambient image with $t_\alpha = 1/10$sec. Comparing the two recovered ambient images and the ground truth, it is obvious that our recovery exhibits less noise defects, more accurate color than Miao and Sim's, thus presenting closer visual quality to ground truth.

Our method also outperforms the Joint Bilateral Filter (JBF), the essential technique used in [Petschnigg et al. 2004] and [Eisemann and Durand 2004]. To compare, we use the flash/no-flash [3] image pair as inputs. Similar to Petschnigg *et al.*'s work, we extract the large scale layer of the no-flash image and detail layer of flash image using bilateral filter, and combine these two layers together based on the same mask as ours (Figure 5.7(b)). We also combine the flash color layer which is decoupled using Eisemann and Durand's method. The inputs and results of JBF are shown in Figure 5.10. As can be seen, the color and details of the subject's body are improved but the face region is still too dark. Moreover, while the ambient noise is reduced, the fine wall texture details have been smoothed away. Compared to our results in Figure 5.1, our reflashing technique achieves higher visual quality.

5.4.2 Outdoor night photography

In outdoor night photography, it is extremely difficult to take a good picture that has a high SNR, even exposure and no motion blur. A typical scene is like this: foreground subjects are usually poorly lit, while background objects are at a distance and exhibit some interesting illumination. To capture a sharp image of the foreground subjects, we turn on the flash and use a short exposure. But flash often ruins the ambient mood, while a fast shutter speed causes the background to be too dark.

Our method solves this problem in a smart way: by brightening (*i.e.* multiplying by a scalar) the recovered ambient radiance, and by re-flashing the foreground

[3] The flash image is the same with our input. The no-flash image is obtained by prolonging exposure time, and NOT the ambient image recovered using our method.

Figure 5.9: Our method can faithfully recover the ambient image. In overall, our result exhibits more accurate color than Miao and Sim's. In details, our ambient recovery presents higher visual quality w.r.t. shadow separation (note the region in the red circle) and noise suppression (note the region in the blue circle).

