Low Dynamic Range Solutions to the High Dynamic Range Imaging Problem

Submitted in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

by

Shanmuganathan Raman
(Roll No. 06407008)

Supervisor:
Prof. Subhasis Chaudhuri

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY–BOMBAY
2011
Dedicated to my Mother.
Thesis Approval

The thesis entitled

Low Dynamic Range Solutions to the High Dynamic Range Imaging Problem

by

Shanmuganathan Raman
(Roll No. 06407008)

is approved for the degree of

Doctor of Philosophy

_________________________  _________________________
Examiner  Examiner

_________________________
Guide

Date: ________________

Place: ________________
Declaration

I declare that this written submission represents my ideas in my own words and where others’ ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

Shanmuganathan Raman
(Name of the student)

06407008
(Roll No.)

Date: ________________
Abstract

High Dynamic Range (HDR) imaging aims at capturing the entire brightness levels of the real world scene using a single image. This is a challenging problem which is addressed well in the analog photography which uses films for recording the scene information. This is achieved as analog cameras enjoy a much higher dynamic range to record the scene information which amounts to capture of an HDR image. A similar argument is not true in the case of digital cameras. The real world brightness differences of scenes with both brightly and poorly illuminated regions are so high that the digital sensors are incapable of recording all the brightness levels. This is due to the limited storage capacity of the digital sensors. Common digital cameras with low dynamic range (LDR) sensors lead to over- or under- saturation while capturing an HDR scene.

Capturing all the brightness levels of the real world scene using a single digital HDR image is a very challenging task. As the problem associated with HDR image generation is due to the sensor, one can design a specialized sensor with proper modifications in order to get an HDR image with a single snapshot. There are indeed HDR sensors which can capture more brightness levels than that of common digital sensors. The digital cameras which use these sensors are quite expensive at present. The other option is to develop algorithms for processing images captured using digital cameras with LDR sensors and use them, in turn, to generate the HDR image of the scene.

Traditional algorithmic approaches for HDR imaging require one to composite multiple differently exposed images of a scene in the irradiance domain. The multi-exposure images together have information of all the brightness levels of the scene which are then reproduced in the generated HDR image. We further have to perform tone mapping of the generated HDR image for displaying on LDR devices such as common displays and printers.

In this thesis, we consider algorithmic solutions for the HDR imaging problem. We first develop algorithms for directly generating an LDR image corresponding to a static scene given
a set of multi-exposure images. The primary goal of these algorithms is to preserve contrast and brightness throughout the image without getting over- or under- saturated. These algorithms are developed using mathematical tools like calculus of variations, edge preserving filtering and gradient domain processing. We show that high contrast LDR images can be generated directly from multi-exposure images using the proposed algorithms.

Real world scenes are dynamic as we do not have control over the movement of objects in the scene while capturing multi-exposure images. Hence, we consider the HDR imaging problem corresponding to a dynamic scene. In the case of dynamic scenes, the algorithms mentioned above may introduce artifacts called ghosts if the scene changes are not accounted for. We develop a novel bottom-up segmentation algorithm through superpixel grouping which would enable us to detect scene changes. We then employ a piecewise patch-based methodology utilizing algorithm for static scenes to directly generate the ghost-free LDR image of the dynamic scene.

The primary advantage of our approach presented in this thesis is that we do not assume any knowledge of camera response function and exposure settings. The generated LDR image has a high contrast and is compatible with common LDR displays and printers. The LDR image can be made compatible with even HDR displays through appropriate choice of inverse tone mapping algorithm. The techniques discussed may serve as an alternative to the “Merge to HDR” tool in Adobe® Photoshop and lead to the development of much improved techniques for processing multi-exposure images. We believe that the approaches presented in this thesis would relieve one of the burden of HDR image generation, tone mapping operation and possible human intervention.
# Contents

Abstract iii

List of Tables ix

List of Figures xi

Nomenclature xiii

1 Introduction 1

1.1 Motivation ................................................. 3
1.2 Contributions of the Thesis ................................. 4
1.3 Thesis Organization .................................... 5

2 Literature Survey 7

2.1 The High Dynamic Range Imaging Problem ................ 7
2.2 Sensor Design Solutions .................................. 9
2.3 Multi-exposure Solutions ................................ 11
2.4 HDR Imaging for Dynamic Scenes ....................... 15
2.5 HDR Image Formats ...................................... 17
2.6 Tone Reproduction ....................................... 18
   2.6.1 Spatial Domain Tone Mapping Operators ................. 18
   2.6.2 Frequency Domain Tone Mapping Operators ............. 20
   2.6.3 Gradient Domain Tone Mapping Operators ............. 21
   2.6.4 Other Operators .................................... 21
2.7 Inverse Tone Reproduction ................................ 22
2.8 HDR Imaging Softwares .................................. 23
3 A Variational Solution ........................................ 25
  3.1 Introduction ................................................. 25
  3.2 Background .................................................. 25
  3.3 Proposed Solution ........................................... 28
  3.4 Other Applications ........................................... 31
    3.4.1 Generation of Pin-hole Image ......................... 31
    3.4.2 Random Texture Generation ........................... 32
    3.4.3 Illumination Compositing for Dark Scenes ......... 35
  3.5 Discussion .................................................... 40

4 Edge Preserving Filter based Solution ......................... 43
  4.1 Introduction .................................................. 43
  4.2 Edge Preserving Filters .................................... 43
  4.3 Bilateral Filter based LDR Image Generation .......... 47
  4.4 Implementation ............................................. 49
  4.5 Discussion .................................................... 49

5 Gradient Domain Solution .................................... 51
  5.1 Introduction .................................................. 51
  5.2 Gradient Domain Processing ............................... 51
  5.3 Gradient Domain Solution ................................ 55
    5.3.1 Pre-processing of Gradients ......................... 55
    5.3.2 Gradient Domain Compositing ....................... 56
    5.3.3 LDR Image Reconstruction ......................... 57
  5.4 Discussion .................................................... 57

6 Results and Discussions ...................................... 59
  6.1 LDR Image Generation for Static Scenes ............... 59
  6.2 Conclusion .................................................... 75

7 Image Compositing for Dynamic Scenes ...................... 79
  7.1 Introduction .................................................. 79
  7.2 Related Work ................................................ 80
  7.3 Bottom-up Segmentation .................................... 81
CONTENTS

7.4 Proposed Approach .................................................. 82
  7.4.1 Estimation of the Decision Regions .................................. 83
  7.4.2 Superpixel Grouping .............................................. 87
  7.4.3 Piece-wise Rectangular Approximation ......................... 89
  7.4.4 Poisson Seam Correction ....................................... 92
7.5 Results ...................................................................... 93
7.6 Conclusions ............................................................. 96

8 Conclusions and Future Directions 103
  8.1 Conclusions .............................................................. 103
  8.2 Future Directions ..................................................... 104

Bibliography 107
List of Tables

6.1 Values of parameters corresponding to different tone mapping operators . . . 62
6.2 Values of parameters corresponding to different tone mapping operators . . . 63
### List of Figures

2.1 Multi-exposure images of a scene ........................................... 12
3.1 Generation of pin-hole equivalent image (synthetic scene) ............... 32
3.2 Generation of pin-hole equivalent image (real scene) ...................... 33
3.3 Random texture generation ...................................................... 34
3.4 Dark scene illuminated by a moving light source .......................... 35
3.5 Binarized frame of the video ................................................... 36
3.6 Light direction and composited image ....................................... 38
3.7 Composited image from a surveillance video ............................... 39
3.8 Composited image of a non-planar scene ................................... 39
3.9 Composited image of a scene with specular reflection .................... 40
4.1 Illustration of bilateral filtering operation ................................ 45
6.1 Multi-exposure images of a static scene .................................... 65
6.2 LDR results of the static scene ............................................... 66
6.3 Computed distortion maps for the LDR images ............................. 67
6.4 Multi-exposure images of a static scene .................................... 69
6.5 LDR results of the static scene ............................................... 70
6.6 Computed distortion maps for the LDR images ............................. 71
6.7 Multi-exposure images of a static scene .................................... 72
6.8 LDR results of the static scene ............................................... 73
6.9 Computed distortion maps for the LDR images ............................. 74
6.10 Multi-exposure images of a static scene ................................... 76
6.11 LDR results of the static scene ............................................... 77
Nomenclature

\[ E \]  \quad \text{image irradiance}

\[ L \]  \quad \text{scene radiance}

\[ f \]  \quad \text{camera response function}

\[ \Delta t \]  \quad \text{exposure time}

\[ d \]  \quad \text{diameter of lens aperture}

\[ h \]  \quad \text{focal length of the lens}

\[ \phi \]  \quad \text{angle between principal ray and optical axis}

\[ \eta_1 \]  \quad \text{capture noise}

\[ \eta_2 \]  \quad \text{quantization noise}

\[ I \]  \quad \text{image intensity value of a multi-exposure image}

\[ \hat{E} \]  \quad \text{high dynamic range (HDR) image}

\[ \hat{I}^{LDR} \]  \quad \text{low dynamic range (LDR) image intensity values}

\[ w \]  \quad \text{weighting function for generating HDR image}

\[ \alpha^{\text{var}} \]  \quad \text{weighting function for variational solution}

\[ \lambda^{\text{var}} \]  \quad \text{regularization parameter}

\[ \sigma^2 \]  \quad \text{local variance}

\[ \alpha^{BF} \]  \quad \text{weighting function for edge-preserving filter based solution}
$G_\rho$ spatial Gaussian function in bilateral Filter

$G_\rho$ range Gaussian function in bilateral Filter

$g$ gradient field corresponding to an image

$\tilde{g}$ modified non-conservative gradient field

$\nabla I$ gradient field corresponding to an image

$\beta$ exposure change parameter for gradient pre-processing

$\hat{g}$ composited non-conservative gradient field

$\alpha_{\text{grad}}$ weighting function for gradient domain solution

$u$ intensity mapping function

$V$ weighted variance measure of multi-exposure images

$\zeta$ Gaussian weighting function for weighted variance computation

$S$ set of all pixel locations on the image grid

$\psi$ set of all pixel locations which do not have scene change

$\gamma$ decision parameter used to identify patches having scene change

AEB Auto Exposure Bracketing

CRF Camera Response Function

HDR High Dynamic Range

IMF Intensity Mapping Function

LDR Low Dynamic Range
Chapter 1

Introduction

Nature has been quite generous in creating the world with a very large number of brightness levels. This has enabled most real world scenes to have a very high contrast. A simple example is the real world scene with both brightly and poorly illuminated regions. The real world scenes with high contrast are visually pleasing. Such scenes are said to have a high dynamic range (HDR). The human eye has also been designed with perception levels which can quickly adapt to visualize such high contrast real world scenes. This helps us visualize the objects in the scene as they are. The reflected light from the object makes a pleasing visual experience. The same cannot be said about the capture devices like the common digital cameras.

Camera controls which are typically varied while capturing a scene are shutter speed, aperture, ISO setting and focus. The shutter speed enables one to control the amount of light entering the lens and has direct correspondence with the amount of exposure received by the sensor. The exposure time is commonly used to define this control which is the inverse of the shutter speed. The aperture regulates the amount of light entering the scene apart from changing the depth of field (the amount of region in the scene being in focus). When the aperture is increased, the exposure increases while the depth of field decreases. ISO setting is usually used to change the sensitivity of the sensor and hence affects the exposure. The focus control enables the photographer to focus on objects present in the scene at different depths.

The cameras were initially used to capture scenes as the recording on analog films which were then developed. Dynamic range of a scene is roughly defined as the ratio between the brightest and darkest regions in the scene. The analog films work on the principle of burning some oxide and hence can be used to capture very high dynamic range of the scene. The major restriction of the analog cameras is that they are not real time and with the advent of digital
computers, their usage became limited. The digital cameras solve this problem and serve to capture, store and share images quickly.

One of the restrictions of the digital cameras is the inability to capture the entire dynamic range of a scene. Consider a scene where a person tries to capture the objects inside a room as well as sunlit areas through the window in a single snapshot. Such a scene has a very high dynamic range. Digital sensors present in common digital cameras cannot capture the entire dynamic range of such scenes because of the limited capacity. This leads to over- and under-saturation of the pixel locations which are brightly and poorly illuminated, respectively. As a result, what we obtain is an image with a low dynamic range.

Although sensors have been designed to capture the dynamic range of the real world scene, they are very expensive at present. The naive way to capture the entire dynamic range of the scene with the help of common digital cameras is to capture multiple differently exposed images of the scene. We assume here that all other camera controls apart from the shutter speed are fixed while capturing these multiple images. These images can then be blended to span the entire dynamic range. This has lead to the advent of a novel digital imaging stream called high dynamic range (HDR) imaging. Computational photography is a research area which blends the techniques of computer vision and computer graphics in order to enable a digital camera capture complex scenes. HDR imaging forms a major branch of computational photography.

HDR imaging aims to capture the brightness levels of the scene as it is. Common way to generate an HDR image is to blend multi-exposure images of the scene. As the scene can be better represented by the irradiance values, one needs to recover the characteristic function of a given camera called the camera response function (CRF). This function maps the intensity levels of a given image to that of the irradiance values. The irradiance values corresponding to the multi-exposure images are weighed appropriately to generate the HDR image.

The HDR image, thus generated, should be stored as floating point values. To visualize the HDR image using common displays and printers, one needs to use an appropriate tone mapping operator. Tone mapping is an operation meant to map the HDR image into an equivalent low dynamic range (LDR) image without much loss of color and contrast. The LDR image, thus generated, would then be compatible with common display devices. There are a variety of tone mapping operators which operate on spatial, frequency or gradient domains.

The multi-exposure images can be shot with a hand-held camera which would need registration. The scene can either be static or be dynamic. In the case of dynamic scenes, the
1.1 Motivation

Consider a set of multi-exposure images of a static scene. We want to solve the problem of representing all the contrast levels of the real world scene using a single image. We, further, want to make the image compatible with common displays and printers. If we further assume that there is no knowledge of the CRF, we perform all the computations on the intensity values of the multi-exposure images. Such an image can never be an HDR image as we do not operate in the irradiance space. This is the solution we really want. If we could somehow design all the operations on the intensity values of the images, we can then generate an LDR image which would convey all the contrast information in the available dynamic range $[0,255]$.

However, we do not have control over the changes in the real world scene we intend to capture using multi-exposure images. There may be new objects coming in and some objects moving out. Further, there may be objects which might have changed their positions. In other words, we can label most real world scenes as being dynamic. We would want to solve the problem of generating an LDR image corresponding to a dynamic scene. This image should not have any artifacts due to the objects in motion and should have high contrast as well. This is a much tougher problem compared to that of the static scene.

We want to solve the HDR imaging problem for both static and dynamic scenes and accommodate the entire real world brightness levels in the available dynamic range, say $[0, 255]$. We want to achieve this task blindly as we assume that we are given just the multi-exposure images and no other information. We desire to avoid artifacts caused due to the moving objects in this process. This is a challenging task as one needs to maintain the high-contrast and properly exposed regions in the LDR image to capture the high dynamic range of the real world scene. Let us now look at the primary contributions of this thesis.
1.2 Contributions of the Thesis

We address the problem of generating an LDR image of the scene directly from a set of multi-exposure images. We assume that we do not have the knowledge of CRF and exposure settings corresponding to multi-exposure images in our work in the thesis. The contributions of the thesis are as below.

1. We attempt to solve the HDR imaging problem for static scenes in the LDR domain itself. We formulate the ill-posed problem in a calculus of variations framework. The unknown LDR image image can be estimated using a smoothness constraint and the iterative solution converges to the desired LDR image. This approach models the weighting function as a data dependent term to weigh the multi-exposure images and there is no explicit calculation of the weighting function. We arrive at an iterative solution using the Euler-Lagrange equation which converges to the high contrast LDR image. This approach is computationally expensive compared to the other two approaches discussed below.

2. Edge-preserving filters are used to define a weighting function to weigh differently exposed images of a static scene in the image domain. The key idea here is to weigh more the pixel locations which have weak edges or textures calculated using a non-linear filter called bilateral filter. It is shown that the approach is faster than the variational solution and LDR images have better contrast. The fast bilateral filters which have evolved recently enable us to perform this weighting operation much faster. This simple technique enables us to generate high contrast LDR image corresponding to a static scene.

3. The task of generating LDR image directly from a set of multiple differently exposed images can also be achieved in the gradient domain. A better solution for direct generation of an LDR image from multi-exposure images of a static scene is proposed in the gradient domain. The gradient domain compositing allows us to perform seamless reconstruction of the desired LDR image. We perform the pre-processing of the gradients followed by the weighting operation in the gradient domain. We employ a Poisson solver to generate the desired LDR image. Apart from generating a high contrast LDR image, even homogeneous regions are well captured using this approach.

4. The proposed approaches for the generation of LDR image corresponding to a static scene lead to artifacts called ghosts when applied directly on the multi-exposure images corre-
sponding to a dynamic scene. We propose a bottom-up segmentation algorithm based on super-pixel grouping for segmenting out scene changes for dynamic scenes. This avoids ghosting artifacts from appearing in the final LDR image while using any of the first three static scene approaches. We show that high contrast LDR image without any artifacts corresponding to a dynamic scene can be generated using the proposed approach. The proposed approach does not require the knowledge of CRF either for motion detection or for LDR image generation and does not need any user interaction as well.

5. The general compositing approach proposed for static scene is shown to solve other relevant applications in computational photography such as all-in-focus image generation, random texture generation, and illumination compositing for dark scenes. We illustrate these applications using the variational approach for static scenes.

1.3 Thesis Organization

The second chapter discusses the background literature on HDR imaging and related techniques. This chapter sets the notation and background terminologies which would be used in later chapters. The next three chapters deal with the generation of LDR image by weighting multi-exposure images corresponding a static scene.

The chapter 3 explains the variational approach for generating an LDR image from multi-exposure images of a static scene. We discuss how an implicit weighting function can be used to weigh the multi-exposure images and an iterative solution can be arrived at. The chapter 4 explains the idea behind edge-preserving filters and how they can be used to composite multi-exposure images of a static scene. We show that the small texture details obtained using edge-preserving filter provide us information for the appropriate design of the weighting function. The chapter 5 provides a gradient domain solution for the static scene. We show that a high contrast LDR image can be generated by gradient domain processing and Poisson solver.

The results from the three approaches for static scenes explained in the last three chapters are discussed in chapter 6. We employ a dynamic range independent quality metric to validate the results. Chapter 7 considers a dynamic scene and explains how multi-exposure images can be combined without any ghosting artifacts. The results for the dynamic scenes are also discussed in this chapter. We conclude the thesis in chapter 8 by providing some future directions for the techniques discussed in the thesis.
Chapter 2

Literature Survey

In this chapter, we shall present an overview of the existing literature. We shall consider the literature related to the HDR imaging problem and review some of the approaches which address the various aspects of the HDR imaging problem. This chapter explains the various terminologies involved apart from providing the necessary background information. We shall first provide an overview of the HDR imaging problem. We shall discuss the existing solutions for the static scene HDR imaging problem using imaging hardware modifications and algorithms. We also explore the existing solutions in the case of dynamic real world scenes and also provide a glimpse of the various file formats used to represent the HDR images. The various tone mapping operators for making HDR images to be compatible with the LDR displays is then provided. We shall provide a brief discussion of the inverse tone mapping operators and their significance in visualizing LDR images using HDR displays. We conclude the chapter with an overview of the existing HDR imaging softwares.

2.1 The High Dynamic Range Imaging Problem

In this section, we shall present a general overview of the HDR imaging problem and discuss some of the common methods proposed to solve this problem. We shall first consider the image formation in a digital camera while capturing a real world scene with high dynamic range. The image irradiance $E$ is related to the scene radiance by the Equation 2.1 [1].

$$E = L \left( \frac{\pi}{4} \right) \left( \frac{d}{h} \right)^2 \cos^4 \phi$$  \hspace{1cm} (2.1)

where $d$ is the diameter of the lens aperture, $h$ is the focal length of the lens, $\phi$ is the angle
between principal ray and the optical axis, and $L$ is the scene radiance. The camera settings which enables one to change the amount of light entering into the camera lens are the shutter speed (exposure setting), ISO setting and the aperture. If we assume the focus, ISO setting and the aperture are fixed, the 2-D image formation can be described by the Equation 2.2.

$$I(x, y) = f(\Delta t E(x, y) + \eta_1) + \eta_2$$  \hspace{1cm} (2.2)

where $I(x, y)$ is the intensity value of the image, $\Delta t$ is the exposure time, and $E(x, y)$ is the image irradiance. The image irradiance $E(x, y)$ and the image intensity $I(x, y)$ are related by a non-linear function called camera response function (CRF) $f$. The noise model $\eta_1$ denotes the capture noise mainly due to shot and thermal noise and the noise model $\eta_2$ denotes the quantization noise mainly caused due to quantization and amplifier ([2], [3]). Generally, CRF is plotted between the logarithm of exposure and the image intensity [4].

The CRF $f$ is a non-linear function introduced by the digital signal processor (DSP) present in the digital cameras. The RAW image obtained using a digital camera image is subjected to various processing modules such as demosaicing, sharpening, white balancing, gamma curve application, and compression. These modules finally lead to the generation of the LDR image which is typically encoded in JPEG format. These modules which are part of the DSP present in the digital camera together account for the non-linearity associated with its CRF [5]. The CRF $f$ corresponding to a digital camera is a proprietary function designed by the camera manufacturer and is not available to the consumers using the digital camera.

A simplified model for the image formation can be arrived at by not considering the noise sources. This is as shown in Equation 2.3.

$$I(x, y) = f(\Delta t E(x, y)).$$  \hspace{1cm} (2.3)

Given a set of multi-exposure images corresponding to a static scene, we can rewrite Equation 2.3 as shown in Equation 2.4.

$$I_k(x, y) = f(\Delta t_k E(x, y))$$  \hspace{1cm} (2.4)

where $I_k(x, y)$ represent the intensity values of the $k^{th}$ image in the exposure stack with exposure time $\Delta t_k$.

The image we capture using the common digital cameras follows Equation 2.4. The CRF is a characteristic function for a given camera and it is not explicitly provided by the camera.
manufacturers. Given a set of differently exposed images and the corresponding exposure times, one can estimate the CRF $f$ [6]. The primary use of the CRF is to map the image intensity values with that of the corresponding irradiance values. The non-linearity associated with the CRF accounts for the limitation of the sensor capacity as well as the compression performed on the captured voltages.

An HDR image is the representation of image irradiance space as it represents the actual brightness levels of the real world scenes. Intelligent design of the camera internals like the lens and the sensor enables one to capture the entire dynamic range using a single snapshot. It is obvious from the discussion above that the sensor must be capable of capturing the entire dynamic range of the scene if one intends to capture a HDR scene using a single snapshot. The advancement in the design of digital CMOS/CCD sensors enables us to capture the entire dynamic range of the scene using a single image [7]. We call these set of solutions as sensor design solutions to the HDR imaging problem.

Alternately, consider a set of multi-exposure images of a scene. These images together can capture the entire dynamic range of the scene. These images can be used to generate the HDR image of the scene. There are approaches which employ such an idea which do not require one to redesign the sensor to enhance its dynamic range. These approaches can be used to generate the HDR image using common digital cameras with LDR sensors once the CRF is known. We call these class of solutions as multi-exposure or algorithmic solutions to the HDR imaging problem.

In the next two sections, we shall discuss a number of approaches which address the HDR imaging problem through sensor design and multi-exposure capture, respectively.

## 2.2 Sensor Design Solutions

The dynamic range of an imaging system is defined as the ratio of the maximum unsaturated voltage level measured by the sensor element to the voltage measured when there is no exposure (dark noise). The sensor well capacity, which is the amount of voltage a sensor element can hold without saturation, plays a major role in determining the dynamic range of an imaging system. The dynamic range can be enhanced by either changing the well capacity or by using multiple sampling methods [7].

Sensor well capacity adjustment involves either increasing the maximum or decreasing the
minimum voltage which can be accommodated in the sensor element. Though this leads to the increase of the sensor dynamic range, it also leads to the decrease of the signal to noise ratio (SNR). While changing the well capacity affects the SNR considerably, there is a necessity for an alternate technique which allows one to increase the dynamic range without affecting the SNR. This SNR - dynamic range trade off can be taken care by using multiple sampling technique. In this technique, the sensor elements which are subjected to large illumination are integrated over a lesser time compared to those sensor elements which are subjected to low illumination. A comparative study of the CMOS sensors with respect to their dynamic range can be found in [8]. An overview of the various attempts to increase the dynamic range of the sensor by different techniques is presented in the tutorial by El Gamal [9].

The sensor with limited capacity can therefore be modified using an intelligent design in order to capture the entire dynamic range of the scene. One can employ spatially varying pixel exposures by employing a mask before the sensor with varying transmittances. The spatial capture of the brightness levels can then be used to reconstruct the HDR image of the scene [10]. Another design inspired from the multiple sampling techniques discussed earlier is the assorted pixel-based approach where the sensor elements with different capacities are placed adjacent to each other. After capturing the scene, sampling the sensor elements which do not saturate among these sensor elements enables one to expand the dynamic range ([11], [12]). These techniques, however, lead to the reduction in the spatial resolution of the captured image as many sensor elements contribute to a single pixel. In other words, there is lot of redundancy introduced while capturing the scene.

Another approach is to have a split aperture for the lens which can capture the entire dynamic range of the scene using a single snapshot [13]. The idea here is to split the beam of light entering the lens into different directions using mirrors which can then be focused on sensors with different sensitivities. This corresponds to the capture of multi-exposure images using a single lens and hence HDR image of the scene can be reconstructed from a single click. A similar idea has been extended to a plenoptic camera. A plenoptic camera captures the 4D light field of a scene. It can be used to generate an HDR image of the scene when successive lenslets are designed with different aperture diameters. This would enable us to capture an HDR image of the scene from a single snapshot apart from benefits of capturing light field of the scene [14]. Commercially available HDR cameras like Spherocam capture HDR image using a single snapshot [15]. These cameras are very expensive at present and are not affordable
by photographers.

We would like to capture the entire dynamic range of the scene with the common digital cameras without the need to modify the camera internals. The overhead in using common digital cameras is that we are forced to capture multiple images of the scene by varying the exposure times as we do not modify the camera internals any more. This relaxes the need of huge expenses for designing complicated camera model. These multi-exposure images can then span the entire dynamic range of the scene. We shall see how this can be achieved in the next section.

### 2.3 Multi-exposure Solutions

Consider a set of multiple differently exposed images of a static HDR scene shown in Figure 2.1. A typical HDR scene would have both brightly and poorly illuminated regions contributing to the very high dynamic range. These images can be obtained from the irradiance of the scene using the Equation 2.4. The smaller values of $\Delta t_k$ let us capture information from the brightly illuminated regions of the scene. The larger values of $\Delta t_k$ let us capture information from the poorly illuminated regions. In this manner, these images would together capture the entire dynamic range of the scene irrespective of the limitations of the camera internals. As the entire dynamic range information is present in this set of multi-exposure images, one needs to appropriately weigh these multi-exposure images in order to generate the HDR image of the scene.

The Wyckoff’s principle which is commonly used by photographers gives the basis for combining multiple images of a scene captured with different exposure times. This principle states that one can capture different information from a scene using differently exposed images which can then be combined to get all the information of the scene irrespective of the illumination [6]. This fact was exploited by Mann and Picard who extended it to the digital images. They were the first to develop a method for combining multi-exposure images of a scene to generate an HDR image. The primary motivation of the work was to let digital cameras also perform as good as analog cameras while capturing HDR scenes. Given a set of multi-exposure images of a static scene, the first step is to recover the corresponding values of the images in the irradiance space. This can be achieved when one has the knowledge of the CRF [6].

It is interesting to note that irradiance maps corresponding to all images together spans
Figure 2.1: (a-i) Multi-exposure images of a scene. Images Courtesy: Erik Reinhard, University of Bristol.
2.3 Multi-exposure Solutions

the entire dynamic range of the real world scene. Some literature refers to the irradiance maps as the radiance maps. We use the term irradiance map to denote the linearized intensity values obtained using a CRF. The HDR image is the weighted linear combination of these irradiance maps. The first step in the generation of HDR image is the estimation of the CRF which would then reveal the irradiance maps. The CRF is a monotonic, invertible function [16].

The HDR image can be generated by weighting these irradiance maps appropriately. Given a set of irradiance maps corresponding to multi-exposure images, the HDR image \( \hat{E}(x, y) \) is given by Equation 2.5.

\[
\hat{E}(x, y) = \sum_{k=1}^{K} w_k(x, y) \frac{f^{-1}(I_k(x, y))}{\Delta t_k}
\]

where

\[
\sum_{k=1}^{K} w_k(x, y) = 1
\]

The weighting function \( w_k(x, y) \) can be designed in various ways for the purpose of generating the HDR image. Mann and Picard suggest the use of the derivative of the CRF, known as certainty function, and use that as weighting function for generating the HDR image [6]. The certainty function resembles that of a Gaussian function which weighs more the irradiance values which are properly exposed. The over- and under- exposed irradiance values are given lesser weights. This enables one to generate an HDR image which is the replica of the real world scene without over- or under- saturation.

Debevec and Malik provide a practical way to estimate the CRF and also recover the HDR image from a set of multi-exposure images [16]. Assuming that the exposure times corresponding to the multi-exposure images are known, a least squares estimation is employed. They suggest the use of a simple hat function instead of the certainty function to weigh the irradiance values corresponding to the multi-exposure images. This function also amounts to weighing the properly exposed regions more compared to the over- or under- exposed regions. This approach is quite popular and is considered the standard method for the generation of the HDR image.

Mitsunaga and Nayar model the inverse of the CRF as a polynomial and estimate the parameters of the polynomial [17]. This enables one to directly invert the intensity values once the polynomial is estimated. This would then recover the irradiance values corresponding to the multi-exposure images when the ratios of exposure are provided. The weighting function suggested by this method is the ratio of the inverse of CRF and its derivative. This weighting function was modeled in order to increase the signal to noise ratio (SNR). The approaches
discussed till now employ the simplified imaging as shown in Equation 2.3. The more advanced approaches for the generation of HDR images use the imaging equation in Equation 2.2. We shall now discuss a few of these approaches.

Considering the uncertainty associated with different types of noise sources in the imaging process shown in Equation 2.2, one can estimate the CRF and also use it to detect scene changes [2]. A more simplified approach involves modeling the additive white Gaussian noise (AWGN) during the capture process to represent $\eta_1$ and $\eta_2$ in Equation 2.2. One can recover the CRF and the HDR image corresponding to a set of multi-exposure images using statistical estimation. This technique which models both the capture noise and the quantization noise as a Gaussian process and employs a maximum likelihood estimation to generate the HDR image and the certainty function can then be used as the weighting function [18]. A typical digital camera suffers from both the temporal and spatial noise. Both these types of noise are taken into account while estimating the optimal weighting function using a stochastic model and the maximum likelihood estimates [19]. This weighting function can then be applied to the linearized intensity values of the multi-exposure images once the CRF is known.

The CRF can be estimated from a set of multi-exposure images by first estimating a function called intensity mapping function (IMF). IMF is assumed to be a polynomial function relating the intensity values of two images captured with different exposure times. This function can be used to even recover the exposure ratios between the images apart from the CRF of the imaging system even when the images are not registered [20]. IMF is also known as comparametric function and its properties are discussed by Mann [21]. The constraints, a given function has to satisfy in order it to be a CRF, are discussed in [22]. An empirical model of CRFs is built using a database of possible real world CRFs and this model can then be used to estimate CRF of a given imaging system from a set of multi-exposure images. The CRF can also be recovered approximately from a single image provided sufficient number of edges are present in the image over the entire range of brightness values in each color channel [23].

Texture details can be transferred from a properly exposed region to over- or under-exposed regions with some user interaction to create a hallucinated image which preserves contrast [24]. The generated image using such a method gives the resemblance of an HDR equivalent image of the scene. HDR imaging techniques discussed for still images can be extended to the generation of an HDR video of the scene. This can be achieved by capturing alternate frames of a video with different exposure times. Appropriate registration techniques can be used between
adjacent frames and HDR frames can be generated [25]. A disadvantage with such a method is that the frame rate of the final HDR video is a fraction lesser compared to that of the captured LDR video. Further one has to employ an appropriate compression scheme for the HDR video as it occupies huge memory.

Higher ISO settings would enable one to reduce the amount of exposure time required to capture the multi-exposure images. However it leads to more sensor noise due to the increase in sensor sensitivity. This problem was addressed in order to generate an HDR image with reduced noise [26]. The trade-off between the ISO setting and the exposure time plays a major role in the capture of an appropriate number of multi-exposure images for spanning the entire dynamic range of the scene. Noise reduction can be achieved by optimally capturing the set of exposures at appropriate exposure times obtained by posing it as an mixed integer programming problem [27].

It is not always possible to mount the digital camera in a tripod while capturing multi-exposure images. While using hand-held cameras, there would be slight shift in the multi-exposure images. This requires one to register the images before combining to generate a HDR image. The most common approach is to convert all the RGB multi-exposure images to grayscale and then align the bitmap versions of these grayscale images [28]. This alignment prevents artifacts arising in the generated HDR image due to hand shake. The longer exposures can lead to motion blur along with the blur due to camera shake. Both can be removed by properly identifying the blur kernels and incorporating these information during the HDR image reconstruction [29].

2.4 HDR Imaging for Dynamic Scenes

While capturing multi-exposure images of a scene, the photographer has least control over the real world scene. Most of the real world scenes we encounter are in fact dynamic. During the time of capture, many objects already present in the scene may change position and new objects may be introduced. When these changes in the scene are not detected and accounted for before combining these images, the final HDR image will have artifacts called ghosts. These artifacts lead to inaccurate visual representation of the real world scene and hence need to be eliminated. The process of eliminating ghosts while generating HDR image corresponding to a dynamic scene is known as deghosting.
Deghosting was first introduced in the panorama photography. Uyttendaele et al. developed a method to generate artifact-free panorama from image sequences of a dynamic scene [30]. The common technique to handle multi-exposure images of a dynamic scene is to detect the pixel locations which are known to have changed in any of the images. This can be achieved by either calculating a weighted variance or the entropy across the images corresponding to a pixel location ([4], [31]). Such pixel locations are then replaced by values from one of the images which has high contrast and does not have any appreciable change. The pixel locations which are static in all the images are composed by traditional HDR techniques discussed above. This approach tends to diminish contrast in the pixel locations which show scene change in any of the multi-exposure images and selecting the appropriate values from the set of images may require manual intervention.

The change detection across the multiple differently exposed images can be merged with the HDR image generation process and appropriately lesser weights can be given to the pixel locations of an image if it is found to have scene change [32]. This technique helps one to eliminate ghosts to certain extent. This approach suffers from the fact that even pixel locations which have scene change are given some weight which may lead to artifacts which are evident while looking closely. This approach also assumes the knowledge of CRF to generate the HDR image.

The most effective approach till date for generating HDR image from multi-exposure images is the one by Gallo et al. [33]. The knowledge of CRF is assumed and one of the multi-exposure images is taken as the reference image in this approach. When there is no change in the scene, the irradiance $E$ remains the same for all the differently exposed images. One can show that the logarithm of linearized intensity values of an image should follow the equation of a straight line with respect to the logarithm of linearized intensity values of a differently exposed image of the same scene in case there is no change in the scene. By permitting some trade-off from the linear function, one can detect changes in the scene in each patch. This enables one to eliminate the ghosts on the generated HDR image.

Consider the case where low ISO settings are used to capture the multi-exposure images. This would require a longer time to capture the highest exposure. If there are moving objects in the scene, the highly exposed image will have motion blur which can also lead to ghosting artifacts. De-blurring needs to be performed to avoid ghosts. In case we use high ISO settings, the exposure times required to capture the multi-exposure images would be lesser than the
previous case. However one has to handle sensor noise in the less exposed images. Proper denoising algorithm needs to be used to remove this noise to recover sharp HDR image. It is observed that the second case stated above yields a more effective solution than that of the first case [34].

2.5 HDR Image Formats

A typical HDR real world scene has a dynamic range upto 10 orders of magnitude. The generated HDR image has floating point values corresponding to each pixel location as it spans the irradiance space. The HDR image needs to be encoded in a standard image format. There are a number standard formats with which an HDR image can be encoded, each of them having a different bit depth and supported maximum dynamic range. An overview of the popular HDR image formats can be seen in [35].

The Pixar log (TIFF) format uses 11 bits per channel which means 33 bits per pixel. The dynamic range supported is 3.8 orders of magnitude which is quite negligible when compared with the dynamic range of real world scenes [36]. Another file format for HDR images is the OpenEXR format by Industrial Light and Magic [37]. This format assigns 16 bits for each channel with one sign bit and 5 bits for exponent. This means there are 48 bits assigned for each pixel. The dynamic range of this format is 10.7 orders of magnitude [38]. The file extension for this format is .exr.

The LogLUV encoding also leads to another TIFF format (.tiff extension). There are two versions of this format - 24 bits per pixel with dynamic range equal to 4.8 orders of magnitude and 32 bits per pixel with dynamic range equal to 38 orders of magnitude [39]. The most popular HDR image format is the Radiance RGBE format which assigns 8 bits per channel with 8 bits assigned for a shared exponent. This format assigns a total of 32 bits per pixel [40]. The maximum dynamic range supported is 76 orders of magnitude and the file extension for this format is .hdr. Radiance RGBE format is the most commonly used file format for representing HDR images.
2.6 Tone Reproduction

The HDR image generated and encoded using any of the formats mentioned in the previous section requires specialized displays. Projector-based displays and LED-based displays with extended dynamic range can be used for the visualization of HDR images [41]. However these displays are very expensive and are still under development. For the common displays which can understand only the LDR content, one has to map the generated HDR image into an appropriate LDR image. This process is known as tone reproduction or tone mapping. The key idea behind tone mapping is to render the scene without any perceptual difference to the human viewer.

There are a variety of tone mapping operators present in the literature. They can be broadly classified into spatial domain operators, frequency domain operators and gradient domain operators. The spatial domain operators can in turn be classified into global and local operators. The proper choice of an appropriate tone mapping operator is a challenge. We shall provide an overview of some of the popular tone mapping operators in each of these categories. A comprehensive account of the tone mapping operators mentioned here can be found in [4].

2.6.1 Spatial Domain Tone Mapping Operators

Global tone mapping operators use a global non-linear function to map the HDR image into an LDR image. The primary advantage of the global operators is that they are very fast and are suitable for real time implementation. However, they suffer from the disadvantage of washing out details of certain parts of the scene as the same global function is applied for all the intensity values. Most of the global operators aim to match either the brightness or the contrast or both of the HDR image and the LDR image.