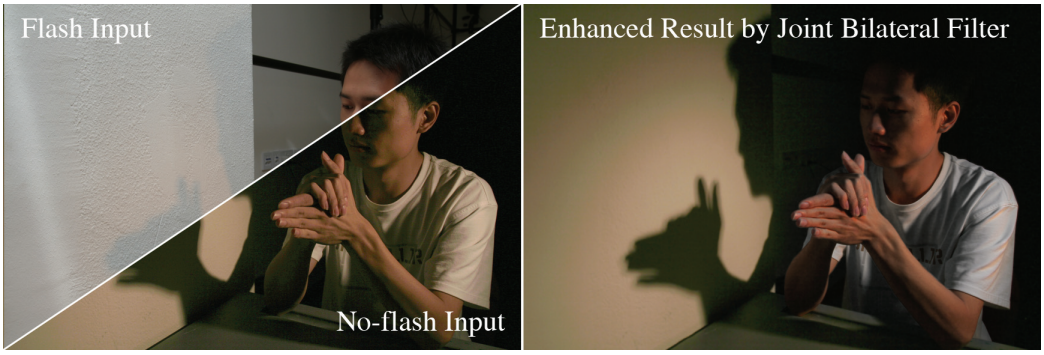


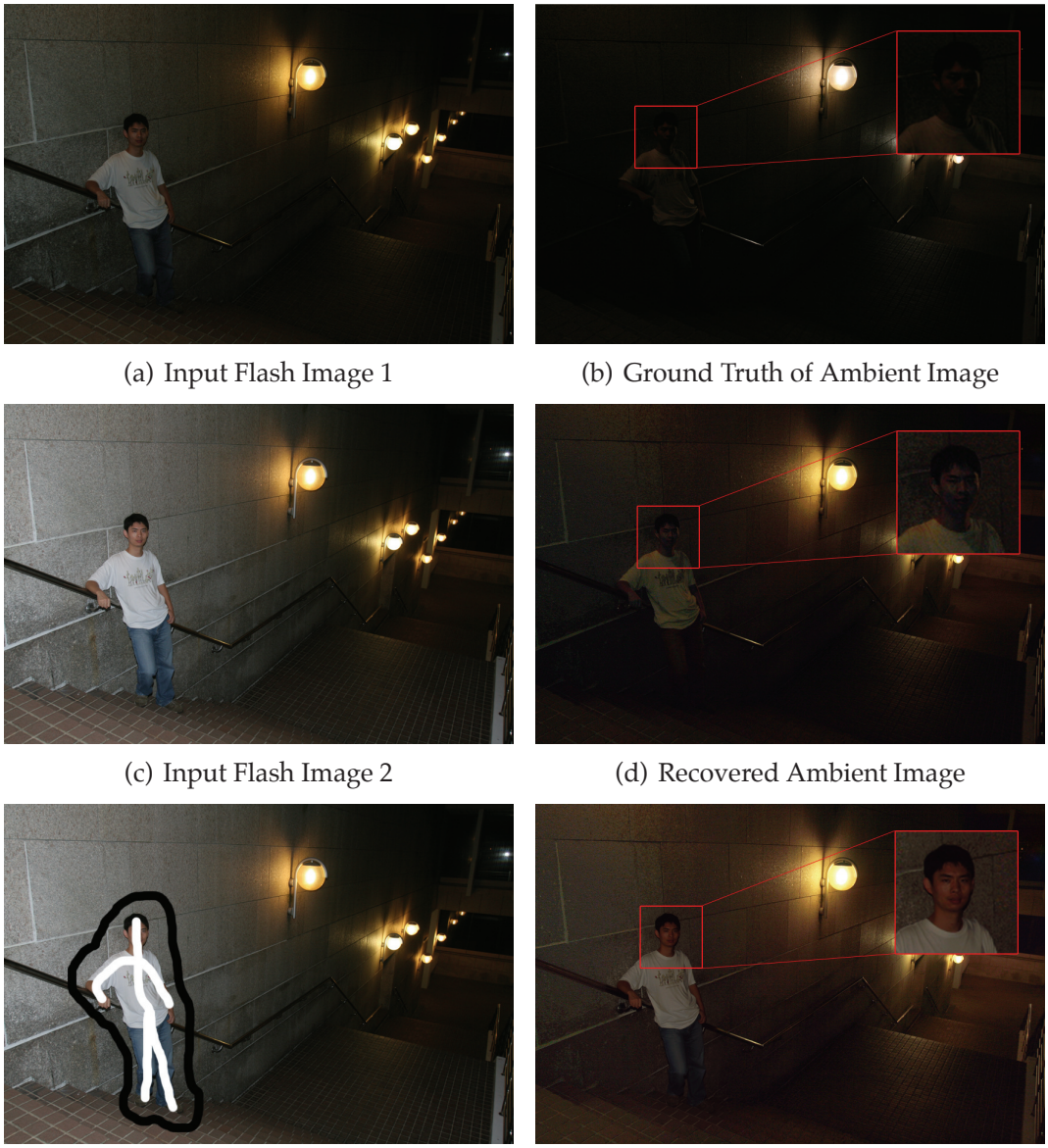
Figure 5.10: Comparison with the result from a flash/no-flash image pair using Joint Bilateral Filtering. The JBF enhancement is achieved using a routine similar to that in [Petschnigg et al. 2004] and [Eisemann and Durand 2004].

subjects (see Figure 5.11). Although the foreground subject is standing in a very dim environment and can barely be seen, our re-flashed image is visually pleasing with a naturally re-lit subject against a brightened ambient background.

5.4.3 Separating inter-reflections

Inter-reflection is a common artifact in flash photography, especially in scenes containing shiny surfaces near objects. To illustrate this problem, we set up a simple experiment as shown in Figure 5.12. We position a desk lamp on the left-front side of a ball with diffused surface to provide ambient illumination. We also point our flash towards the ceiling and use the bounced light as the second illuminant. We then place a mirror on the desk to reflect light onto the ball, thus creating inter-reflections.