The first global operator is the brightness ratio preserving operator. It considers that the ratio between a pair of brightness values of the HDR image and the ratio between the corresponding pair of brightness values of the LDR image should be equal. This operator models the brightness as a power function of luminance. This approach provides a global function relating the luminance of the scene and the intensity (brightness) values of the LDR image [42]. This operator is inspired from lighting design and hence most suitable only for indoor scenes.

Instead of preserving ratios of brightness values, the brightness preserving operator tries to preserve the brightness values themselves before and after tone mapping [43]. This operator
produces better results compared to the brightness ratio preserving operator. The brightness values of the LDR image can be modeled as a scaled version of the real world luminance. The scale factor can be estimated in such a way as to preserve the contrast instead of the brightness [44]. Ferwerda et al. used a bias term based on the luminance to the the term specified in [44]. This model helps one to model the scotopic vision apart from the photopic lighting conditions as the luminance term added is an achromatic component.

Simple logarithm and exponential mappings similar to gamma correction can also be employed to map the HDR image on to the LDR image. This approach works well when the dynamic range of the real world scene just exceeds the dynamic range of the common display devices. The results are not that good when logarithmic or exponential functions are applied to the high dynamic range scenes which have both brightly and poorly lit regions.

The logarithmic operator can be extended to have a variable base between 2 and 10 and thereby can be adjusted to compress high dynamic range scenes[45]. A bias parameter is introduced to modify the contrast of the tone mapped LDR image in a desirable manner. Inspired by the photoreceptor physiology of the human eye, Reinhard and Devlin designed a tone mapping operator based on photoreceptor response of the human eye [46]. This operator has to be applied on the R, G, B channels separately similar to what happens inside the human eye. Another operator which exploits the human visual adaptation mechanism for building a tone mapping operator is the time dependent visual adaptation operator [47]. This operator uses a similar model to the one discussed earlier by Tumblin and Rushmeier [43].

Histogram adjustment can also be used to match the features of the LDR and HDR image [48]. Such an operator apart from preserving visibility, can be used to preserve the contrast, color and brightness of the LDR image with respect to the HDR image. Uniform rational quantization can be used to tone map the HDR image with only two parameters which can be estimated accurately through experiments [49].

Global operators are simple and fast, however there is a limit to the maximum dynamic range of the HDR image handled by them. This leads us to design tone mapping operators taking local information also into account. Local operators employ a measure of neighborhood around a pixel location and hence the operation required varies from one pixel location to another pixel location. This helps one to tone map HDR images of higher dynamic range.

The basic local operator is the spatially varying operator one which divides the HDR image by the blurred version of the image at each pixel location [50]. This effectively amounts
to introducing halos at some pixel locations which is due to the blurring process involved in the operation. The same operation can be performed in the logarithm of the HDR image using the retinex theory to reduce halo artifacts [51]. This approach works on different color channels separately. Further multi-scale operation can be performed with Gaussian functions of different kernel sizes.

Alternately, color appearance model can also be employed to tone map medium dynamic range images [52]. Another local operator which employs color appearance model is the multi-scale observer model based on color appearance correlates [53]. This model takes into account the complete human visual processing steps and tries to use them for compressing the dynamic range. This model is complex and not suited for real time implementation. It can be used to handle HDR images of scenes with very high dynamic range. Another local operator which employs simplified model of the human visual system is the one by Ashikhmin [54].

The photographic tone mapping operation functions as a global operator for the medium dynamic range HDR image and as a local operator for those images with very high dynamic range [55]. This operator aims at reducing artifacts such as visible halos. This operator uses multiple Gaussians with increasing kernel sizes and is similar to the concept of edge preserving filter which will be examined in detail in chapter 4. There is a local tone mapping operator which segments the HDR image into various regions using the histogram [56]. After grouping similar regions, grouped regions are handled separately.

### 2.6.2 Frequency Domain Tone Mapping Operators

For diffuse surfaces, the HDR image of the scene can be split into reflectance and illuminance components. A homomorphic filtering operation can then be employed. The illuminance component leads to the high dynamic range and hence needs to be compressed. The low frequencies corresponding to the illuminance component are attenuated in the frequency domain to achieve the desired tone reproduction [57]. This approach is the earliest attempt known to reduce the dynamic range of an image.

Bilateral filter is an edge preserving filter which splits an HDR image into a base layer and a detail layer. The detail layer is inherently LDR while the base layer is HDR. The base layer needs to be compressed using an appropriate function and combined with the detail layer in order to reduce the overall dynamic range. Such an approach is employed for tone mapping HDR image into an LDR image [58]. The edge-preserving filter for three dimensions is known
as the trilateral filter. This filter provides a piece-wise linear approximation of the image while
the bilateral filter provides a piece-wise constant approximation of the image. The trilateral filter
uses gradient information to compute the base layer effectively [59]. This filter can be employed
to achieve better tone mapping compared to the bilateral filter. Edge avoiding wavelets can
also be employed to split the HDR image into base and detail layers and thereby perform tone
mapping [60].

2.6.3 Gradient Domain Tone Mapping Operators

The first work to explore the possibility of tone mapping in the gradient domain is the one by
Horn [61]. As already discussed in the previous section, it is assumed that the HDR image can
be split into its reflectance and illuminance components. The gradient of the logarithm of the
HDR image is then computed which is similar to the computation of contrast ratios of the image
itself. The gradients can then be subjected to simple thresholding making the higher gradients
zero. The resulting Poisson equation can be solved to reconstruct the desired LDR image. A
more detailed account of the gradient domain processing is provided in chapter 5.

This approach was further extended later by designing an appropriate weighting func-
tion which attenuates the larger gradient smoothly [62]. The modified gradient field in the log
domain is then employed in the Poisson equation to recover the underlying scalar field. The ex-
ponentiation of the scalar field leads to the generation of the desired LDR image. This approach
is popularly known as the gradient domain HDR compression and is widely used. Another
tone mapping operator operates on the visual response space [63]. The idea here is to arrive
at a global tone mapping operator based on a transducer function which maps the input image
onto the contrast information. The inverse of this transducer function can then be used to map
the contrast information present in the HDR image onto the corresponding LDR image. This
approach is similar to gradient domain tone mapping though it employs a pyramidal contrast
representation which relieves one from solving the Poisson equation.

2.6.4 Other Operators

An interactive tone mapping operator which adjusts local tonal values with the help of few user
brush strokes is proposed by Lischinski et al. [64]. Tone mapping operator can be generated for
a given display device by minimizing the amount of distortion [65]. This can be achieved by
using the information from a human visual system (HVS) perception model. This approach can be employed to retain details in the region of interests like human faces. A generic tone mapping operator approximates various local and global spatial operators by simple image processing operations like a curve and a spatial modulation functions [66]. This generic operator can then be used to analyze the performance of different tone mapping operators. Color correction techniques for the already existing local and global operators is discussed in [67].

2.7 Inverse Tone Reproduction

We may occasionally need a methodology to display the already existing LDR images on the HDR displays. This process of converting the LDR images so that they can be visualized in the HDR displays is known as inverse tone mapping [68]. This approach uses the inverse of the photographic tone mapping operator discussed earlier to expand an LDR image into an HDR image. A similar approach can be applied to an LDR video to convert it into an HDR video [69].

A more practical real time algorithm for performing inverse tone mapping involves inverse gamma correction and brightness enhancement [70]. This method can be implemented in a GPU or HDR display hardware and can be used for images as well as videos. It is found that these inverse tone mapping operators do not perform that well if the LDR images are not exposed properly. Over- and under- exposures degrade the performance of the inverse tone mapping operators. In such situations, it is convenient to use simple methods such as gamma expansion as substitute for more complex inverse tone mapping techniques [71]. The complexity of the different inverse tone mapping operators are analyzed using psychophysical studies and their complexities are investigated in [72].

A more detailed account of the inverse tone mapping operators can be found in the state of report by Banterle et al. [73]. However, through psychophysical evaluation it is shown that just boosting the LDR image to an HDR image provides a better visual experience compared with the application of an inverse tone mapping operator [74]. These methods provide a means to display LDR images into HDR displays. This means that the generation of high quality LDR images becomes as important as the generation of HDR images.
2.8 HDR Imaging Softwares

There are many commercial softwares which have incorporated the techniques reviewed in the previous sections. The Adobe® Photoshop CS4 has a HDR image generation tool called “Merge to HDR” which enables one to create an HDR image from a set of multi-exposure images [75]. It recovers CRF from the exposure times recorded in the header of the images and hence HDR image of the static scene is reconstructed. The generated HDR images can be saved as Radiance RGBE (.hdr) or OpenEXR (.exr) formats. The latest release from Adobe® Photoshop called CS5 also has the ability to deal with multi-exposure images corresponding to dynamic scenes albeit with some user interaction. The tone mapping utility of “Merge to HDR Pro” tool in Photoshop CS5 enables one to save the HDR image as an LDR image as tone mapping is integrated with HDR image generation. However, this tool requires one to specify the exposure times corresponding to the multi-exposure images.

The recent releases of Matlab® numerical computational software by Mathworks Inc. enables one to read, write, create and tone map an HDR image. These tasks are achieved using the following functions present in the image processing toolbox - hdrread (read an HDR image), hdrwrite (write an HDR image), makehdr (create an HDR image from a set of multi-exposure images) and tonemap (create an LDR image from the given HDR image). The multi-exposure images corresponding to a dynamic scene cannot be handled using the Matlab® functions mentioned above [76].

Photomatix® is another novel commercial software by HDRsoft exclusively designed for the generation of HDR image from a set of multi-exposure images. Many photographers use this tool for obtaining the photographic look of the HDR image. This software also provides functions such as tone mapping, direct LDR image generation and deghosting for dynamic scenes [77]. This software can either be used as a stand alone version or be used as a plugin for other softwares such as Adobe® Photoshop, Lightroom and Aperture.

There is also an open source software called pfs tools available through sourceforge for performing basic tasks of HDR imaging such as HDR image generation, tone mapping and deghosting [78]. The tone mapping is achieved through an utility called pfs tools which incorporates many widely used tone mapping operators. We shall be using this tool to generate the results of different tone mapping operators in Chapter 6. The HDR image generation is achieved using an utility called pfscalibration which implements algorithms by Robertson et al. [18], and
Mitsunaga and Nayar [17]. This software can be freely downloaded along with the source code from the sourceforge website [79].

HDR videos are very memory occupying and therefore require some technique for compression. GoHDR is a new company dedicated to the development of HDR video compression algorithms [80]. They claim to achieve 100:1 compression ratio for HDR videos. The compression algorithms are available for HDR camera models by Spheron-VR [15].

This chapter explained the background theory behind HDR imaging. We focus on solving the HDR imaging problem using post-processing algorithm rather than modifying the camera internals. We would now want to generate the LDR image directly from a set of multi-exposure images of the scene. We shall first discuss three different methods for achieving this task while capturing static scenes. We shall present some other related applications which can be addressed using the algorithms for static scenes. We shall then extend it to address more complex dynamic scenes where there is appreciable change in the scene while capturing multiple images of the scenes.

The primary objective is to generate the desired high contrast LDR image assuming that we do not have any knowledge of CRF, exposure settings and scene details. This would enable us to by-pass the HDR image generation and tone reproduction operations. Employing the inverse tone reproduction techniques, we can make the LDR images compatible even with HDR displays. We shall present a variational solution to the LDR image generation from multi-exposure images corresponding to a static scene in the next chapter.
Chapter 3

A Variational Solution

3.1 Introduction

In this chapter, we shall present the general compositing problem and look at a possible formulation for solving the compositing problem. We shall then extend this approach to solve the HDR imaging problem for static scenes. We want to generate an LDR image of a static scene from a set of multi-exposure images assuming that the exposure times and the CRF are not known. We shall start with the motivation behind the proposed approach and then employ a variational framework to solve this problem. We finally arrive at an iterative solution to the problem which converges to the desired LDR image.

Consider the basic image formation equation discussed in chapter 2 and shown in Equation 3.1.

\[ I_k(x, y) = f(\Delta t_k E(x, y)) \]  

(3.1)

where \( I_k(x, y) \) represents the intensity values of the \( k^{th} \) image in the exposure stack with exposure time \( \Delta t_k \). Consider the case of static scene in which \( E(x, y) \) does not change. We would like to find the high contrast LDR image of the scene given \( I_k(x, y) \) when CRF \( f \) and exposure times \( \Delta t_k \) are not known. The approach presented in this chapter and the approaches in the next two chapters try to address this problem.

3.2 Background

Compositing multiple registered images of a scene can be used for a variety of applications in computer vision. The first work which addressed this idea provides a general overview of how
multi-scale image fusion can be used as a tool for compositing [81].

In this section, we develop an algorithm which performs compositing on multiple input images of the scene. Let the algorithm take $K$ images as input. Let $I_k(x, y)$ be the intensity value of the pixel at location $(x, y)$ of the $k^{th}$ image, where $1 \leq k \leq K$. Let $\hat{I}^{LDR}(x, y)$ be the unknown image to be constructed from $I_k(x, y)$.

General compositing problem can be formulated as shown in Equation 3.2.

$$\hat{I}^{LDR}(x, y) = \sum_{k=1}^{K} \alpha_k(x, y) I_k(x, y)$$

(3.2)

where $\alpha_k(x, y)$ is our representation for weighting function similar to the alpha matte function employed in alpha blending. This weighting function can be modeled in different ways. We shall consider the different ways and refine it as we arrive at the best design for the weighting function which would enable us to solve the HDR imaging problem. As a simple case, we can model $\alpha_k(x, y)$ as shown in Equation 3.3.

$$\alpha_k(x, y) = \begin{cases} 1 & \text{for } k = m \text{ (some index)} \\ 0 & \text{otherwise} \end{cases}$$

(3.3)

Here our objective is to select a value for a pixel from a specific observation and it is a pixel-based approach. This approach is carried out by many artists in computer graphics. In some cases, this approach is carried out automatically, when this becomes a problem of combinatorial optimization. However, we do not know which of the multi-exposure images are properly exposed in a particular region. Even white and black objects present in the scene might wrongly be classified as over- and under-exposure respectively. Given a single image, it is tough to identifying the pixel locations which are properly exposed. Even if identified using some recent approach like [82], solving a combinatorial optimization problem to arrive at the desired LDR image is a challenge by itself.

Alternatively, we can select $\alpha_k(x, y)$ using a region based approach as shown in Equation 3.4.

$$\alpha_k(x, y) = \begin{cases} 1 & \text{for } (x, y) \in R_k \\ 0 & \text{otherwise} \end{cases}$$

(3.4)

where, $R_m$ is the distinct region to be selected from the $k^{th}$ input image. The layer based approach explained earlier uses such a matte. As discussed earlier, detecting the regions of the multi-exposure images which are properly exposed is a challenge. There are algorithms which
correct the exposure in a particular region of an image ([83], [82]). These algorithms detect over- and under-exposed regions and correct the exposure in these regions. Our task is to stitch the properly exposed regions from all the multi-exposure images. As stitching of irregular regions into a single image without any seams is itself a challenge, we propose to employ a much simpler and effective approach to address this problem. The stitching of irregular regions is addressed in Chapter 7 when we need to deal with the real world scenes which are dynamic.

This problem of getting an LDR image from a set of multi-exposure images can be solved using the least squares [84]. The least squares approach yields a solution for the desired LDR image which minimizes the squared norm of the difference between the LDR image and the weighted sum of the multi-exposure images. The solution, thus obtained, may not be visibly pleasing as natural images are generally smooth. Hence we need to add some constraints on the desired LDR image so that the image depicts the high contrast real world scene. We need to adopt an approach based on calculus of variations in order to incorporate smoothness constraints on the desired LDR image [85]. We shall provide an overview of some of the classic vision problems which have been solved using the calculus of variations.

Calculus of variations has been widely used to solve key vision problems over the years. A typical variational formulation involves a data term and few constraints on the unknown function to be estimated. The major problem is shape from shading where one attempts to estimate the 3-D information of the object from a single image making use of the shading of the object surface. This classic problem in vision has been addressed using calculus of variations [86]. Another problem is photometric stereo where one tries to estimate the 3-D information of an object from multiple images of the object captured using illuminations from different directions. This problem can also be solved using the variational methods [1].

Depth of the objects located at different layers of the scene can be estimated from multiple defocused images of the scene. This task can also be achieved by variational methods ([87], [88]). Optical flow computation between successive frames of a video can be achieved using variational methods [89]. A recent highly accurate optical flow algorithm employs over-parameterization along with the variational methods to obtain dense optical flow [90]. Recently, optical flow was employed to create synthetic shutter speed frames from a video of a natural scene [91]. The variational methods can also be used to perform motion segmentation on a video [92].

We shall now formulate the compositing problem in a variational framework and try to
generate the desired LDR image from a set of multi-exposure images. The salient feature of the proposed approach is that the variational framework enables us to make the weighting function computation implicit. This gives us freedom to model the weighting function as a function of both the multi-exposure images as well as the unknown LDR image. However, the actual weights corresponding to the different multi-exposure images can not be computed separately. As the weighting function is a function of the unknown LDR image, we need to start with one of the multi-exposure images as the initial value for the LDR image and iteratively solve for the desired LDR image until convergence.

3.3 Proposed Solution

We require multiple, differently exposed images $I_k(x, y)$, where $1 \leq k \leq K$, as the $K$ input images corresponding to the static scene. We want to design an algorithm to produce the desired high contrast LDR image $\hat{I}^{LDR}(x, y)$ which is similar to the corresponding HDR image generated using standard HDR techniques ([6],[4]). We shall now design our weighting function $\alpha_{\text{var}}^k(x, y)$ with regard to this application. We want the selected pixel to have a high local contrast, yet blending smoothly across different regions while compositing from different images. Therefore we require a variational approach to solve this. The basic formulation of this problem can be designed as shown in Equation 3.5.

$$\hat{I}^{LDR}(x, y) = \arg \min_{\hat{I}} \left[ \int \int \left\{ \left( \hat{I}(x, y) - \sum_{k=1}^{K} \alpha_{\text{var}}^k(x, y)W_k(I_k(x, y)) \right)^2 + \lambda_{\text{var}} \left( \frac{\partial \hat{I}(x, y)}{\partial x} + \frac{\partial \hat{I}(x, y)}{\partial y} \right)^2 \right\} dx dy \right]$$ (3.5)

where $\lambda_{\text{var}}$ is a regularization parameter which appropriately weighs the smoothness term $(\frac{\partial \hat{I}(x, y)}{\partial x} + \frac{\partial \hat{I}(x, y)}{\partial y})^2$, applied everywhere and $W_k$ is the warping function which compensates for the motion of camera corresponding to the $k^{th}$ input image [93]. If we assume that the warping has already been carried out (note that this may require camera calibration) and incorporated into the input images $I_k(x, y)$, the basic formulation of the problem can be given as Equation 3.6.

$$\hat{I}^{LDR}(x, y) = \arg \min_{\hat{I}} \left[ \int \int \left\{ \left( \hat{I}(x, y) - \sum_{k=1}^{K} \alpha_{\text{var}}^k(x, y)I_k(x, y) \right)^2 + \lambda_{\text{var}} \left( \frac{\partial \hat{I}(x, y)}{\partial x} + \frac{\partial \hat{I}(x, y)}{\partial y} \right)^2 \right\} dx dy \right]$$ (3.6)
3.3 Proposed Solution

Our main objective in the development of this algorithm is to model the weighting function \( \alpha_k(x, y) \) which weighs each of the pixels of the input images \( I_k(x, y) \) in such a way so that it is amenable to an iterative solution, avoiding a combinatorial search, while \( \alpha_k(x, y) \) is still a data dependent term.