In the two input flash images, we can see that both the ambient light and bounced flash are reflected to the bottom of ball, producing an overlaid inter-reflection artifact. However, even for such a complex scene, our method correctly separates the inter-reflection due to flash versus those due to ambient illumination (see the bottom row of Figure 5.12). Compared to existing works, for example, [Nayar et al. 1993; Nayar and Gong 1992; Tan et al. 2003], we do not need a set of polarized images to detect inter-reflections, nor do we require knowledge of the scene geometry, nor any user inputs. No special processing is necessary to handle inter-reflections as our model correctly separates them.



(a) Input Flash Image 1 (b) Ground Truth of Ambient Image

(c) Input Flash Image 2 (d) Recovered Ambient Image

(e) User Marking (f) Re-flashed Result

Figure 5.11: Enhancing outdoor night photography: This challenging scene is solved in a smart way — by combining a brightened ambient image with the re-lit foreground subject. The zoomed-in shots of the subject's head shows that our re-flashed result has a higher SNR than the recovered ambient component. It is also brighter and clearer than the ground truth ambient image. The texture on the wall behind the subject is clearly visible. In this example, we can even tune the color of the flash radiance, so that the re-lit subject exhibits a hue similar to that in the ambient background.

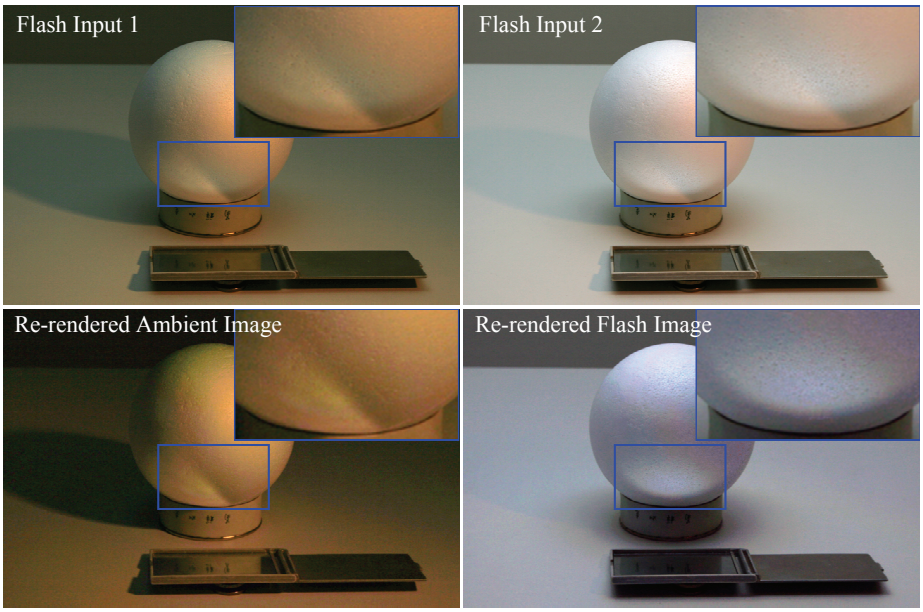


Figure 5.12: Top row: two input flash images showing an overlaid inter-reflection artifact on the bottom of the ball. Bottom row: these show the clean separation of inter-reflection caused by the ambient illuminant from those caused by the flash.

5.5 Limitation and discussions

Our method suffers from some limitations. First, to compute gradient variation, we require that both input images be pixel-aligned. But this is unlike other flash/no-flash techniques. If the total exposure time is too long, any mis-alignment due to scene or camera motion could increase the error in the recovered gradients. But compared to flash/no-flash methods, capturing two successive flash images can be performed very quickly, thereby mitigating this problem. In theory, we can capture both flash images with just a single burst of flash, by accurately timing the camera sensor(s) with the decay of the flash burst. This is a hardware issue which we plan to tackle in the future. Using a single flash also conserves battery power, and spares the human subject from the annoyance of enduring multiple flashes.

Second, our specularity correction algorithm may fail in cases where specular regions are very large in both the inputs. Since the gradient information in such regions are irrecoverably lost, the original color and texture cannot be recovered faithfully. One way to tackle this problem is to employ image inpainting with user

hinted structure information, such the work of [Sun et al. 2005].

5.6 Chapter Summary

It is difficult to capture high quality photographs when the ambient light is weak. Based on the mathematical relationship between radiance gradient and illumination studied in Chapter 3, we present a novel method that can faithfully recover the ambient and flash-only radiance components from just two flash photographs. Our method is flexible in that we allow users to re-flash selective regions instead of the whole image. The intensity of the re-flashing can also be controlled.

This work, again, proves that gradient variation is key of enhancing photographs across illumination. Unlike other researchers, we do not ignore such gradient variation, but make use of illumination change to decompose radiance gradient. Our gradient decomposition method is efficient yet simple, and can neatly separate the radiance gradient field of the flash and ambient components of a scene. Based on that, we also introduce a way to fuse them together on parts of the image, selected with minimal user input, thereby creating novel re-flashing effects. Finally, our method naturally handles specularities, shadows and inter-reflections, producing results of high visual quality.

Chapter 6

Conclusions and Future Work

6.1 Summary

Our goal in this thesis was to find a key to enhancing digital photography under extreme illumination situations, such as HDR and low light environment. Under those situations, capturing and representing satisfactory photographs are very difficult due to the limited sensing capability of digital cameras. In Computational Photography — an emerging area across computer vision and graphics — many approaches have been proposed, but none of them solved the problems completely. Our research revealed the fact that the radiance gradient variation according to the illumination change, could provide important clues to the problem. Based on this idea, we introduced solutions for photography enhancement in HDR and low light environment, respectively.

In the theory chapter, we mathematically showed how radiance gradient changes under the variation of single or two illuminants. We pointed out that the illumination variation does not only change the gradient magnitude, but also the gradient orientation. However, in many recent works, researchers often assumed that the radiance gradient is invariant to the illumination change. Therefore, we further discussed the conditions under which such assumption holds and when not. Such discussion naturally leads to our solutions presented in Chapter 4 and 5.

First, we presented a novel solution to enhance photographs in HDR scenes by transferring contrast and textures from their corresponding near infrared photos.

The key idea was to manipulate image gradient magnitude, which not only adjusts local brightness contrast but also preserves overall illumination consistency. We built a dual-camera prototype that can capture both the visible image and its near infrared counterpart in a single shot. Thus our method does not depend on the static scene assumption. Moreover, the whole process of our approach is adaptive and fully automatic. Compared to the simple alpha blending and HDR + Tone-mapping, our results exhibit higher visual quality.

Second, we introduced an approach that can recover and re-flash ambient radiance from two flash photographs. The input flash photos are taken with different flash intensities, thus providing the gradient variation of the captured radiance. By exploiting such gradient variation, we successfully decomposed gradient field of the ambient and flash-only radiance, allowing users to selectively re-flash the ambiance. Compared to the existing flash/no-flash techniques, our method can effectively suppress many artifacts such as noise, cast shadows, interreflection, specularities, *etc*. Moreover, as the two flash inputs can be taken quickly and within a very short time interval, our approach is less dependent on the static scene assumption. We tested our method in both indoor and outdoor low lighting environment, and achieved visually pleasing results.

6.2 Review of Contributions

The major contributions of this thesis are summarized as follows.

1. Our study on the variation of radiance gradient points out a new perspective to computational illumination. By successfully enhancing photographs in HDR and low light environment, we showed that gradient variation can be a key to enhancing photographs across different illuminations.

2. We invented a novel approach which uses NIR information to adjust the contrast of photographs adaptively and to enrich texture details automatically with one single shot.

3. We proposed a simple yet efficient computational re-flashing method to enhance low light ambient radiance by letting the user re-flash it selectively.