One easy way to model \( \alpha_k(x, y) \) is to pick one pixel from \( K \) images corresponding to the pixel location \((x, y)\) using Equation 3.3. So, \( \alpha_k(x, y) \) takes one of the two values in \( \{0, 1\} \) at any location \((x, y)\). This leads us to a combinatorial problem and the variational framework has no proper solution. This cannot be solved using the Euler-Lagrange formulation which offers an iterative solution. Hence, we relax the condition on the weighting function \( \alpha_k(x, y) \) allowing it to take real values in the range \([0, 1]\) subject to the constraint

\[
\sum_{k=1}^{K} \alpha_k(x, y) = 1. \tag{3.7}
\]

We propose a weighting function which would optimally weigh pixels of each of the \( K \) images so that the unknown image \( \hat{I}_{LDR}(x, y) \) is very close to the desired image with uniform illumination or contrast. The new weighting function is as shown in Equation 3.8.

\[
\alpha_k(x, y) = \frac{B_k(x, y)}{A(x, y)} \tag{3.8}
\]

where

\[
B_k(x, y) = (C_{var} + \sigma_k^2(x, y))(\hat{I}(x, y) - I_k(x, y))^2 \tag{3.9}
\]

\[
A(x, y) = \sum_{k=1}^{K} B_k(x, y) \tag{3.10}
\]

where \( \sigma_k^2(x, y) \) is the local variance around the pixel location \((x, y)\) in the \( k^{th} \) image, \( C_{var} \) is a real number meant to vary the influence of \( \sigma_k^2(x, y) \) on the weighting function as well as to prevent the condition of \( \alpha_k(x, y) \) being zero in all images at homogeneous regions, and \((\hat{I}(x, y) - I_k(x, y))^2\) is a measure of how the composited image is different from the \( k^{th} \) observation at a given location. One can replace \( \sigma_k^2(x, y) \) by other contrast measures such as local average of the gradient magnitudes for faster computation. The division by the sum over all the input images is done to ensure that the weights assigned to \( K \) images sum to unity for a given pixel \((x, y)\). Similar weighting function has been employed in colorization [94].

The choice of the above weighting function \( \alpha_k(x, y) \) stems from the fact that if the pixel in a particular observation has a high local variance \( \sigma_k^2(x, y) \), then it should have a higher
weight while compositing. Similarly, we want to emphasize selection of the over-exposed view at an under-illuminated pixel or the under-exposed view at an over-illuminated region. This is achieved through the choice of the second term \((\hat{I}(x, y) - I_k(x, y))^2\). The composited image would be away from being over- or under-exposed. This particular choice of the weighting function has an additional mathematical justification. If the \(K\) input images correspond to the same scene with different noisy (assumed Gaussian) observations, then the solution would be an optimal noise smoothing filter.

This optimization problem is solved using calculus of variations using the corresponding Euler-Lagrange equation \([1]\). The iterative discretized version of the solution is as shown in Equation 3.11 \([95]\).

\[
\hat{I}^{l+1}(i, j) = \bar{I}(i, j) - \frac{1}{\lambda_{\text{var}}} \left[ \sum_{k=1}^{K} B_k (i, j) I_k(i, j) \right] \left\{ 1 - \sum_{k=1}^{K} I_k(i, j) \left( \frac{A^l(i, j)D^l_k(i, j) - B^l_k(i, j)F^l(i, j)}{(A^l(i, j))^2} \right) \right\} 
\] (3.11)

where

\[
D^l_k(i, j) = 2 \left[ (C_{\text{var}} + \sigma_k^2(i, j)) (\hat{I}^l(i, j) - I_k(i, j)) \right] 
\] (3.12)

\[
F^l(i, j) = \sum_{k=1}^{K} D^l_k(i, j). 
\] (3.13)

The suffix ‘\(l\)’ denotes the value of the variable at the \(l^{th}\) iteration and \(\bar{I}(x, y)\) denotes the average value of \(\hat{I}(x, y)\) over its nearest 4-connected neighborhood.

We have experimented on the suitability of the proposed algorithm on a number of practical applications. For reasons of brevity, we show experimental results for only three specific applications that are highly relevant currently. We have used exactly the same computer code in all applications while maintaining the same value of the regularization parameter \(\lambda_{\text{var}}\).

As the natural images are generally smooth, we assumed that the final LDR image is also smooth. In case of noisy multi-exposure images, higher value of \(\lambda_{\text{var}}\) can be used to reduce noise in the final LDR image. For the results shown in the thesis, we assigned the value of 10 to \(\lambda_{\text{var}}\) to show the robustness of our variational approach. We observed that the higher values of \(\lambda_{\text{var}}\) tend to smooth out fine texture details which are important to preserve the overall contrast. Typical value of the parameter \(C_{\text{var}}\) used in our study is 100.
3.4 Other Applications

We shall now look at some of the applications in vision other than the HDR imaging where
the compositing techniques we have discussed thus far can be used. The results corresponding
to HDR imaging problem will be presented in chapter 6 along with that of two other methods
discussed in chapter 4 and chapter 5. We shall consider enhancing depth of field by combining
various images of the same scene with different images of lesser depth of fields as the first
application. As all-in-focus image is called a pin-hole image, we call this application as the
generation of pin-hole image. We shall then see how a random texture can be generated by
combining a set of heterogeneous textures. We also present an approach for compositing frames
of a video in which a dark scene is illuminated by a moving light source. We employ the
variational compositing method discussed in this chapter to illustrate these applications.

3.4.1 Generation of Pin-hole Image

Depth from defocus ([96],[97]) is an active area of research where two or more different obser-
vations are used to recover the dense depth map. One can use a suitable restoration technique
(by measuring the relative defocus) [98] to recover the corresponding pin-hole image. On the
other hand, if one is given a large number of such defocused observations along with individual
camera parameter settings, the corresponding depth recovery technique is called depth from
focus [99]. This approach was further enhanced and used to increase the depth of field of the
image by compositing multiple images of the scene with different layers of the scene in focus
([96], [98], [100]). Another approach for refocusing by active method involves projecting a
dot pattern at the time of capture [101]. We show that the proposed technique can be used to
combine these observations to obtain the pin-hole equivalent image very efficiently in the case
of not so large depth variation in the scene. This enables us to avoid the need to estimate the
blur kernels and perform deconvolution operations.

Consider multiple defocused images of a scene with each image having different parts of
the scene in focus. Our algorithm can combine these images to produce an image which has all
parts of the scene in focus. The weighting function designed earlier in this chapter will enable
us to pick up the pixel intensity at a point from that particular observation where the point is in
focus so that it has the maximum local contrast.

The images in Figures 3.1(a-c) (synthesized using POV-Ray toolkit [102]) of the scene
have three objects (a ball, a cuboid, and a cylinder). Each of these input images have one of the objects in focus. These images are composited by our algorithm and finally we get an image in Figure 3.1(d) in which all the three objects are in focus (pin-hole equivalent). In Figures 3.2(a-c), we have three defocused images of a real world scene as the input images. The scene consists of planer surfaces at three distinct distances. Different layer in the scene is focused in each of the input images in Figures 3.2(a-c). Our algorithm works on these images and generates an output image in which all parts of the scene are properly in focus as shown in Figure 3.2(d).

**3.4.2 Random Texture Generation**

In certain applications, such as template preparation in textile industry, it is often desired to combine different existing textures to generate a new random texture. Our algorithm can also be employed to generate a randomly textured image from multiple different textures. The nature
Figure 3.2: (a-c) Defocused input images with different depth layer in focus, and (d) compositing pin-hole equivalent image.
of the output image texture is controlled by varying the parameter $C_{\text{var}}$ defined earlier. This varies the influence of the local variance $\sigma_k^2(x, y)$ on the output image texture. The composited image will have the appearance of having picked up small regions from different constituent textures in a repetitive manner while picking locally high contrast pixels from the constituent images. Given a set of $J$ heterogeneous texture images, assume that we choose to composite $r$ textures at a time using variational compositing. This would lead to the generation of $J^C_r$ different textures.

Figures 3.3(a-e) show a set of texture images used as input. Figure 3.3(f) shows the composited output from five input images. One does see a certain correlation in the composited image with each of the five input images. In some sense, one can relate this application to the problem of cross-dissolve [103] in graphics.

Figure 3.3: (a-e) Different texture images as input, and (f) composited new texture. Data Courtesy: Brodatz Texture Database.
3.4.3 Illumination Compositing for Dark Scenes

Consider the situation where a completely dark and static scene which cannot be illuminated entirely with an available light source, is being videographed by moving the light source. The light source is moved in such a manner that parts of the scene become visible as soon as the light falls on them. This happens due to the fact that the field of view is much larger than that can be illuminated by the directional light source. See Figure 3.4.

![Figure 3.4: Thumbnails of some of the video frames used for compositing (left to right and top to bottom).](image)

We propose a novel application which uses appropriate compositing techniques to solve this unique problem. Our method composites specific video frames of a scene illuminated by a moving light source onto a single image consisting of all illuminated regions of the scene. The initial scene as mentioned earlier is dark with very low or no illumination and the moving light source acts like a painting brush on the scene to reveal details to be captured on a single image. The method is innovative in the sense that it does not require any information regarding camera calibration, modeling of the scene reflectance and light source estimation.

In the implementation from the video sequence of the scene, we use selected video frames to generate a single photorealistic composited image which would show the maximally illuminated scene. We also show the approximate path in which the light had transversed across the scene. Further, from the shape of the incident light beam on the scene, we estimate the inclination of the light source with respect to the scene. Considering the situation where a person needs to focus only on a particular section of the scene, we give the option to the user to select
a specific sub-path from the entire path of the moving light. Our approach would then generate an image corresponding to that part of the scene. Our approach has some similarities with the compositing based relighting techniques which however produce non-photorealistic images [104]. This method is different from the concept of VideoBrush, where field of view of the camera is enhanced through the process of scene mosaicking [105].

We use a flash light with a parabolic reflector and with minimal focusing artifacts in the light beam as our moving light source. The torch is a standard, commercially available flashlight which collimates the light beam within an angle of a 3 degree cone. The scene is captured using a digital video camera. The camera is held in the same position while the flashlight is moved along an arbitrary trajectory over the scene to illuminate it. The video is captured till the flash light has illuminated almost all parts of the scene. The frames of the video would then have bright circular or elliptical regions illuminating only those points of the otherwise unlit scene.

Having acquired the video for a duration of a few seconds, we are left with a large number of frames to be composited to obtain the maximally illuminated scene. While selecting the frames to be composited, we stressed on minimizing the number of frames in such a manner that the details of the entire scene are available. In other words, we require that the illuminated regions captured across some selected frames span the entire scene. In this manner, we would be able to produce an image which has all the regions of the scene illuminated.

![Figure 3.5: (a) The binarized frame, and (b) after closing operation.](image)

Thus to select a proper set of frames for compositing, we start with finding the coordinates of the center of light beam at each frame. Each frame is converted into a grayscale image and then to a binary image using Otsu’s threshold [106]. The binarized version may not clearly be an ellipse or a circle. This is due to the presence of edges of the objects in the scene (Figure 3.5(a)). We perform a morphological closing operation on the binarized image with a disc shaped structuring element to fill out any holes (Figure 3.5(b)). Using the Canny edge detector, the outer edge of the light beam region is obtained. The region inside the edge thus separates
the well illuminated region from the low illumination region. This provides us a set of edge points \((x^e_i, y^e_i)\) which can further be processed. We calculate the centroid \((\bar{x}, \bar{y})\) of the resulting binary image and the average spread of the light pattern. The root mean square spread \(\kappa_{rms}\) of the edge points around the centroid is given by Equation 3.14.

\[
\kappa_{rms} = \sqrt{\frac{\sum_{i=1}^{n} (x^e_i - \bar{x})^2 + (y^e_i - \bar{y})^2}{n}}
\]  

(3.14)

where \(n\) is the number of edge points. Video frames are selected using a greedy scheme so that the overlap of the illumination pattern is approximately half of each light beam pattern. In other words, the distance between the centroids for consecutively selected video frames should be approximately half of the sum of spreads. This is as shown in Equation 3.15 for two frames.

\[
\left| \frac{\kappa_{rms,1} + \kappa_{rms,2}}{2} \right| - \sqrt{(\bar{x}_1 - \bar{x}_2)^2 + (\bar{y}_1 - \bar{y}_2)^2} \leq \vartheta
\]

(3.15)

where \(\kappa_{rms,1}\) and \(\kappa_{rms,2}\) are the beam spreads, \((\bar{x}_1, \bar{y}_1)\) and \((\bar{x}_2, \bar{y}_2)\) are the centroids corresponding to the two candidate frames and \(\vartheta\) is a small positive real number. We have considered the above procedure only for the consecutively selected frames and not for all the frames so that a real time and sequentially updatable matte can be prepared. If such a constraint is not required to be met one can search over the entire image space to drop additional redundant frames. We composite all the selected frames to get a completely illuminated image of the scene which is used as a reference image while showing the light path as traced out by the respective centroids.

With the selected video frames at hand our main task is to composite them onto a single image that will comprise of all the details of the scene. We found that the methods which composite scene using multi-exposure images are quite suitable for this application. Properly exposed regions of the scene will have high contrast, well-exposedness and no saturation. These properties are quite similar to those of the regions which are illuminated by the flash light in our application. We decide to employ the weighting function decided earlier in this chapter to perform compositing of the selected frames.

Sometimes it is quite important in selected applications that the user needs to observe in detail only a specific section of the scene. Keeping this aspect in mind, we designed an interface wherein the user can select any particular segment of the path of the moving light. The calculated centroids are denoted on the final image using numbers in an increasing order. The increasing order represents the path traveled by the light. The path is thus marked by the red line (Figure 3.6(a)).
Now, if the user wants to select a sub-path along the light path, clicks can be made near any subset of points. The program automatically calculates the composite image using corresponding frames. Finally, the original path will also appear and one can again choose a different subset of points on the same image, if so desired [107].

![Figure 3.6: (a) Composited image with the maximally illuminated region along with the path in which the flashlight moved, and (b) Composited image along with the selected sub-path.](image)

Further, we use the estimated centroids in all the frames to get the path of the moving flashlight (Figure 3.6(a)). We can composite all these images also using the gradient domain compositing to be discussed in Chapter 5. The composited image along with the estimated flashlight path are as shown in Figure 3.6(a). The path is indicated in red color in the figure. Now, we let the user select an arbitrary sub-path. The resultant composited image along with the selected sub-path (yellow) is as shown in Figure 3.6(b).

The scene considered so far had a planar surface. This enabled us to estimate the centroids and the path of moving light source precisely. Figure 3.7 shows the composited image for frames from an outdoor surveillance video with large variations in depth.

Figure 3.8 shows the composited image of a scene comprising of some objects, wall surfaces almost perpendicular to each other and a cable. This cable gives rise to shadow on the wall behind it. Since the shadow is dependent on the direction of the light source, we notice that the composited image in Figure 3.8 has multiple shadows. This artifact is similar to the ghosting effect one gets while compositing multi-exposure images of a dynamic scene [4].

In Figure 3.9, we observe that there are regions in the scene which produce glare effects
Figure 3.7: Composited image from a surveillance video.

Figure 3.8: Composited image of a scene showing multiple shadows of the cable on the wall.
when illuminated due to specularity of the surface.

Figure 3.9: Composited image of a scene which shows glare effects from an oil painted wall.

The proposed technique is a very effective tool that can composite a video of a dark scene illuminated part by part by a moving light source. In future, we would like to extend the capability of the tool to determine the direction of incidence of the light beam to the scenes which have high depth variations. The knowledge about the direction of the incident light, image plane and scene modeling would help us to remove glares and multiple shadows.

3.5 Discussion

The variational approach converges to a high contrast LDR image given a set of multi-exposure images corresponding to a static scene. The approach is simple and does not require estimation of the weighting function explicitly. The approach is also simple to implement. However, calculating local variance is a computationally expensive operation and can be substituted by some other local contrast measure like local average of gradient magnitudes. This approach also leads to the blurring of the edges in the final image as the smoothness based regularization term is applied globally. This problem can be addressed by using the work by Geman and Geman as the regularization term instead of the first order derivatives of the unknown [108].
In the next chapter, we shall present an alternate approach which involves explicit computation of the weighting function. As we are going to use edge preserving filter for determining the weights corresponding to the different multi-exposure images, we show that the edges will not be blurred as in this approach.
Chapter 4

Edge Preserving Filter based Solution

4.1 Introduction

Edge preserving filters are non-linear filters employed to smoothen out small textures in the image while preserving strong edges. These filters can be used to manipulate images without any blocky artifacts. Edge preserving filters have wide range of applications in computer vision, especially computational photography. We shall explain how these edge preserving filters can be employed to reconstruct LDR image of a scene from multi-exposure images. To start with, we shall provide the background regarding various edge preserving filters describing many commonly used classes of edge preserving filters. We then discuss a type of edge preserving filter called bilateral filter and show how it can be used to generate a high contrast LDR image from a set of multi-exposure images. This approach is simple and faster compared to the variational approach discussed in the previous chapter.

4.2 Edge Preserving Filters

Edge preserving filters have a strong connection with the scale-space techniques used in vision. The conventional Gaussian kernel is a low pass filter which does not take into account strong edges in the image while blurring it. Alternatively, one can use anisotropic diffusion to implement scale-space in order to detect pixel locations which have strong edge information and preserve them while smoothing [109]. Edge preserving smoothing can also be achieved using an adaptive smoothing operation which involves averaging a function taking into account its continuity at a given location [110]. Adaptive smoothing also leads to a scale-space representa-
Another important edge preserving filter is the popular non-linear filter called bilateral filter. Bilateral filtering introduced by Tomasi and Manduchi in 1998 [111] is a non-linear technique which employs product of a Gaussian kernel in the spatial domain and a Gaussian kernel in the intensity. This makes the filtering operation edge preserving and only the fine textures present in the image are smoothed out.

Let \( I(x, y) \) be the image which needs to be operated by a bilateral filter. Let \( G_\phi \) be the 2-D Gaussian spatial kernel and \( G_\rho \) be the 1-D Gaussian range kernel (on the intensity values). If we denote the bilateral filtered image by \( I^{BF}(x, y) \), the bilateral filtering operation is as shown in Equation 4.1.

\[
I^{BF}(x, y) = \frac{\sum_{x'} \sum_{y'} I(x', y') G_\phi(x - x', y - y') G_\rho(I(x, y) - I(x', y'))}{\sum_{x'} \sum_{y'} G_\phi(x - x', y - y') G_\rho(I(x, y) - I(x', y'))}
\]

(4.1)

\[
G_\phi(x, y) = \exp\left(-\frac{x^2 + y^2}{2\phi^2}\right)
\]

(4.2)

\[
G_\rho(a) = \exp\left(-\frac{a^2}{2\rho^2}\right)
\]

where \((x', y')\) corresponds to the neighborhood of pixel location \((x, y)\), \(\phi\) denotes the extent of the spatial kernel and \(\rho\) denotes the minimum amplitude to be defined as an edge. We shall illustrate the edge-preserving filtering operation using this bilateral filter.

Consider the L channel (of CIE Lab space) of an image which represents the gray values present in the image as shown in Figure 4.1(a). Applying the bilateral filter to this image provides us the edge-preserved image shown in Figure 4.1(b). As one can observe, the bilateral filtering operation preserves strong edges and removes all small texture details. This image is generally called the base layer obtained using the edge-preserving filtering operation. The absolute of the difference image between Figure 4.1(a) and 4.1(b) is shown in Figure 4.1(c). One can see the finer details obtained in this image. This image is usually referred to as the detail layer of the given image. An image can be split into a base layer and a detail layer using an edge-preserving filter. These layers together capture both the coarse and fine details present in a given image.