6.3 Future Directions

In this thesis, we demonstrated the usefulness of gradient variation across different illuminations in two applications: HDR and low light photography. In the future, this work can be extended in several ways:

Enhancement using multiple flashes In Chapter 3, we have shown how the radiance gradient varies under the change of one and two illuminants. In fact, the mathematical relationship could be extended to incorporate more illuminants with varying intensities and directions. As a result, our computational re-flashing technique could be extended by using a camera equipped with multiple flashes, similar to [Raskar et al. 2004c]. By recovering the radiance emitted by flashes in different positions, it is possible to interpret 3D scene structures and to create more flexible re-flashing effects.

Peripheral illumination correction Peripheral illumination fall-off, *a.k.a.* vignetting, is a common artifact that often jeopardizes computer vision algorithms. The incident light falls differently from optical center to peripheral region, creating the uneven exposure effect. The state-of-the-art methods usually model such nonlinear fall-off with a nth polynomial equation and estimate the parameters for each term. However, such approaches inherently assume that (1) optical center and image center coincide, and (2) vignetting is radially symmetric. As we know, under many situations, both of these assumptions are invalid. Based on our study, such illumination fall-off could be modeled with gradient variation, thus the original radiance (without vignetting) could be easily recovered by simple gradient decomposition. Such solution would not require the above assumptions, hence be more applicable and efficient.

Bibliography

Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., and Cohen, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics 23*, 294–302. 1, 2.1.2, 2.3, 5.3.7

Agrawal, A., and Raskar, R. 2007. Resolving objects at higher resolution from a single motion-blurred image. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 1–8. 1

Agrawal, A., Chellappa, R., and Raskar, R. 2005. An algebraic approach to surface reconstruction from gradient fields. In *Proceedings of IEEE International Conference on Computer Vision*. 2.1.3

Agrawal, A., Raskar, R., Nayar, S. K., and Li, Y. 2005. Removing Photography Artifacts using Gradient Projection and Flash-Exposure Sampling. *ACM Transactions on Graphics 24*, 3 (Jul), 828–835. 1, 1.2.2, 2.2.2, 2.2.2, 3.1, 3.4.2, 5.2

Agrawal, A., Raskar, R., and Chellappa, R. 2006. Edge suppression by gradient field transformation using cross-projection tensors. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. 2.2.2, 3.1

Agrawal, A., Raskar, R., and Chellappa, R. 2006. What is the range of surface reconstructions from a gradient field? In *Proceedings of European Conference on Computer Vision*. 2.1.3, 3.1, 4.4.3, 5.3.2

Agrawal, A., Xu, Y., and Raskar, R. 2009. Invertible motion blur in video. *ACM Transactions on Graphics*. 1

Bae, S., Paris, S., and Durand, F. 2006. Two-scale tone management for photographic look. *ACM Transactions on Graphics*. 2.1.1, 4.2, 4.4.3, 4.4.3

Bando, Y., Chen, B.-Y., and Nishita, T. 2008. Extracting depth and matte using a color-filtered aperture. *ACM Transactions on Graphics 27*, 5, 1–9. 1

Ben-Ezra, M., and K.Nayar, S. 2004. Motion-based motion deblurring. *IEEE Trans. on PAMI 26*, 6 (June), 689–698. 1, 5.2

Bushberg, J. T., Seibert, J. A., Leidholdt, E. M., and Boone, J. M. 2002. *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins. 5.3.3

Canon. Canon professional network. http://cpn.canon-europe.com/. 2.2.2, 5.2

Cho, S., and Lee, S. 2009. Fast motion deblurring. *ACM Transactions on Graphics (SIGGRAPH Asia 2009) 28*, 5. 1

Coe, B. 1976. *The birth of photography: the story of the formative years, 1800-1900*. Taplinger Pub. Co. 1

Debevec, P. HDR Shop. http://www.hdrshop.com/. 4.5, 5.4

Debevec, P. E., and Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH*, 369–378. 1, 1.2.1, 2.2.1, 3.1, 3.2, 1, 3.2

Durand, F., and Dorsey, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*. 2.1.1

Eisemann, E., and Durand, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics*. 1, 1.2.2, 2.1.1, 2.2.2, 2.2.2, 5.2, 5.4.1, 5.10

Fattal, R., Lischinski, D., and Werman, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*. 1, 1.2.1, 2.1.2, 2.2.1

Forsyth, D. A., and Ponce, J. 2002. *Computer Vision: A Modern Approach*. Prentice Hall. 3.3.1

Fredembach, C., and S usstrunk, S. 2008. Colouring the near infrared. In *Proceedings of the IS&T/SID 16th Color Imaging Conference*, 176–182. 4.2

Gonzalez, R. C., and Woods, R. E. 2002. *Digital Image Processing*. Prentice Hall. 4.4.3

Guo, D., and Sim, T. 2009. Face makeup by example. *Proceedings of IEEE Computer Vision and Pattern Recognition*. 2.1.2

Hays, J., and Efros, A. 2007. Scene completion using millions of photographs. *ACM Transactions on Graphics*. 2.1.2

Holst, G. 2006. Noise in imaging: The good, the bad and the right. *Photonics Spectra* (Dec). 5.3.3

Huang, J., and Mumford, D. 1999. Statistics of natural images and models. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 1541–1547. 4.2, 4.3.2, 4.3.2

Jia, J., Sun, J., Tang, C. K., and Shum, H.-Y. 2006. Drag-and-drop pasting. *ACM Transactions on Graphics*. 2.1.2, 2.3

Kazhdan, M., Bolitho, M., and Hoppe, H. 2006. Poisson surface reconstruction. In *Eurographics Symposium on Geometry Processing*, 61–70. 2.1.2

Kolb, C., Mitchell, D., and Hanrahan, P. 1995. A realistic camera model for computer graphics. In *Proceedings of ACM SIGGRAPH*, 317–324. 1

Krishnan, D., and Fergus, R. 2009. Dark flash. *ACM Transactions on Graphics*. 1

Kwatra, V., Schödl, A., Essa, I., Turk, G., and Bobick, A. 2003. Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics 22*, 3 (July), 277–286. 5.3.7

Lefkowitz, L. 1981. *The Kodak Workshop Series - Electronic Flash*. Eastman Kodak Company. 2.2.2

Levin, A., Fergus, R., Durand, F., and Freeman, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*. 1

Levin, A., Hasinoff, S., Green, P., Durand, F., and Freeman, B. 2009. 4d frequency analysis of computational cameras for depth of field extension. *ACM Transactions on Graphics*. 1

Li, Y., Sun, J., Tang, C.-K., and Shum, H.-Y. 2004. Lazy snapping. *ACM Transactions on Graphics*. 1, 2.3

Lu, Z., Wu, Z., and Brown, M. S. 2009. Directed assistance for ink-bleed reduction in old documents. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. 2.3

Lumsdaine, A., and Georgiev, T. 2009. The focused plenoptic camera. In *International Conference on Computational Photography*. 1

Maher, C., and Berman, L. Fine art infrared photography. http://www.infrareddreams.com/. 4.2

Mahmoudi, M., and Sapiro, G. 2005. Fast image and video denoising via non-local means of similar neighborhoods. *IEEE Signal Processing Letters 12*, 12, 839–842. 1

Miao, X., and Sim, T. 2005. Ambient image recovery and rendering from flash photographs. In *Proceedings of International Conference on Image Processing*, vol. 2, 1038–1041. 1, 1.2.2, 2.2.2, 2.2.2, 5.2, 5.4.1

Mitsunaga, T., and Nayar, S. K. 1999. Radiometric Self Calibration. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, vol. 1, 374–380. 2.2.1

Morris, N. J. W., Avidan, S., Matusik, W., and Pfister, H. 2007. Statistics of infrared images. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, vol. 17–22, 1–7. 4.2, 4.2, 4.3.2

Nayar, S. K., and Gong, Y. 1992. Colored Interreflections and Shape Recovery. In *DARPA Image Understanding Workshop (IUW)*, 333–343. 5.4.3

Nayar, S. K., and Mitsunaga, T. 2000. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, vol. 1, 472–479. 1

Nayar, S. K., Fang, X., and Boult, T. 1993. Removal of Specularities using Color and Polarization. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, 583–590. 5.4.3