Another detail preserving technique used to compress high contrast scenes for the low contrast displays is the low curvature image simplifier (LCIS) [112]. This approach is a variation
4.2 Edge Preserving Filters

Figure 4.1: (a) L-Channel of one of the multi-exposure images \( L \), (b) after bilateral filtering \( L^{BF} \), and (c) the difference image \( |L - L^{BF}| \). Image intensities are scaled for display purpose. (Data Courtesy: Erik Reinhard, University of Bristol.)
of the anisotropic diffusion equation discussed earlier with a gradient threshold parameter used to change the details which have to be preserved. This parameter is analogous to the range kernel used in the bilateral filter. The application of LCIS is in HDR tone mapping.

The fundamental relation between bilateral filter and anisotropic diffusion was addressed using adaptive smoothing [113]. It is first shown that anisotropic diffusion and adaptive smoothing are equivalent operations as adaptive smoothing is just an implementation of the anisotropic diffusion. Then it is shown that bilateral filtering and adaptive smoothing are equivalent as the weights used in both the schemes are the same. Mean shift can also be employed to perform edge preserving smoothing [114]. This filter is also known as local mode filter [115]. The relation between mean shift filter and bilateral filter can be established by considering a restricted mean shift filtering method which operates only on the range values [116]. A similar relation is also established in [115]. The bilateral filter has also close connection with Bayesian approach and can be derived from a Bayesian framework [117].

There are a variety of other edge-preserving filters which have close connection with scale space and bilateral filter. The trilateral filter uses a filter window tilted by the gradient to achieve better denoising [59]. The trilateral filter can also be used for HDR tone mapping as discussed in chapter 2. The performance of bilateral filter can be improved by making it operate at various scales [118]. This helps us separate the base layer from the detail layer at every scale. This multi-scale bilateral filter can be used to enhance detail while combining images of the same object captured under different lighting conditions.

Edge preserving filter at various scales can also be built using weighted least squares (WLS) approach [119]. The results of this approach are comparable to that of the bilateral filter, but this approach is computationally expensive. WLS approach can be used to solve applications such as multi-scale tone manipulation, detail exaggeration and HDR tone mapping. A more recent multi-scale edge-preserving filter is designed using the local extrema of a signal [120]. This approach helps one treat the edges and textures of high contrast differently and hence preserve details better by capturing oscillations. It can be used in HDR image tone mapping and to boost coarse and fine details of the image separately.

Among these edge-preserving filters, bilateral filter is the most commonly used tool for a variety of applications. Some of the applications of bilateral filter include HDR image tone mapping [58] already discussed in chapter 2. The look of a photograph can be enhanced by just processing detail layer obtained using bilateral filter. This technique can be used to generate
high contrast images which are visually very pleasing [121]. The digital image captured with flash have less noise and do not capture the ambient illumination of the scene. The same scene captured without flash will be able to capture the ambient illumination of the scene and will have more sensor noise. Combining flash and no-flash images using bilateral filter will help one to generate noise-free image of the scene with actual illumination ([122], [123]).

The bilateral filter is a non-linear filter with $O(n^2)$ complexity where $n$ is the number of pixels in the image. We shall now discuss some of the methods which help to accelerate the computation of bilateral filtering operation. The common approach involves solving the bilateral filter in the frequency domain which reduces the complexity to $O(n \log n)$ [58]. This approach was further enhanced to result in a more robust $O(n \log n)$ algorithm for computing bilateral filter [124]. A different approach for bilateral filter with similar complexity was proposed in [125]. The constant time $O(1)$ bilateral filter computation algorithms are recently proposed in ([126], [127]).

4.3 Bilateral Filter based LDR Image Generation

Consider a set of multi-exposure images corresponding to a static scene. We shall discuss how edge-preserving filters can be employed to generate a high contrast LDR image of the scene. We shall illustrate the compositing methodology by considering the special case of bilateral filter though any of the edge-preserving filters discussed in the previous section can be used.

The fundamental goal of compositing is to obtain mattes for the input images so that the final image has the desired features. The mattes are obtained using a function called matting function which enables one to generate an appropriate matte for a given input image. It is mandatory that the matting function must be a function of the input image itself or its features for automatic compositing approach.

The desired qualities for the matting function for solving HDR problem in non-irradiance domain is that the final image must have proper contrast, well-exposedness and should not have saturation in the upper and lower intensity values [128]. For a particular pixel location, the matting function must assign higher weights for intensity values from those images which have higher contrast, are well-exposed and have minimal saturation. We will now design our matting function based on bilateral filtering keeping these criteria in mind.

Our objective is to composite multiple differently exposed images into a single image.
which is as close as possible to the original scene. Bilateral filtering serves our purpose in the
design of appropriate mattes to achieve this task. Consider a gray scale image and the bilateral
filtered version of the same image. The strong edges in the image are preserved while weak
edges or textures are completely smoothed out. If we calculate the difference between the input
image and the bilateral filtered image, it would only have weak edges or texture information
which are crucial for compositing purposes.

Weak edges and textures in images are the first casualty whenever under (or over) exposure
takes place. These weak edges are lost locally. Hence they serve as ideal markers to detect over
(or under) exposure. If at a given location, the weak edges are relatively strong, compared to the
rest of observations, this region in that particular observation should be given a higher weight
while compositing. This is the motivation behind this work.

Figure 4.1 shows our approach for the L channel corresponding to one of the multi-exposure
images. The bilateral filtered image (Figure 4.1(b)) shows that the small textures of the original
image (Figure 4.1(a)) have been removed by the operation while the strong edges are preserved.
The difference image in Figure 4.1(c) shows that the lost details due to bilateral filtering can
be recovered. This provides the motivation for deciding upon the weighting function \( \alpha_{BF} \) for
our approach.

Consider \( K \) multi-exposure images. We design our weighting function as the function of
the difference image as shown in Equation 4.3. This weighting function is directly dependent
on the difference image between the original image and the bilateral filtered image. As already
discussed, this would enable us to give more weight to those pixel locations of the given multi-
exposure image with more texture details. Also, the edge preserving nature of the bilateral
filter prevents us from over-weighing the edges. The final LDR image will have all the edges
preserved while more contrast being obtained in the less textured regions which are lost due to
over- and under- saturation.

\[
\alpha_{m}^{BF}(x, y) = \frac{(C^{BF} + |I_m(x, y) - I_m^{BF}(x, y)|)}{\sum_{k=1}^{K}(C^{BF} + |I_k(x, y) - I_k^{BF}(x, y)|)} \tag{4.3}
\]

where \( \alpha_{m}^{BF}(x, y) \) is the matting function, \( I_m(x, y) \) are the intensity values of the multi-exposure
images, \( I_m^{BF}(x, y) \) are the corresponding bilateral filtered images, and \( C^{BF} \) is a real number
(assigned a value of 70 in this study). The composited LDR image is given by Equation 4.4.

\[
\hat{I}(x, y) = \sum_{k=1}^{K} \alpha_{k}^{BF}(x, y)I_k(x, y) \tag{4.4}
\]
where
\[ \sum_{k=1}^{K} \alpha_k^{BF}(x,y) = 1. \] (4.5)

The parameter \( C^{BF} \) has two roles to play. It prevents numerical instabilities at homogeneous regions. It can also be used as a possible tuning parameter if certain interactivity is desired by an user.

### 4.4 Implementation

The implementation of the designed algorithm is quite straightforward. We use the approximate bilateral filter by Paris and Durand [124]. Other faster versions of bilateral filter can also be used to composite the images in lesser time ([126], [127]). For an images \( I_k(x,y) \) of size \( M \times N \), we use the following functions to obtain \( \rho \) and \( \rho_k \) which represent the standard deviations for spatial and range Gaussian functions, respectively, in a bilateral filter.

\[ \rho = K_1 \times \min(M,N) \] (4.6)

\[ \rho_k = K_2 \times (\max(I_k(x,y)) - \min(I_k(x,y))) \] (4.7)

where \( K_1 \) and \( K_2 \) are positive real constants. We vary \( K_1 \) and \( K_2 \) to obtain varying amount of smoothing and to vary the threshold for retaining edges, and they are assigned the values of 1 and \( \frac{1}{10} \), respectively, in this study [129].

One way to implement our algorithm for color images is to operate on R,G, and B channels separately. Alternately, one can work in the CIELab space and work on the L channel alone for each image to obtain the respective matte. This matte can then be used to composite L,a,b channels. This approach would reduce the computation time marginally as only one bilateral filtering operation needs to be performed per image. The results are found to be quite similar to that when R, G, and B channels are processed separately to generate the weights.

### 4.5 Discussion

We have shown how edge preserving filters can be used to weigh the multi-exposure images appropriately to obtain a high contrast LDR image. The approach can also be applied at various
scales using any of the multi-scale edge preserving filters for improved results. With new fast approximations of the non-linear filters such as the bilateral filters, one can even aim for real time compositing of multi-exposure images in the capture device itself.

As we operate only on the texture information for determining the weights, this approach would lead to slight over- or under- saturations of certain regions in the final LDR image. Further the real constant used in the weight needs to tuned for each set of multi-exposure images to get the best results. These overheads can be avoided when this problem is addressed in the gradient domain. We shall see in the next chapter how best results out of the approaches for static scenes can be obtained while compositing multi-exposure images in the gradient domain.
Chapter 5

Gradient Domain Solution

5.1 Introduction

The problem of compositing multi-exposure images of a scene to generate the high contrast LDR image can also be addressed in the gradient domain. The gradient domain processing enables one to seamlessly reconstruct the scene. In this chapter, we shall present an overview of the gradient domain processing methodology and some of the applications already employing this technique. We shall then present the automatic compositing technique which can be performed on the gradients corresponding to multiple images. We then extend this technique to composite multiple differently exposed images of a static scene with high dynamic range. We show that this approach produces the best results compared to the methods discussed in previous chapters.

5.2 Gradient Domain Processing

Consider an image $I(x, y)$ and the corresponding gradient field $g(x, y)$. The gradient field $g(x, y)$ is obtained from the image $I(x, y)$ through the $\nabla$ operator considering $I(x, y)$ as a scalar field. This relation is shown in Equation 5.1.

$$g(x, y) = \nabla I(x, y).$$  \hspace{1cm} (5.1)

The vector field $g(x, y)$, thus obtained, is generally referred to as the conservative field as the curl of this gradient field is zero. However, when we perform some processing on $g(x, y)$, the resulting gradient field $\tilde{g}(x, y)$ may no longer be conservative. This is due to the fact that
the curl operator on \( \tilde{g}(x, y) \) may lead to a non-zero value. The recovery of the scalar field corresponding to \( \tilde{g}(x, y) \) is a non-trivial problem.

Suppose we want to find a scalar field \( \tilde{I}(x, y) \) which is closest to the modified vector field \( \tilde{g}(x, y) \), we would like to solve the least squares estimation problem shown in Equation 5.2.

\[
\min_{\tilde{I}} \int_{x} \int_{y} \| \nabla \tilde{I}(x, y) - \tilde{g}(x, y) \|^2 \, dx \, dy \tag{5.2}
\]

This problem can be addressed using calculus of variations and employing Euler-Lagrange equation, one would get a Poisson equation shown in Equation 5.3.

\[
\nabla^2 \tilde{I}(x, y) = \nabla \cdot \tilde{g}(x, y) \tag{5.3}
\]

where \( \nabla^2 \tilde{I}(x, y) \) is the Laplacian corresponding to the scalar field \( \tilde{I}(x, y) \) and \( \nabla \cdot \tilde{g}(x, y) \) denotes the divergence of the modified vector field \( \tilde{g}(x, y) \). The Poisson equation has physical significance such as electric field at a given point in space due to a number of charges can be estimated by solving this equation. For more detailed account of the significance of the Poisson equation, one can refer to the classic book by Schey [130]. This elliptic partial differential equation needs to be solved using appropriate boundary conditions in order to recover the scalar field which is closest to the vector field.

There are two types of boundary conditions one normally uses to solve a given partial differential equation (PDE). When the function to be estimated is known in the boundary, we can use those values at the boundary and estimate the function only on the inside region. This boundary condition is known as Dirichlet boundary condition. This type of boundary condition is used in seamless cloning where one wants to blend an irregular patch of one image onto another image [131]. When the function to be estimated is not known on the boundary, we assume that the derivative of the unknown function is orthogonal to the normal vector to the boundary. This boundary condition is known as Neumann boundary condition. The use of either the Neumann or the Dirichlet boundary condition depends on the requirements of the given application to be solved. Since the values at the boundary are unknown for the LDR image, we employ Neumann boundary condition in the proposed approach.

The Poisson equation can either be solved by direct methods or by iterative methods. The iterative solutions are more accurate compared to the direct solution. However, the direct solutions are faster. The direct method for solving Poisson involves modelling the PDE along with its boundary condition as a linear system. Then the linear system can be solved using
5.2 Gradient Domain Processing

traditional methods like successive over-relaxation to obtain the desired scalar field. An account of the direct solutions for the Poisson equations can be found in ([132], [133]). Recently, iterative solutions also became popular and are being widely used ([134], [135]). Most of the real-time applications prefer to use direct Poisson solver to recover the scalar field from a non-conservative vector field.

The Poisson equation can be improved by some variations in the formulation. Including a data term along with the least squares term would lead to a screened Poisson equation. This equation is explained and its applicability to various applications in computer vision is demonstrated by Bhat et al. [136]. They provide a spatial and Fourier domain analysis of the screened Poisson equation to arrive at a closed form solution and the applications addressed include image sharpening and image blending.

More generalized framework from which screened Poisson equation and the Poisson equation can be obtained is discussed in [137]. The tool, known as GradientShop, has a framework which involves separate gradient term each weighed using weighting functions. This optimization framework can then be solved to address various image and video processing applications. The design of the weighting functions in different ways lead to solutions of different applications such as image sharpening, non-photorealistic rendering, deblocking and pseudo relighting.

Another popular commercial software which employs a variant of Poisson equation is Adobe® Photoshop [75]. The technique behind the healing brush tool for image blending in Photoshop is explained in [138]. The detailed account of the applications of Poisson equation for image manipulation tasks such as seamless cloning, texture flattening, local illumination change and local color change are explained in [131]. This work popularly known Poisson image editing led to host of other applications solved in the gradient domain. Seamless cloning involves smooth blending of a region of one image into another without any introduction of visible seams. This application is a challenge to solve in the spatial domain as images of natural scenery have lot of variations in terms of texture, gray-level and color. However, the gradient domain solution offers a seamless generation of the composite image from two images when appropriate boundary conditions are used. An alternate method for image editing based on coordinates can be used for seamless cloning [139].

There are a number of computational photography applications which employ processing in the gradient domain. We shall now present an overview of some of the key applications which use gradient domain processing. As discussed in chapter 2, gradient domain processing has
been used in tone mapping [62]. In this work the gradients of the logarithm of the HDR image is computed. The idea here is to design a weighting function which penalizes large gradients while preserving the small gradients. This leads to effective compression of the dynamic range of the given HDR image.

Image stitching to generate panorama from a set of successive images with a spatial shift can also be achieved in the gradient domain [140]. The similarity between a given pair of successive images is taken into account while defining the cost function in the gradient domain. Further, the visibility of seams along the stitching boundary due to photometric and geometric inconsistencies can be minimized by solving the optimization problem in the gradient domain. Quad-trees can be employed to achieve a faster image stitching with reduced memory requirements [141].

Extraction of foreground and background separately from a natural image along with the matte is a challenging problem in vision. Most of the techniques assume the user to provide a region which is a definite foreground and another region which is a definite background. The remaining region adjoining the boundary is the ambiguous region where matte needs to be estimated. This information is called trimap. Given a trimap and the corresponding natural image, one can employ processing in the gradient domain to extract foreground, background and the matte. This approach is popularly known as Poisson matting [142].

Non-photorealistic rendering involves generation of images which cannot normally be captured using a camera. One example is the combined look of the same scene during day and night. Due to sunlight finer details of the scene will not be visible that well while these details will be more prominent in nights because of artificial illumination. One can combine two images of the same scene captured at day and night to reproduce better visual effects. This application can be solved in the gradient domain by an appropriate designed weighting function [143].

The separation of an image into reflectance and illuminance components provide the intrinsic images corresponding to the given image. The intrinsic images can be estimated from a sequence of images of the same scene with different illumination by gradient domain techniques [144]. A camera which captures the gradients directly instead of the intensity values are useful to capture HDR images of static scenes. The image captured as gradients can be subjected to a Poisson solver in order to reconstruct the high contrast image [145]. Other applications which involve processing in the gradient domain include interactive digital photo montage [146], re-
moving photography artifacts such as self-reflections from mirror [147], edge suppression to extract foreground from background [148], and removing shadows from images [149].

5.3 Gradient Domain Solution

In this section, we shall discuss about the approach to solve the HDR problem in the gradient domain. Assume that we have a set of multi-exposure images of a static scene. We would like to composite these images in the gradient domain and reconstruct the high contrast LDR image of the scene. This process involves two steps. The gradients corresponding to the multi-exposure images are first pre-processed in order to reduce saturated regions. These modified gradients are then composited using an appropriate weighting function in the second step. A preliminary solution was addressed by us in [150].

5.3.1 Pre-processing of Gradients

This step is meant to process the gradients in such a way that saturation is reduced if the patch is over- or under-exposed. This is done by carefully modifying their corresponding gradient fields. A method to correct the nearly saturated regions in the image domain was recently proposed by Masood et al. [83] and Guo et al. [82]. We perform a similar operation in this work in the gradient domain.

Given a set of multi-exposure images, we subject their gradients to an illumination change function specified in [131]. This operation is performed on each patch separately to brighten the less exposed images and darken the more exposed images as a pre-processing operation. We modify the gradient fields corresponding to each patch location as shown in Equation 5.4.

\[ \tilde{g}_k(x, y) = \left( \frac{Mean^p[\nabla I_k(x, y)]}{|\nabla I_k(x, y)|} \right)^{\beta_k} \nabla I_k(x, y). \]  \hspace{1cm} (5.4)

where \( Mean^p[|\nabla I_k(x, y)|] \) is the mean gradient magnitude of the \( p^{th} \) patch in the \( k^{th} \) image, \( \nabla I_k(x, y) \) is the gradient vector of the \( k^{th} \) image, \( \tilde{g}_k(x, y) \) is the corresponding modified gradient vector, and \( \beta_k \) is a real number signifying the relative exposure time corresponding to each image. This parameter has to be modeled in such a way as to reduce the over- and under-exposure in the patches. We define \( \beta_k \) as given in Equation 5.5.

\[ \beta_k = 0.5(0.5 - Mean^p[I_k(x, y)]) \]  \hspace{1cm} (5.5)
where $\text{Mean}^p[I_k(x, y)]$ is the mean intensity corresponding to the $p^{th}$ patch in the $k^{th}$ image. This formulation enables us to penalize the gradients whenever there is over- or under-saturation in the intensity domain.

### 5.3.2 Gradient Domain Compositing

The gradient domain compositing involves weighing the gradients of input images at each patch location appropriately to obtain the resultant gradient field $\hat{g}(x, y)$. The compositing process can be better explained by Equation 5.6.

$$\hat{g}(x, y) = \sum_{k=1}^{K} \alpha^\text{grad}_k(x, y)\hat{g}_k(x, y).$$

(5.6)

where $\alpha^\text{grad}_k(x, y)$ is the weighting function which weighs the gradients $g_k(x, y)$ and

$$\sum_{k=1}^{K} \alpha^\text{grad}_k(x, y) = 1.$$

Consider the images $I_k(x, y)$ to have intensity values in the normalized range $[0, 1]$. We composite the modified gradients $\hat{g}_k(x, y)$ using the weighting function in Equation 5.7.

$$\alpha^\text{grad}_k(x, y) = \delta\left(b_k(x, y) + c_k(x, y)\right) + (1 - \delta)\left(b_k(x, y)c_k(x, y)\right).$$

(5.7)

where $0 \leq \delta \leq 1$, $b_k(x, y) = e^{-\frac{(I_k(x, y)-0.5)^2}{2\times0.2^2}}$ is the brightness term used to provide less weight to over-exposed and under-exposed regions and $c_k(x, y)$ is the normalized local contrast term which equals the mean of the gradient magnitudes in the neighborhood of $(x, y)$ as defined in Equation 5.8.

$$c_k(x, y) = \frac{1}{N_{\Omega}} \sum_{k=1}^{K} \hat{g}_k(x, y)$$

(5.8)

where $N_{\Omega}$ is the neighborhood of the pixel location $(x, y)$. We use the 8-neighbors in this work when $N_{\Omega} = 9$. Equation 5.7 has significance in always assigning more weight to that image which has less saturation and more local contrast. Note that the term $b_k(x, y)$ is high whenever the observation is in the middle of the available dynamic range, providing a larger weight for the weighting function. Similarly, if the local contrast $c_k(x, y)$ is high, the gradient is given a higher weight. Since presence of noise in the observation $I_k(x, y)$ may adversely affect the local contrast, the measure is smoothed out locally through an averaging process. We found that the sum and the product of the brightness and the contrast terms are both equally effective in achieving the task. Hence, a convex combination of the sum and the product is used in Equation 5.7.
5.3.3 LDR Image Reconstruction

We employ overlapping patches with one pixel overlap in both horizontal and vertical directions. The composited gradient field $\hat{g}(x, y)$ can be arranged on the original image grid and it will not be conservative. In other words, there will not be a unique scalar field which can be determined by integrating the gradient field $\hat{g}(x, y)$. This vector field must be subjected to Poisson solver with Neumann boundary conditions. One can either use an iterative solver or a direct solver to reconstruct the scalar field closest to this vector field. We employ a direct Poisson solver in order to achieve this task as it is faster compared to other solvers ([132], [133]). The scalar field obtained as a result of the Poisson solver will represent the desired LDR image of the scene when mapped to the range $[0, 255]$. This LDR image will have high contrast present in both the brightly and poorly illuminated regions of the scene.

5.4 Discussion

The gradient domain compositing method yields better compositing for a set of multi-exposure images corresponding to a static scene. This method composites gradient fields corresponding to multi-exposure images and thereby enables one to seamlessly reconstruct the final LDR image. One can even use overlapping patches for better results as discussed in the chapter 7 for dynamic scenes. Further, Poisson solver can be made to reconstruct the LDR image faster by using multigrid methods. We employ this gradient domain compositing for multi-exposure images corresponding to dynamic scenes in Chapter 7 as it provides better results. The results corresponding to different approaches for static scenes are analyzed in the next chapter using a dynamic range independent quality metric.
Chapter 6

Results and Discussions

In this chapter, we shall present the results of the proposed methods in the last three chapters corresponding to a set of multi-exposure images for a static scene. The purpose of putting all the results in a single chapter is to enable us to compare the performances of these methods. There is also an existing method which performs compositing on the Laplacian pyramid known as exposure fusion which was developed at the same time as the variational approach [128]. This method also does not assume the knowledge of CRF and exposure settings. We shall compare the results of the various approaches described in the last three chapters with this method. Since many of the results and comparisons are shown in color, the readers are requested to go through the softcopy of this thesis.

6.1 LDR Image Generation for Static Scenes

Consider the multi-exposure images of a static scene captured with increasing exposure times as shown in Figure 6.1. This scene has a very high dynamic range as the lizard and the surrounding regions are poorly illuminated while the foliage in the background is brightly illuminated. As can be seen, the lizard and rock can be better depicted through images captured with high exposure times while the details in the foliage can be better depicted through images captured with low exposure times. We do not assume the knowledge of CRF and exposure times. We would like to composite these images using the techniques discussed earlier to generate an LDR image.

We also consider another competitive method called exposure fusion [128]. Under this
scheme, the unnormalized matte function $\alpha_{k}^{EF}(x, y)$ for the $k^{th}$ frame is given by Equation 6.1.

$$\alpha_{k}^{EF}(x, y) = CON_{k}(x, y) \times SAT_{k}(x, y) \times EXP_{k}(x, y)$$

(6.1)

where $CON_{k}(x, y) = |\nabla^{2}(I_{k}^{gray}(x, y))|$ is the absolute value of Laplacian on the grayscale version of the $k^{th}$ image and provides a measure of local contrast,

$SAT_{k}(x, y) = \text{stddev}(I_{R}^{k}(x, y), I_{G}^{k}(x, y), I_{B}^{k}(x, y))$

is the standard deviation within the $R, G, B$ channels of the $k^{th}$ image measuring the saturation level, and $EXP_{k}(x, y) = \prod_{Q \in \{R, G, B\}} \exp\left(-\frac{(I_{Q}^{k}(x, y) - 0.5)^{2}}{2 \times 0.012^{2}}\right)$ is the measure of well-exposedness.

We can either use a real world HDR scene or a HDR image for evaluating the proposed approaches. Assuming a CRF one can generate multi-exposure images from a given HDR image. We can then use any of the approaches discussed in the previous chapters to generate the LDR image of the scene. These images will have high contrast and we can use an inverse tone mapping operator to generate the nearest HDR image to the LDR image. This would enable us to check how well the LDR images are able to generate the full dynamic range of the scene. However, we have a dynamic range independent metric by Aydin et al. to evaluate the LDR images with that of the HDR image [151]. Further, there is still not a single inverse tone mapping operator which enables one to preserve full contrast information. This approach can be used only when a reliable inverse tone mapping operator is designed which can faithfully convert an LDR image into an HDR image without any distortion. Therefore, we would use only images of a real world HDR scene captured using a common digital camera and use them for the evaluation of the proposed approaches.

In this chapter, we consider 4 sets of multi-exposure images corresponding to static scenes. These sets of images are labeled as lizard, desk, star and house. For the first 3 sets, we know the exposure times used to capture the images and hence HDR image of the scene can be generated. We use “Merge to HDR” tool in Adobe® Photoshop CS4 to generate the HDR images corresponding to the first 3 sets. We use the open source HDR imaging software called pfs tools discussed in Chapter 2 for tone mapping the HDR images using different tone mapping operators [78]. The parameters of the different tone mapping operators are usually set to default values in pfs tools and they are changed only when the default values produce visually bad LDR images. The tone mapping operators and the values of the parameters used for generating the
LDR results are tabulated in Tables 6.1 and 6.2. In the table, for the ‘lizard’ and the ‘star’ set of multi-exposure images, we employ Ashikhmin’s tone mapping function along with the local contrast preservation but without the threshold vs intensity function [54]. For the ‘desk’ set of multi-exposure images we employ the simple tone mapping function suggested in [54].

Apart from the tone mapping operators we also consider the LDR results obtained using the methods discussed earlier - variational solution (Chapter 3), edge-preserving filter based solution (Chapter 4) and gradient domain solution (Chapter 5). We also consider the LDR image obtained using the exposure fusion method discussed earlier [128]. We present the LDR image results obtained using these approaches with the parameters being assigned the same values for all the datasets ($\lambda^\text{var} = 10$, $C^\text{var} = 100$, $C^\text{BF} = 70$). These approaches do not require the knowledge of exposure times. Hence the LDR image results are shown for set 4 using the approaches which involve direct LDR image generation without going through the HDR imaging methodology.

Aydin et al. have developed a metric which can generate a distortion map by comparing two images having different dynamic ranges [151]. The online metric requires one to provide both the HDR image and the LDR image obtained from the HDR image using a tone mapping operator. As LDR images are supposedly shown in an LDR display, we need to specify a typical display on which all the LDR images are going to be visualized. We assume that the LDR images are shown in a typical LCD display with maximum luminance 100 and gamma 2.2. We also assume that for all the LDR images, the viewing distance is 0.5 meters and the number of pixels per unit visual cone (in degree) is 30. We use default parameters in this metric and the significance of the choice of other display parameters can be found in [151].

The distortion maps show the probability of the occurrence of different distortions at every pixel location - the amplification of visible contrast (blue), loss of visible contrast (green) and reversal of visible contrast (red). This metric measures the distortions suffered at each pixel location when the dynamic range of an HDR image is compressed. Though this method is applicable for the tone mapping operators which obtain an LDR image from the corresponding HDR image, we use the same metric to obtain the distortion maps for the LDR images generated using the approaches presented in this thesis as well as exposure fusion. One would expect that
<table>
<thead>
<tr>
<th>Tone Mapping Operators</th>
<th>Lizard (Figure 6.2)</th>
<th>Desk (Figure 6.5)</th>
<th>Star (Figure 6.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast bilateral filtering [58]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatial kernel sigma</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Range kernel sigma</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Base contrast</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gradient domain [62]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Beta</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Color saturation</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Noise reduction</td>
<td>0</td>
<td>0</td>
<td>0.881</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Contrast domain [63]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contrast factor</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>Saturation factor</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Detail factor</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Photographic [55]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Key value</td>
<td>0.18</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Phi</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Range</td>
<td>8</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Lower scale</td>
<td>1</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Upper scale</td>
<td>43</td>
<td>43</td>
<td>-</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Photoreceptor [46]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brightness</td>
<td>10</td>
<td>10</td>
<td>2.39</td>
</tr>
<tr>
<td>Chromatic adaptation</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Light adaptation</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6.1: Values of parameters corresponding to different tone mapping operators
### 6.1 LDR Image Generation for Static Scenes

<table>
<thead>
<tr>
<th>Tone Mapping Operators</th>
<th>Lizard (Figure 6.2)</th>
<th>Desk (Figure 6.5)</th>
<th>Star (Figure 6.8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashikhmin’s [54]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local contrast Threshold</td>
<td>0.5</td>
<td>simple</td>
<td>0.38</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>1</td>
<td>2.061</td>
</tr>
<tr>
<td><strong>Adaptive logarithmic [45]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>1</td>
<td>0.727</td>
<td>1</td>
</tr>
<tr>
<td><strong>Time-dependent Adaptation [47]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplier</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Cones</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Rods</td>
<td>-</td>
<td>0.5</td>
<td>-</td>
</tr>
<tr>
<td>Pre Gamma</td>
<td>0.909</td>
<td>2.091</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6.2: Values of parameters corresponding to different tone mapping operators

This metric would lead to more distortions for the direct LDR image generation methods as they are not tone mapped from the reference HDR image with which the LDR is going to be compared. Nevertheless, we present the distortion maps for LDR images obtained using both the popular tone mapping methods as well as the direct LDR image generation methods in this chapter.

Consider the set of images termed as ‘lizard’ multi-exposure set shown in Figure 6.1. We shall now present the computational time required for the direct LDR image generation methods. For compositing 9 images shown in Figure 6.1 of size 2464 X 1632 using Matlab® in an Intel Xeon® machine with 4 GB RAM, variational compositing took more than 15 minutes, exposure fusion took 107 seconds, bilateral filter based compositing took 160 seconds, while the gradient domain compositing took 113 seconds. We can observe that two of the proposed solutions for this HDR imaging problem are as fast as the exposure fusion method.

Let us now discuss the scene which is captured using the 9 multi-exposure images shown in Figure 6.1. There is a lizard sitting on a rock which is located in a tree shade while the bushes behind the rock are sunlit. This is a typical scene which has a very high dynamic range where both brightly (bushes) and poorly (lizard and rock) illuminated regions are present in the same scene. As explained in Chapter 2, common digital cameras can not capture the entire scene information using a single snapshot due to the limited capacity of the sensor elements. This is
more evident while observing the images in Figure 6.1 where the images are arranged in the increasing order of exposure times. The images in Figure 6.1(a-d) have low exposure times and hence capture the details present in the brightly illuminated regions (bushes) of the scene. The images in Figure 6.1(e-g) have high exposure times and hence capture the details present in the poorly illuminated regions of the scene (lizard and rock). The images in Figure 6.1(h,i) are so highly exposed that all details are completely saturated.

The LDR image results of the approaches discussed in last three chapters along with that of the existing method (exposure fusion) for multi-exposure images of a static scene are shown in Figure 6.2(a-d). We also provide the LDR results of some of the popular tone mapping methods for comparison in Figure 6.2(e-l). On visual examination, we can observe that the tone mapped LDR image shown in Figure 6.2(f) yields the best result in terms of both brightness and contrast. This method is the gradient domain tone mapping operator proposed by Fattal et al. [62]. One of the proposed approaches, gradient domain solution, discussed in Chapter 5 yields the best LDR image after this approach. This can be observed in Figure 6.2(d). The adaptive logarithmic tone mapping operator is also observed to produce an equally good result as shown in Figure 6.2(k) [45].

The other methods do not either preserve brightness in both brightly and poorly illuminated regions or suffer from reduction in contrast. A typical example of the under-saturation of the poorly illuminated regions (lizard and rock) is shown in Figure 6.2(l) where time dependent visual adaptation tone mapping operator is used [47]. An approach which leads to a loss of contrast for this scene is the exposure fusion approach as can be seen in Figure 6.2(b) [128]. These are the inferences one can arrive at through visual examination.

Let us now evaluate the LDR image results shown in Figure 6.2 against the corresponding HDR image generated using Adobe® Photoshop CS4. We employ the online metric called dynamic range independent quality metric by Aydin et al. [151]. The results of the quality metric are presented in Figure 6.3. The smaller number of pixels marked red, green, or blue, the better the LDR image is. According to this metric, adaptive logarithmic tone mapping (Figure 6.3(k)) [45], exposure fusion (Figure 6.3(b)) [128], and one of the proposed solutions based on bilateral filter (Figure 6.3(c)) yield the least distortion. The methods with the most distortion of the LDR images are the Ashikhmin’s tone mapping operator (Figure 6.3(j)) [54], time dependent visual adaptation operator (Figure 6.3(l)) [47], and the proposed gradient domain solution (Figure 6.3(d)).
The observations based on the results of the quality metric do not agree with the best LDR images picked through visual examination. We had seen that the LDR image obtained through the proposed gradient domain solution preserved both brightness and contrast on brightly and poorly illuminated regions of the scene. Further, exposure fusion method is found to produce the LDR image with less distortion according to this metric. However, the image in Figure 6.2(b) does not look better than some of the LDR images obtained using other methods.

Figure 6.1: (a-i) Multi-exposure images of a static scene. Images Courtesy: Erik Reinhard, University of Bristol.

Consider the set of images labeled ‘desk’ obtained using the auto-exposure bracketing (AEB) feature of the digital camera shown in Figure 6.4. In this scene, the lamp and the jar on which it is located are brightly illuminated. Further, part of the books which are placed proximal to the lamp is brightly illuminated while the other part of the books is poorly illuminated. This scene does not have that high dynamic range as the scene shown in Figure 6.1. Hence, only 3
Figure 6.2: LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
Figure 6.3: Computed distortion maps for LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
images captured with different exposure times are sufficient to capture the entire dynamic range of the scene as shown in Figure 6.4. The brightly illuminated regions of the scene are captured well in the images in Figure 6.4(a) and Figure 6.4(b) while the poorly illuminated regions are captured well in Figure 6.4(c).

We shall now discuss the LDR images obtained using direct methods and using tone mapping methods from the HDR image shown in Figure 6.5. We can observe that Ashikhmin’s tone mapping operator yields the best result (Figure 6.5(j)) [54]. This method preserves contrast in both the brightly and poorly illuminated regions of the scene. Among the direct LDR image generation methods, variational solution presented in Chapter 3 yields the best result (Figure 6.5(a)) closely followed by exposure fusion (Figure 6.5(b)) [128]. The gradient domain tone mapping [62] yields the LDR image which is under-saturated as can be seen in Figure 6.5(f).

Let us now discuss the distortion maps obtained for these LDR images using the dynamic range independent quality metric [151] shown in Figure 6.6. Ashikhmin’s tone mapping operator (Figure 6.6(j)) [54] and the exposure fusion method (Figure 6.6(b)) [128] yield the least distortion among the LDR images shown in Figure 6.5. This is in complete agreement to the visual examination of the images in Figure 6.5. Also, gradient domain tone mapping method [62] (Figure 6.6(f)) shows more distortion compared to the other methods which also agrees with the visual examination. The proposed approaches such as variational solution (Figure 6.6(a)) and bilateral filter based solution (Figure 6.6(c)) are found to have lesser distortion compared to most of the tone mapping methods.

Consider the ‘star’ set of multi-exposure images shown in Figure 6.7. These images are captured in a Starbucks shop somewhere in San Antonio. The dynamic range of the scene is not that large as the sunlit scene and hence 3 images are sufficient to capture the entire dynamic range of the scene as shown in Figure 6.7. In this scene, the lamp and the adjacent regions such as paintings mounted on the wall and the boxes on the cupboard are brightly illuminated. The chairs located on the floor which are far away from the lamp are poorly illuminated. The images are arranged in the increasing order of exposure times. The lamp and the adjoining regions are better captured in Figure 6.7(a) while the regions far away from the lamp are better captured in Figure 6.7(b) and Figure 6.7(c).

The LDR images generated using the proposed approaches along with that of other tone mapping methods and exposure fusion method are presented in Figure 6.8. It can be seen that the best LDR image result is the one generated using the proposed solution based on bilateral
Figure 6.4: (a-c) Multi-exposure images of a static scene. Images Courtesy: Sam Hasinoff, MIT CSAIL.
Figure 6.5: LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
Figure 6.6: Computed distortion maps for LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
filter (Figure 6.8(c)). For this scene, all the direct LDR image generation methods yield better results compared to that of the tone mapping operators as can be observed in Figure 6.8(a-d). The bilateral filter based tone mapping operator produces an over-saturated LDR image irrespective of the values for different parameters (Figure 6.8(e)) [58]. The best LDR result among the tone mapping operators is obtained while using the gradient domain tone mapping operator (Figure 6.8(f)) [62].

The distortion maps obtained for the LDR images in Figure 6.8 with respect to the corresponding HDR image are presented in Figure 6.9. The gradient domain tone mapping operator yields the LDR image with the least distortion (Figure 6.9(f)) [62]. Except the proposed gradient domain solution, other direct LDR image generation methods are found to have lesser distortion (Figure 6.8(a-d)). As expected, the bilateral filter based tone mapping operator shows maximum distortion as the LDR image in this case was completely over-saturated (Figure 6.9(e)) [58].

![Multi-exposure images of a static scene](image)

Figure 6.7: (a-c) Multi-exposure images of a static scene. Images Courtesy: Gopinath Sivanesan, San Antonio.

Consider a scene with a very high dynamic range shown in Figure 6.10. In this scene,
Figure 6.8: LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
Figure 6.9: Computed distortion maps for LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, (d) Gradient domain compositing, (e) Fast bilateral filtering tone mapping operator [58], (f) Gradient domain tone mapping operator [62], (g) Contrast domain tone mapping operator [63], (h) Photographic tone mapping operator [55], (i) Photoreceptor tone mapping operator [46], (j) Ashikhmin’s tone mapping operator [54], (k) Adaptive logarithmic tone mapping operator [45], and (l) Time-dependent visual adaptation tone mapping operator [47].
the interior of the room is very poorly illuminated while the trees visible through the window and the door are very brightly illuminated by the sun. Such scenes are quite challenging as the dynamic range of the entire scene is very large. For this ‘house’ set of multi-exposure images, we do not have the knowledge of exposure times. The HDR image can not hence be reliably estimated for such a scene because of its enormous dynamic range. We will now apply the direct methods of LDR image generation and compare the performance of these methods for this scene. The CRF estimated for such scenes without exposure times can be erroneous. Therefore, we do not assume the presence of accurate HDR image corresponding to this scene and we will not use the dynamic range independent distortion measure for evaluation as earlier.