Neumann, L., and Neumann, A. 2005. Color style transfer techniques using hue, lightness and saturation histogram matching. In *Computational Aesthetics in Graphics, Visualization and Imaging*, 111–122. 4.2

Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., and Hanrahan, P. 2005. Light field photography with a hand-held plenoptic camera. Tech. rep., Stanford University. 1

Oh, B. M., Chen, M., Dorsey, J., and Durand, F. 2001. Image-based modeling and photo editing. *ACM Transactions on Graphics*. 2.1.1

Pérez, P., Gangnet, M., and Blake, A. 2003. Poisson image editing. *ACM Transactions on Graphics*. 1, 2.1.2, 2.3

Petschnigg, G., Agrawala, M., Hoppe, H., Szeliski, R., Cohen, M., and Toyama, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics 23*, 3 (Aug), 664–672. 1, 1.2.2, 2.1.1, 2.2.2, 2.2.2, 5.2, 5.4.1, 5.10

Raskar, R., and Tumblin, J. 2009. *Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors*. A K Peters. 2

Raskar, R., han Tan, K., Feris, R., Yu, J., and Turk, M. 2004. Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics*. 1

Raskar, R., Ilie, A., and Yu, J. 2004. Image fusion for context enhancement. In *ACM Nonphotorealistic Animation and Rendering (NPAR)*. 3.1

Raskar, R., Tan, K. H., Feris, R., Yu, J. Y., and Turk, M. 2004. Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics* (Aug). 1.2.2, 2.2.2, 2.4, 6.3

Reinhard, E., and Devlin, K. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics 11*, 1, 13–24. 2.2.1

Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*. 1, 1.2.1, 2.2.1, 4.5

Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. 2006. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann. 2.2.1

Sun, J., Jia, J., Tang, C.-K., and Shum, H.-Y. 2004. Poisson matting. *ACM Transactions on Graphics*. 2.1.2, 2.3

Sun, J., Yuan, L., Jia, J., and Shum, H.-Y. 2005. Image completion with structure propagation. *ACM Transactions on Graphics 24*, 3, 861–868. 2.1.2, 5.5

Tai, Y.-W., Du, H., Brown, M. S., and Lin, S. 2008. Image/video deblurring using a hybrid camera. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. 1

Tan, P., Lin, S., Quan, L., and Shum, H.-Y. 2003. Highlight removal by illumination-constrained inpainting. In *Proceedings of IEEE International Conference on Computer Vision*. 5.4.3

Telea, A. 2004. An image inpainting technique based on the fast marching method. *J. Graphics Tools 9*, 1. 5.3.6

Tomasi, C., and Manduchi, R. 1998. Bilateral filtering for gray and color images. In *Proceedings of IEEE International Conference on Computer Vision*, 839–846. 1, 2.1.1, 4.4.3

Tumblin, J., and Rushmeier, H. E. 1999. Tone reproduction for computer generated images. *IEEE Computer Graphics and Applications 13*, 6, 42–48. 2.2.1

Tumblin, J., and Turk, G. 1999. Low curvature image simplifiers (lcis): A boundary hierarchy for detail-preserving contrast reduction. *Proceedings of ACM SIGGRAPH*. 2.1.1

Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., and Tumblin, J. 2007. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*. 1

Wang, J., Agrawala, M., and Cohen, M. 2007. Soft scissors : An interactive tool for realtime high quality matting. *ACM Transactions on Graphics*. 2.3

Wikipedia. BRDF. http://en.wikipedia.org/wiki/Bidirectional_reflectance_distribution_function. 3.3

Wu, T.-P., Tang, C.-K., Brown, M. S., and Shum, H.-Y. 2007. Shapepalettes: interactive normal transfer via sketching. *ACM Transactions on Graphics*. 2.3

Yuan, L., Sun, J., Quan, L., and Shum, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. *ACM Transactions on Graphics*. 5.2

Zhang, X., Sim, T., and Miao, X. 2008. Enhancing photographs with near infrared images. In *Proceedings of IEEE Computer Vision and Pattern Recognition*. 4.1