The LDR images generated using different approaches are shown in Figure 6.11. For this set of images, exposure fusion method (Figure 6.11(b)) and the proposed gradient domain solution (Figure 6.11(d)) give the best results. The variational solution yields an LDR image (Figure 6.11(a)) which is over-saturated in the brightly illuminated regions which is due to the inability of this approach to handle highly over-exposed regions in Figure 6.10(d). The bilateral filter based solution provides an LDR image with better contrast compared to the other methods and is slightly under-saturated in the poorly illuminated regions (Figure 6.11(c)).

The variational approach fails to generate an artifact-free LDR image as the details in the brightly illuminated regions are completely washed out in Figure 6.10(d). In these regions, the iterative solution is not able to transfer details from other images in order to preserve contrast. The edge-preserving bilateral filter solution mainly relies on preserving small textures lost due to the limited dynamic range of the imaging device. In Fig 6.10, all the images in the multi-exposure has low exposure in the poorly illuminated regions except the last image (Figure 6.10(d)). Since, there are not enough images which are properly exposed in this poorly illuminated region, the solution is slightly under-exposed.

### 6.2 Conclusion

We presented the results of different methods we developed in the last three chapters for generating an LDR image corresponding to a static scene. We compared the results of these methods with that of different tone mapping operators. We showed that the proposed solutions for the LDR image generation in the previous chapters perform as well as the most tone mapping operators even without any CRF information. We showed that the direct LDR image generation
(a) (b)

(c) (d)

Figure 6.10: (a-d) Multi-exposure images of a static scene. Images Courtesy: Min H. Kim.
Figure 6.11: LDR results of a static scene by (a) Variational compositing, (b) Exposure fusion [128], (c) Bilateral filter based compositing, and (d) Gradient domain compositing.
methods to be as effective in reproducing all the details of the scene as the tone mapping operators. The primary advantage of the proposed solutions for generating an LDR image is that we have not assumed any knowledge of the CRF, exposure times, and the presence of HDR image. We would now like to develop an approach for the more complex case of dynamic scenes. We shall look at this problem along with the details of the proposed solution for dynamic scenes in the next chapter.
Chapter 7

Image Compositing for Dynamic Scenes

7.1 Introduction

Most real world scenes are dynamic. While capturing multiple images of a scene, one does not have much control over the movement of objects in the scene. If the changes in the scene are not detected before compositing multi-exposure images, the generated LDR image would have artifacts called ghosts. It is imperative that we need to detect any scene changes across these multi-exposure images to prevent ghosts from appearing in the generated LDR image. In this work, we address the problem of generating a high contrast LDR image of a dynamic scene directly from a set of multi-exposure images. Our contributions are to develop a robust algorithm for detecting scene changes and to compose different regions seamlessly. Additionally, we show that this can be achieved in the absence of any knowledge of both the CRF and the camera parameter settings. Further, we do not need to manually tune various parameters based on heuristics, as is common with existing techniques.

We develop a novel bottom-up segmentation algorithm based on superpixel grouping to segment out the scene changes [152]. A characteristic function between a given pair of observations with different exposures enables us to identify decision regions for grouping the superpixels which belong to the foreground (scene changes). After detecting the regions of the image which show change with respect to a reference image, we compose the multi-exposure images to generate the LDR image without any ghosts. The primary advantage of our approach is that we do not assume any knowledge of the scene and camera settings. Further, we show that seamless LDR image can be generated even when there is an appreciable scene change across the multi-exposure images.
To start with, we shall provide an overview of the segmentation techniques. We shall then look at the proposed approach for generating an LDR image from a set of multi-exposure images of a dynamic scene in detail. We shall then present the results of the proposed approach for a few dynamic scenes. We show that high quality LDR images can be generated using the proposed approach. We compare the LDR image results of the proposed approach with that of state-of-the-art approaches by Gallo et al. [33] and Adobe® Photoshop CS5. The generated LDR images can be made compatible with HDR displays using inverse tone mapping [68].

### 7.2 Related Work

In the case of static scenes, capturing multi-exposure images with a hand-held camera can lead to registration mismatch. The images can be registered using bitmaps computed on image pyramids [28]. However while capturing multi-exposure images of a scene, we cannot guarantee that the scene would not change. There are chances of new objects being introduced in the scene between the exposures due to motion. Also, objects such as leaves and branches of a tree would move when there is a presence of wind in the scene. In other words, the scene would most probably be dynamic. When the methods mentioned above are employed for compositing multi-exposure images of a dynamic scene, the objects in motion in the scene would give rise to artifacts called ghosts. It is required that the changes in the scene are detected apriori before compositing is performed on the multi-exposure images. We shall first look at the methods previously used for removing ghosting artifacts.

The change detection across the multiple differently exposed images can be merged with the HDR image generation process and appropriately lesser weights can be given to the pixel locations of an image if it is found to have scene change [32]. This technique helps one to eliminate ghosts to certain extent. This approach suffers from the fact that even pixel locations which have scene change are given some weight which may lead to artifacts which are evident while looking closely. This approach also assumes the knowledge of CRF to generate the HDR image.

Jacobs et al. proposed a method to identify the regions on the image grid which change across multi-exposure images using weighted variance and entropy measures [31]. This method fills the motion regions using details from one of the observations and thereby reducing contrast in such regions. Gallo et al. proposed a method to detect motion regions in multi-exposure
images when CRF is known and eliminate them while compositing [33]. This approach preserves contrast in the motion regions as regions from multiple images are combined. However, estimation of CRF of the imaging system required by this approach is a challenge and any error in estimation will lead to poor detection of dynamic regions of the scene. Another approach for eliminating ghosting artifacts while creating mosaic from images having possible exposure change was proposed in [30]. However, the emphasis of the work has been primarily on seam removal and not on HDR imaging.

Segmenting foreground from background in a single image is a classic vision problem. The segmentation algorithm can either be automatic or be interactive. One popular approach to achieve interactive segmentation is the Grabcut by Rother et al. [153]. Interactive segmentation depends on the user input such as a bounding box or scribbles to perform segmentation. Another class of interactive algorithms which extract the alpha matte along with the foreground mask is known as matting. An example is the natural image matting [154]. We focus more on the automatic segmentation approaches in this work as we intend to develop an automatic method to compensate for scene change across different exposures.

Automatic segmentation approaches can be broadly classified into top-down and bottom-up methods. In the top-down approach, one tries to capture the entire object boundary directly by the learned features of the desired object class ([155], [156]). This approach tries to get the foreground separated from the background as a whole. While in the bottom-up approach, the entire image is split into homogeneous regions based on color, contours, and texture details. These homogeneous regions are then grouped to segment the foreground from the background [157]. The bottom-up segmentation methods have derived much interest within the computer vision community of late as they lead to better segmentation of the foreground objects.

7.3 Bottom-up Segmentation

In this chapter, we specifically focus on bottom-up segmentation. The algorithm by Shi and Malik uses normalized cuts to split a given image into multiple regions which are homogeneous [157]. This approach was later extended to define different homogeneous regions of the image as superpixels and they are grouped for segmentation [152]. Each superpixel is a collection of a set of pixels inside a closed contour signifying uniformity in terms of color, intensity and texture. Object recognition systems can then work on the level of superpixels instead of image
pixels which can help in designing faster algorithms. For segmentation tasks, one can group the superpixels belonging to foreground object based on some criteria to segment out the foreground ([158], [159]). Even neighborhood can be defined for superpixels to improve the segmentation [160].

A typical bottom-up approach for segmentation relies on the efficient grouping of the superpixels and recovering the foreground. Apart from basic segmentation, grouping of superpixels have a number of applications in computer vision. Consider the problem of estimating the depth map of a scene from a single image. Superpixels corresponding to objects at different depths of the scene can be grouped to recover the 3D information from a single image [161].

In the present work, we apply superpixel grouping for detecting scene change in the multi-exposure images of a dynamic scene. We assume one of the multi-exposure images as the reference and employ grouping of superpixels to recover the scene change in the other multi-exposure images. The key challenge here is the identification of appropriate decision regions to group the superpixels. Given a set of $K$ multi-exposure images corresponding to a dynamic scene, we need to define $K-1$ different decision regions with respect to the reference image. We shall explain the steps involved in the modeling of decision regions, the grouping of superpixels and LDR image compositing in more detail in the next section.

### 7.4 Proposed Approach

We shall revisit the basic image formation equation discussed in chapter 2. The 2-D image formation can be described by the Equation 7.1.

$$I(x, y) = f\left(\Delta t E(x, y)\right) \quad (7.1)$$

where $I(x, y)$ is the intensity value of the image, $\Delta t$ is the exposure time, and $E(x, y)$ is the image irradiance. The image irradiance $E(x, y)$ and the image intensity $I(x, y)$ are related by a non-linear function called camera response function (CRF) $f$. The term $(\Delta t E(x, y))$ is generally called the exposure. Generally, CRF is estimated from a set of multi-exposure images and plotted as the logarithm of exposure vs the image intensity.

While capturing multi-exposure images corresponding to a dynamic scene, we can rewrite Equation 7.1 as shown in Equation 7.2.

$$I_k(x, y) = f\left(\Delta t_k E_k(x, y)\right) \quad (7.2)$$
where \( I_k(x, y) \) represent the intensity values of the \( k^{th} \) image in the exposure stack with exposure time \( \Delta t_k \). Here \( E_k \) is the irradiance of the scene corresponding to the \( k^{th} \) exposure as the scene changes across the images. Given a set of observations \( I_k(x, y) \) corresponding to a dynamic scene, we would want to generate an artifact-free but high contrast LDR image of the scene when the CRF \( f \) and exposure times \( \Delta t_k \) are not known. Further we have the issue of irradiance changing as well.

We assume that we do not have the knowledge of CRF of the camera and the exposure settings. Given a set of multi-exposure images, our task is to identify the regions which have moving objects in each of the images and eliminate them while compositing. The salient feature of the proposed approach is that we composite patches from multiple images even in regions which show scene change, thereby preserving overall contrast of the scene in the generated LDR image. We shall first estimate the decision boundaries to classify dynamic and static regions and use that to reconstruct the ghost-free LDR image of the dynamic scene ([162], [163], [164]).

### 7.4.1 Estimation of the Decision Regions

Consider any two observations of a scene \( I_1(x, y) \) and \( I_2(x, y) \) of a static scene which differ only in the exposure times. These two images can be related by the linear equation in the log domain shown in Equation 7.3.

\[
\log f^{-1}(I_2(x, y)) = \log f^{-1}(I_1(x, y)) + \log \left( \frac{\Delta t_2}{\Delta t_1} \right) \quad (7.3)
\]

This equation shows us that the knowledge of CRF \( f \) and hence its inverse will prove vital in the estimation of dynamic regions of the scene. This fact was employed by Gallo et al. to generate HDR image corresponding to a dynamic scene without any ghosting artifacts [33]. In this work we do not assume the knowledge of CRF \( f \). We shall now discuss how dynamic regions can be determined in each of the multi-exposure images with respect to a reference image without the knowledge of CRF \( f \) and the exposure times.

The intensity values of these two images can also be related by Equation 7.4.

\[
I_2(x, y) = f \left( f^{-1}(I_1(x, y)) \left( \frac{\Delta t_2}{\Delta t_1} \right) \right) \quad (7.4)
\]

which is of the form

\[
I_2(x, y) = u_{2,1}(I_1(x, y)) \quad (7.5)
\]
The intensity values of these two images can be characterized by a function called comparametric function $u_{2,1}$ [21]. The comparametric function is also referred to as intensity mapping function (IMF) and is estimated from the histograms when there is minimal scene change between the two images [20]. We use the term IMF to refer to this function henceforth. This function defines how the intensity values of two images of a static scene should relate when there is only a difference of exposure times. IMF is a non-linear function whose slope can be computed and this would be greater than 1 if the exposure time of one image is greater than that of the reference image. This function can be accurately estimated when the scene is static as intensity values in any pixel locations can be used to estimate the IMF. However in the case of a dynamic scene, we need to have a different technique to estimate IMF.

Consider a set of multi-exposure images corresponding to a dynamic scene shown in Figure 7.1. These images are taken at different times of the day with different exposure settings. These images together are sufficient to recover the entire dynamic range of the scene. However the scene changes appreciably in the 7.1(b) and 7.1(d) due to the movement of people in the scene. When CRF is known, we can recover the HDR equivalent of the scene using the technique mentioned in [33]. In the absence of an accurate estimate of the CRF corresponding to the camera used, we need to figure out pixel locations which do not change in any of the multi-exposure images. The intensity values of the multi-exposure images corresponding to these pixel locations are used to estimate the IMF between a pair of images from the exposure stack in Figure 7.1.

It is worth noting that the normalized intensity values of the multi-exposure images are in the range $[0, 1]$. The weighted variance measure $V(x, y)$ can be computed for $K$ differently exposed images $I_i(x, y)$ using the Equation 7.6.

$$V(x, y) = \frac{\sum_{k=1}^{K} \zeta_k(x, y)I_k^2(x, y)/\bar{\zeta}(x, y)}{\left(\sum_{k=1}^{K} \zeta_k(x, y)I_k(x, y)\right)^2 / (\bar{\zeta}(x, y))^2} - 1 \quad (7.6)$$

where $\bar{\zeta}(x, y) = \sum_{k=1}^{K} \zeta_k(x, y)$ and the weight is given by the Gaussian function

$$\zeta_k(x, y) = e^{- \frac{(I_k(x, y) - 0.5)^2}{2 \times 0.2^2}} \quad (7.7)$$

As explained in chapter 5, this Gaussian function $\zeta_i(x, y)$ is used as weight in order to provide lesser weight to the over-exposed and the under-exposed pixel locations. We use an appropriate
Figure 7.1: (a-e) Multi-Exposure images of a dynamic scene. Images Courtesy: Orazio Gallo, UCSC.
threshold (0.25 times the maximum weighted variance) to detect pixel locations which show low weighted variance measures. In the case of noisy multi-exposure images, simple Gaussian spatial smoothing or anisotropic diffusion can be used prior to the computation of weighted variance [109]. The weighted variance measure has already been used to fill intensity values from one of the images in the exposure stack in the dynamic regions ([4], [31]). However, such an approach reduces contrast in regions where there is scene change in any of the multi-exposure images.

The weighted variance measure provides us the pixel locations of the scene where there are no appreciable changes in any of the multi-exposure images. We assume that we have enough pixel locations on the image grid where there is no scene change in any of the images. This assumption is true for a general natural scene and may not apply to complex scenes such as those involving crowd motion in most of the pixel locations. However, such occurrences are a rarity while processing natural scenes and we employ the weighted variance measure as a tool to determine the pixel locations from where we would get data points to estimate the IMFs.

Now, we would be able to estimate a unique IMF for a given pair of images using the intensity values at these pixel locations. Let \( S \subseteq \mathbb{R}^2 \) be the set of all pixel locations in the image grid. Now, the pixel locations where there are no motion in any of the images is given by the set \( \psi \subseteq S \). As a generality, we select one of the multi-exposure images as representing the static scene. We would now estimate IMFs between the intensity values of this image and the rest of the images in \( \psi \). Given a set of \( K \) multi-exposure images, we would have a total of \((K - 1)\) IMFs with respect to the reference image.

We need to fit a polynomial of order four (chosen empirically) in order to estimate IMF. We employ a polynomial of order four which is used in the previous work of modeling IMF and is reported to fit the function accurately [20]. The pixel locations of the multi-exposure images in \( \psi \) should follow this IMF with respect to the reference image in order to get classified as static. The pixel locations which have some appreciable scene change from the reference would not follow this IMF. The estimated IMF between the images Figure 7.1(b) and Figure 7.1(e) (reference) is shown in Figure 7.2(a).

These IMFs, thus estimated, are now employed to find the decision regions to perform bottom-up segmentation. Having estimated the IMFs for each of the multi-exposure images with respect to the reference image, we need to define a constant width region around the IMF. This constant width region would represent the pixel locations of the test image which do not
Figure 7.2: (a) The IMF between a pair of images in Figure 7.1(b) and Figure 7.1(e) (reference), and (b) The constant width region which defines the decision boundary for static and dynamic regions.

have any appreciable change with respect to the reference as shown in Figure 7.2(b). This constant width region provides us with a decision boundary between static and dynamic regions of the given multi-exposure image with respect to the reference. We exploit this constant width region for decision regions as this provides us a means to deal with noisy images. The width of the region around the IMFs can be increased in the case of multi-exposure images with more noise.

### 7.4.2 Superpixel Grouping

We now propagate the decisions from pixel level to region level. We exploit over-segmentation of images using superpixels to recover regions of images which have homogeneous color and texture. We compute superpixels on all multi-exposure images excluding the reference image [152]. Alternately, one can use a fast algorithm to speed up the over-segmentation process ([165], [166]). We use superpixels as it saves us from the huge load of classifying each pixel and grouping them. Further, it allows us to take care of the object boundaries while segmentation which is not possible when patches are used. The superpixels corresponding to the images shown in Figure 7.1 are shown in Figure 7.3. As can be observed in Figure 7.3(e), the superpixels do not cross the object boundaries and grouping them would enable us to recover the exact
Figure 7.3: (a-d) Superpixels estimated corresponding to the first four multi-exposure images in Figure 1, and (e) Magnified superpixels corresponding to a region of (d).
silhouette of the scene change. Instead of classifying every pixel, we classify the superpixels for the possible scene change with respect to the reference (Figure 7.1(e)).

Given the reference image and any other multi-exposure image, we now find out the fraction of pixels in a given superpixel which lie inside the constant width region ($0.12$ in this work). We define a parameter $\gamma$ which defines the minimum fraction of pixels which should be present inside the constant width region for the superpixel to be classified as having no change. We used $\gamma$ to be equal to 0.9 in our experiments. We classify all the superpixels corresponding to the given image with respect to the reference as either dynamic or static.

This operation would effectively let us group all the superpixels of the given image which show change with respect to the reference image unlike other methods based on square patches [33]. This would enable us to recover exact boundary of the object present in the dynamic region of the scene. This is the novel bottom-up segmentation algorithm we have developed for multi-exposure images using the estimated IMF for the computation of decision boundaries. The bottom-up segmentation which has been performed on the multi-exposure images is shown in Figure 7.4. One can clearly see that the proposed algorithm for segmentation is able to group all the superpixels which convey an appreciable scene change. We need to ignore these regions while compositing multi-exposure images in order to avoid ghosting artifacts.

### 7.4.3 Piece-wise Rectangular Approximation

Having detected the scene changes with respect to the reference image, we need to develop a method to reconstruct the final LDR image of the scene. The LDR image needs to be generated from the reference image and the superpixels from other images marked as having no scene change (static). The segmented images shown in Figure 7.4 present a new challenge while compositing the final LDR image. As one can observe from Figure 7.4(d), the segmented regions can be very irregular which leads to problems when we want the LDR image to be generated without any visible seams. Also, one cannot guarantee that these grouped regions would be closed as evident in Figure 7.4(d). Possible solution can be the use of gradient domain processing with Dirichlet boundary conditions and employ Poisson solver to reconstruct the LDR image. However, we do not have the actual values of the LDR image on the segmentation boundaries which makes this approach infeasible. We need to adopt a different strategy to handle these irregular boundaries.

We shall now discuss the piece-wise approximation approach to handle these irregular
Figure 7.4: (a-d) Bottom-up segmentation through superpixel grouping performed on the images in Figure 7.3.
Figure 7.5: (a-d) Piece-wise rectangular approximation of the segmentation boundaries shown in Figure 7.4.
boundaries. We split the images into overlapping patches of certain size (say 6 X 6) with an overlap of one pixel in each direction. We detect patches of the images which have more than 90 percent of the pixels lying inside the static superpixels. These patches are classified as not having appreciable motion. Such an operation would result in piece-wise rectangular approximation of the bottom-up segmentation boundaries as visible in Figure 7.5. As one can visualize from Figure 7.5(d), we have the freedom to choose the size of the patches which approximate the segmentation boundary. If faster compositing is desired, one can use larger patches to approximate the segmentation boundaries.

We can now use any of the static multi-exposure compositing algorithms on the patches marked as static regions ([167], [95], [129]). But they tend to generate seams in the composited image when unequal number of images being stitched at different patch locations. We shall use gradient domain compositing on those patches which are labeled as static which was discussed in detail in chapter 5. The gradient domain compositing helps us to prevent seams from appearing in the final composite image apart from enabling us to perform faster compositing compared to the existing methods. To avoid these seams across patch boundaries while combining different number of images at each patch location, we use overlapping patches (of size 6 X 6) with an overlap of one pixel in each direction.

7.4.4 Poisson Seam Correction

Had we used non-overlapping patches, there would not have been any information passage between adjacent patches. This would have resulted in visible seams on the patch boundaries. The use of overlapping patches for approximating the segmentation boundaries enables us to avoid these visible seams on the patch boundaries and reduce artifacts in the LDR image. The seams across the patch boundaries which are avoided by the use of overlapping patches is called Poisson seam correction. We shall now use the gradient domain solution obtained in Chapter 5 to operate on the patches which do not have scene change. We will get a composited gradient patch at each of the patch location. These gradients corresponding to patch locations are now used for the reconstruction of the desired high contrast LDR image.

The gradients corresponding to the composited patches (of size 5 X 5, one less than the size of overlapping patches) are arranged on the image grid $S$. The resultant vector field may not be conservative. We employ a direct Poisson solver with Neumann boundary conditions to generate the scalar field closest to the vector field ([135], [131]). The scalar field obtained
through Poisson seam correction operation is the desired LDR image. The LDR image will not have any ghosting artifacts as we have eliminated the patches of the multi-exposure images which correspond to the dynamic regions of the scene. The complete schematic representation of the steps involved for dynamic scenes is shown in Figure 7.6.

![Schematic representation of the proposed approach.](image)

Figure 7.6: Schematic representation of the proposed approach.

### 7.5 Results

In this section, we consider multi-exposure images of a dynamic scene. We shall present the LDR images generated using the proposed approach and compare the results with that of the tone-mapped LDR image obtained using the method of Gallo et al. [33] and “Merge to HDR pro” tool in Adobe® Photoshop CS5 [75]. Gallo et al. employ an interactive tone mapping method proposed by Lischinski et al. [64]. It is worth noting again that the proposed approach does not require the knowledge of CRF and the exposure settings. Further, we do not explicitly generate the HDR image of the scene and therefore do not require tone reproduction. These are the key advantages of the proposed approach over that of Gallo et al. [33] and Adobe® Photoshop CS5 [75]. Both these methods require the knowledge of exposure settings in order to generate the HDR image and then the LDR image. In case exposure settings of the multi-
Figure 7.7: (a-e) Multi-Exposure images of a dynamic scene. Images Courtesy: Orazio Gallo, UCSC.
exposure images are unavailable, the results would be erroneous.

We will not use the dynamic range independent metric by Aydin et al. for producing distortion maps for the LDR images in the case of dynamic scenes [151]. This metric requires a reference HDR image to be used as the exact replica of the dynamic scene captured. The quality of the HDR image corresponding to a dynamic scene depends on the scene change detection algorithm used. We do not yet have a method which produces completely artifact-free HDR image in the case of dynamic scenes. We, therefore, employ visual examination to compare LDR images obtained using different approaches. The slight loss in contrast in the results while using the proposed approach may be due to the fact that we are performing all the processing in the LDR domain without linearizing the intensity values of the images. Also, the difference in color tone between the results is mainly due to the type of tone mapping employed by Gallo et al. [33] and Adobe® Photoshop CS5 [75].

Consider the multi-exposure images of a dynamic scene in Figure 7.1. Figure 7.1(e) is picked as the reference image. Figure 7.9(a) shows the result of an existing approach [129]. As scene change is not accounted for, ghosting artifacts are clearly visible in the generated LDR image. The tone mapped image corresponding to the HDR image generated by Gallo et al. is as shown in Figure 7.9(b). It can be seen that they are some artifacts near the bottom of the pillar and the floor. “Merge to HDR Pro” tool in Adobe® Photoshop CS5 yielded the LDR image shown in Figure 7.9(c). One can see loss of contrast details on the floor which is over-saturated. The LDR image generated using the proposed approach is shown in Figure 7.9(d). One can observe that the proposed approach is able to generate an artifact-free LDR image from a set of multi-exposure images. The details in both the brightly and poorly illuminated regions are clearly visible and there are no artifacts.

Consider another set of multi-exposure images of a dynamic scene as shown in Figure 7.7. This scene is complex in the sense that there are many people moving in and out of the scene across these images. Further there is a sun-lit region (brightly illuminated) and tree shade (poorly illuminated). As expected, common digital cameras cannot capture the entire dynamic range of this scene with varied levels of brightness. We assume the image in Figure 7.7(c) to be the reference image. We detect scene changes on the other images with respect to this image. Figure 7.10(a) shows the tone mapped LDR image of Gallo et al. [33]. Though it has higher contrast, this image shows some blue colored artifacts between the walking people. “Merge to HDR Pro” tool in Adobe® Photoshop CS5 provided the LDR image shown in Figure 7.10(b).
One can observe that the overall contrast is far better than the method by Gallo et al. [33]. However, the ghosting artifacts are more visible and the LDR image is not visually pleasing. Compositing and Poisson seam correction using the proposed approach enabled us to generate the LDR image shown in Figure 7.10(c).

Let us consider another dynamic scene captured using differently exposed images shown in Figure 7.8. This scene has the branches of the tree moving and there is a person introduced in the scene while capturing the last image (Figure 7.8(d)). This is an example of a scene where there is a significant motion in the majority of the pixel locations. We pick the image Figure 7.8(b) as the reference image. The tone mapped result of Gallo et al. [33] is shown in Figure 7.11(a). Figure 7.11(b) shows the LDR image generated using “Merge to HDR Pro” tool in Adobe® Photoshop CS5. For this scene, the LDR image generated using this tool has very good contrast. One can observe that there are tiny ghosting artifacts in the place where the person was located in Figure 7.8(d). The reconstructed LDR image using the proposed approach is shown in Figure 7.11(c). We can see that the proposed approach is able to reconstruct the scene without any artifacts even in the case of significant motion of tiny objects across the multi-exposure images. The contrast of the generated LDR image is slightly less compared to the other two methods.

7.6 Conclusions

We have proposed a novel bottom-up motion segmentation approach for detecting motion in multi-exposure images corresponding to a dynamic scene. We then approximated the segmentation boundaries and reconstructed the dynamic scene with an LDR image without any artifacts. The proposed approach is a quite useful tool in digital photography where photographers like to use multiple differently exposed images to capture a dynamic natural scene. The proposed approach has added advantages of not requiring both the CRF and the tone mapping operation. Further, exposure settings of the multi-exposure images are also not needed. The generated high contrast LDR image is compatible with common displays and occupies lesser memory compared to the corresponding HDR image.

The proposed approach can either be included in the digital camera firmware or common image manipulation tools like Adobe® Photoshop. The proposed approach would be a worthy alternative to the “Merge to HDR” tool available in latest Photoshop releases (CS2 onwards).
7.6 Conclusions

Figure 7.8: (a-d) Multi-Exposure images of a dynamic scene. Images Courtesy: Orazio Gallo, UCSC.
The generated LDR image can be made compatible with HDR displays by using any of the inverse tone mapping algorithms. In this chapter, we explained how the algorithms for static scenes can be extended to handle dynamic scenes through bottom-up segmentation approach which does not require any additional information along with the multi-exposure images. In the next chapter, we shall present the conclusions of this thesis along with few pointers to improve the approaches discussed in this thesis.
Figure 7.9: (a) LDR image generated by bilateral filter based solution without motion detection showing ghosts, (b) Tone mapped LDR image using [33], (c) LDR image obtained using Adobe® Photoshop CS5, and (d) LDR image generated using the proposed approach.
Figure 7.10: (a) Tone mapped LDR image using [33], (b) LDR image obtained using Adobe® Photoshop CS5, and (c) LDR image generated using the proposed approach.
Figure 7.11: (a) Tone mapped LDR image using [33], (d) LDR image obtained using Adobe® Photoshop CS5, and (c) LDR image generated using the proposed approach.
Chapter 8

Conclusions and Future Directions

8.1 Conclusions

We have addressed the HDR imaging problem for static and dynamic scenes. We aimed to directly generate high contrast LDR images corresponding to these scenes. This is a highly challenging problem as it is ill-posed. We have showed that high contrast LDR images can be generated directly from a set of multi-exposure images without the knowledge of CRF and exposure settings. We further showed how an artifact-free LDR image can be generated corresponding to a dynamic scene. The primary advantage of our approaches for both static and dynamic scenes is that we do not go through the conventional step of HDR image generation and tone mapping. This enables us to minimize computational and memory requirements. The generated LDR image is compatible with common displays and printers. Further, the LDR image can be made compatible with HDR displays using any of the existing inverse tone mapping operators.

As we have defined our weighting function in such a way as to minimize the over- and under- saturation and also to enhance local contrast at a given pixel location, the proposed approaches can be extended to a wide range of other computational photography applications. The approaches discussed in this thesis are computationally fast and can be implemented in real time using fast GPUs. The proposed approaches free us from making significant changes in the sensor to capture the entire dynamic range of the scene. Common digital cameras can be used to capture scenes with high dynamic range using the proposed LDR solutions.

The primary contributions of this thesis are as follows. We have proposed three algorithms for generating high contrast LDR image from a set of multi-exposure images corresponding to a
static scene. These algorithms are developed using mathematical tools such as calculus of variations, edge-preserving filtering, and gradient domain processing. We showed that artifact-free LDR images can be generated even in the absence of the knowledge of CRF and the exposure settings. We extended the approach for static scenes to the more complex case of dynamic scenes. We have proposed a bottom-up segmentation algorithm followed by gradient domain processing in order to generate a high contrast LDR image without any ghosting artifacts.

The key contribution of the algorithms developed in this thesis is that we do not assume any knowledge of the CRF, exposure times and the scene information. The salient feature of the proposed algorithms is that we are able to generate high contrast LDR image of the scene without any artifacts directly from a set of multiple differently exposed images. The proposed framework shown in Figure 7.6 can handle even the complex case of dynamic scene where there is appreciable scene changes in the multi-exposure images. The generated LDR image can readily be visualized using common devices such as digital displays and printers as most of such devices can handle only LDR images. The overhead of HDR image generation and tone mapping is not present while using the proposed approaches for the generation of LDR image corresponding to a scene. We have compared the LDR images generated using the proposed approaches with the other recent approaches and have shown that the proposed approaches are able to generate high quality LDR images while not requiring exposure settings and other camera information.

We have further explained applications such as pin-hole image generation, random texture generation, and illumination compositing using the algorithms for compositing static scenes which we have developed in this thesis. Though the algorithms were not developed keeping these applications in the mind, we showed that appreciably good results can be obtained using the proposed algorithms. The proposed approaches can also be extended to a host of other applications in computational photography which require one to composite multiple images while preserving contrast.

8.2 Future Directions

Generating an LDR image from a set of multi-exposure images corresponding to a dynamic scene is a challenging task. There is still not a single algorithm which can solve this problem completely as the real world scenes have very high dynamic range. The techniques developed
8.2 Future Directions

in this thesis make a strong step towards identifying an unique solution for the HDR imaging problem by relaxing the need for the estimation of CRF, thereby enabling one to directly generate an LDR image from a set of multi-exposure images. We shall now look at some of the future directions which can aid one to arrive at a more robust solution for the HDR imaging problem.

The algorithms discussed in this thesis can be used to work on more complex datasets which would enable one to model the solutions better. The computational complexity of the algorithms can be reduced once faster bilateral filtering and Poisson solvers are integrated into the solutions as and when they are available. The results obtained using the proposed algorithms can be validated more thoroughly based on the human visual system model for accurate compositing of LDR images. This would also require us to arrive at a suitable algorithm for a given digital display device. The proposed approaches can be suitably modified to perform region based compositing of multi-exposure images by defining boundary conditions on the regions. We can also learn the different parameters using an appropriate model which would serve as an alternative to the non-parametric approaches developed in this thesis.

The LDR image generation algorithms developed for static and dynamic scenes can be implemented in real time using a GPU. Parallel implementation would help in fast computation of the desired LDR image. More complex scenes such as crowd motion can be addressed by improving the scene change detection algorithm. A plugin for Adobe® Photoshop can be developed using the proposed techniques as an alternative to the existing “Merge to HDR Pro” tool. This would enable the user the overhead of identifying proper tone mapping operator for a given scene. The proposed algorithms can also be used to build freewares like Enfuse and Tufuse which perform exposure fusion for static scenes.

The final LDR images can be visually evaluated by employing an user study consisting of a number of people. These people can be asked to rate the LDR images generated using various approaches. This would enable one to determine the quality of LDR images generated using direct LDR image generation methods with that of the popular tone mapping operators. The major advantage with such an approach is that it would relieve one from the need to use a dynamic range independent quality metric.
Bibliography


obopixels: Fast superpixels using geometric flows,” *Pattern Analysis and Machine Intelli- 

[167] T. Mertens, J. Kautz, and F. V. Reeth, “Exposure fusion: A simple and practical alter- 
native to high dynamic range photography,” *Computer Graphics Forum*, vol. 28, no. 1, 
Publications from this Work

Journals


Conferences


**Symposiums**

Acknowledgments

I would like to thank my advisor Prof. Subhasis Chaudhuri for his motivation and help during the course of this dissertation. He was very helpful with key suggestions while solving different problems addressed in the thesis. He was highly supportive and motivated me to publish the completed works at top conferences. I thank for the help rendered by him to bring out the thesis in good shape. I would like to thank the reviewers of the thesis Prof. Subhashis Banerjee, Prof. A. N. Rajagopalan, and Prof. Hendrik Lensch. Their valuable comments helped to improve the thesis.

I would like to thank Prof. Harish Pillai and Prof. SN Merchant for agreeing to be in my RPC committee and providing useful suggestions in the annual progress seminars. I am very thankful to Prof. Rohit Manchanda, my MTech advisor, who was very helpful throughout my stint at IIT Bombay. Special thanks are reserved for Prof. S. Jayaraman (PSG Tech), presently at BITS Dubai, who kindled interest in me to take signal processing as my research area. I would like to thank Prof. S. Arunkumar (IITB) and Prof. G. Sankarasubramanian (PSG Tech) who were very supportive.

I thank Microsoft Research India for offering me the PhD fellowship which gave me support financially during the course of my research. I thank Dr. P. Anandan and Dr. Manik Varma who encouraged and guided me when I developed a machine learning application while interning at Microsoft Research India. I also thank Dr. Lihi Zelnik-Manor who was also a collaborator while developing the algorithm. The internship at Microsoft Research India was very helpful in my progress later on.

I appreciate the help rendered by the following students who worked with me during the course of my thesis - Nikhil Pande and G. Suriya. I especially would like to thank Kumar Vishal who worked with me on a problem addressed in the thesis. We had a lot of discussions which helped me to enhance the work.

My seniors in the lab Dr. Vinay Namboodiri and Dr. V. Srikrishnan helped me during the
initial stages of my work. I thank them for providing me valuable pointers on various aspects of vision research. I was fortunate to get some of the bright students as my lab colleagues. I want to extend my gratitude to the following wonderful labmates who were helpful at various instants - Dr. Tanmay Pawar, Deven Patel, Shilpa Inamdar, Anup Shetty, Gourav Dhavale, NS Ananthakrishnan, Trupti Thete, Smita Nair, Amit Tilwankar, Ronak Shah, Kanchan Bahirat, Prashant Saxena, Anish Garg, Rishabh Iyer, Soham Mehta, Ranjith, and Sameer Hiware.

My friends in IIT Bombay deserve special mention as I had fun in the stressful environment due to their association. I was lucky to get some wonderful students as my friends at IIT. I especially thank my friends for spending memorable moments with me - Ramaprakash Bayadi, Swanand Khare, Srinivasan Krishnaswamy, Prasad Bokil, Harsha Simha, Debasattam Pal, Subashish Datta, Dr. VK Jadhav, Venkateswara Pai, Dr. Narayan Manjarekar, Dr. Rambabu, Dr. John Eric Steephen, Harish Katti, Naveen Shankar, Dr. Venkatsubramanian, Rachel Kalpana, Santhanakrishnan, Dr. Vijay Venu, Deepak Patil, and Jessy Roy.

I am highly thankful to my friend Dr. H. Priyadarshan for helping me at various critical junctures during my research. He has been an amazing researcher and a very good friend. I still relish the numerous discussions we had over time. I am also thankful to Dr. Sanjay Modi, Homeopath who helped me overcome several health problems during my research.

I thank the EE Department office staff at IIT Bombay, especially Mrs. Vaishali Deshpande, Mrs. Madhu Shetty, Devidas Bagul and Vilas for their kind help at different times. I also thank the open source operating systems especially Ubuntu and \LaTeX which made the thesis possible in quick time.

I would like to express my sincere thanks to the elite Computer Vision and Computer Graphics community for providing me with abundant literature which helped me solve the problems addressed in this thesis. I especially loved the works of Prof. Shree Nayar (Columbia), Prof. Richard Szeliski (MSR, Washington), Prof. Fredo Durand (MIT), Prof. William Freeman (MIT), Prof. Jitendra Malik (UC Berkeley) and Dr. Todor Georgiev (Adobe) and got motivated by their amazing works in computer vision. I enjoyed reading parts of the computer vision book by Prof. Richard Szeliski which is freely available for download online.

Reading general books helped me to relax a bit. I especially loved the books by Simon Singh, RK Narayan, Richard Feynman, ET Bell and George Gamow. I also thank cricinfo and atpworldtour tennis websites for updating me with latest scores. I loved exploring and connecting with like minded people using LinkedIn professional network and I acknowledge
their innovation for professional networking. Facebook and LinkedIn friends also deserve a special mention here.

Roger Federer is a great tennis legend who has inspired me in many ways. I admire him as a player and a person, he has motivated me to work towards excellence. Some of the old Hindi film songs by Mohammed Rafi, Lata Mangeshkar and Kishore Kumar are really great to listen. I would like to acknowledge these legends for the comfort I received listening to their songs. I could relax a bit by playing cricket against many online users at Howzat and it helped me to chill out on occasions.

Finally, I would like to acknowledge the support provided by my parents throughout the course of completion of the thesis. I missed their company for the past 12 years. I am happy that I will be living with them from now on.

Date: ________________________________ Shanmuganathan Raman