

StereoHDR: OPTIMAL FRAMEWORK FOR STEREO CAMERA-BASED HDR AND DEPTH

A Project Report

submitted by

PRADYUMNA VENKATESH CHARI (EE15B122)

in partial fulfilment of requirements

for the award of the degree of

BACHELOR OF TECHNOLOGY



**DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS**

June 2019

THESIS CERTIFICATE

This is to certify that the thesis titled **StereoHDR: OPTIMAL FRAMEWORK FOR STEREO CAMERA-BASED HDR AND DEPTH**, submitted by **Pradyumna Venkatesh Chari**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Kaushik Mitra
Research Guide
Assistant Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Place: Chennai

Date: 7th June 2019

ACKNOWLEDGEMENTS

I would like to thank Prof. Kaushik Mitra for allowing me to work on this topic of my interest. I would like to thank Mr. Anil Kumar Vadathya for helping me with understanding various parts of this project. Additionally, I would like to also thank the various members of the Computational Imaging Lab at IIT Madras for helping me with various aspects related to operating experiments and collecting the required data for experimentation.

ABSTRACT

KEYWORDS: High Dynamic Range Imaging; Stereo; Optimal Capture Sequence.

Dual cameras have added a new modality to mobile camera imaging. However, most of the recent applications have limited themselves to extracting only 3D information from them, or for view diversity during imaging. In this work, we utilize the dual camera setup’s simultaneous capture ability for HDR recovery in addition to the depth estimation. The simultaneous use of dual cameras enables faster HDR recovery, thereby reducing native ghosting artifacts as a result of moving objects in the scene. To solve for HDR and depth, we propose a *novel optimization framework* for obtaining an optimal exposure sequence, under the constraints of *minimal* capture time, *minimal* disparity error and *high* SNR for each exposure. Using the optimal stereo exposure stack captured we alternate between HDR reconstruction and disparity estimation. We utilize the estimated disparity to identify point correlations for estimating the Inverse Camera Response Function. The disparity is also used to warp the secondary (wide) view into the primary (narrow) view, for HDR recovery. Our experiments show that our optimal stereo exposure sequence performs better than most other sequences covering the dynamic range of interest. It also performs comparably with the full stereo stack spanning the entire dynamic range of interest, in terms of disparity and HDR.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF FIGURES	vi
ABBREVIATIONS	vii
NOTATION	viii
1 INTRODUCTION	1
1.1 Background	1
1.2 Objective	1
1.3 Structure of Thesis	2
2 LITERATURE REVIEW	3
2.1 HDR Imaging	3
2.2 Disparity Estimation	3
2.3 Stereo Camera based HDR	4
2.4 Other Related Work	4
3 BACKGROUND CONCEPTS	5
3.1 Camera Response Functions	5
3.2 Exposure Compensation	6
3.3 Noise in Imaging and the Signal to Noise Ratio	7
4 OPTIMAL FRAMEWORK FOR HDR AND DEPTH	8
4.1 HDR and Depth estimation	8
4.1.1 Capture Time	10
4.1.2 Dynamic Range Coverage	10
4.1.3 Disparity Error	11

4.1.4	Joint Optimization	11
4.2	ISO Control	11
4.3	Implementing the Optimization	12
5	PIPELINE STEPS	14
5.1	Scene Radiance Distribution Estimation	14
5.1.1	Accurate Radiance Distribution Estimation	14
5.1.2	Approximate Radiance Distribution Estimation	14
5.2	Iterative Disparity and CRF Estimation	16
5.2.1	CRF Estimation	17
5.2.2	Disparity Estimation	18
5.3	Image Fusion	19
6	EXPERIMENTS, RESULTS AND CONCLUSION	20
6.1	Validating the Framework	20
6.1.1	HDR Performance	20
6.1.2	Optimal HDR+Depth	23
6.2	Validating Pipeline Steps	24
6.2.1	ICRF Estimation	24
6.2.2	Iterative CRF and Disparity Fine-tuning	25
6.2.3	Simulated Saturation	25
6.3	Applications	26
6.3.1	Image Refocusing	26
7	CONCLUSIONS AND FUTURE DIRECTIONS	31

LIST OF FIGURES

1.1	An overview of the StereoHDR algorithm, along with a downstream application.	2
3.1	A pair of inverse Camera Response Functions. Note that the x-axis corresponds to the pixel value, while the y-axis corresponds to the log of the scene Radiance.	6
4.1	Description of the StereoHDR Algorithm Flow, highlighting various steps in the pipeline and the corresponding results at those steps. . .	9
4.2	Scene Description and Camera Performance Analysis Using Radiance Maps.	10
5.1	Figure description of the Accurate Scene Radiance Distribution Estimation. Interleaved images are captured from both cameras to cover the entire radiance range of interest.	15
5.2	Figure description of the Approximate Scene Radiance Distribution Estimation. A subset of the distribution is estimated, and the closest distribution to it is identified from a set of pre-defined distributions. . .	16
5.3	Figure description of the disparity estimation algorithm. While the core algorithm for disparity estimation is the one proposed by Mozerov and Weijer (2015), we add appropriate steps to accomodate the HDR scene requirements.	18
6.1	Description of the various scenes captured for experiments.	21
6.2	Comparison of the StereoHDR algorithm against other single camera HDR algorithms.	22
6.3	This figure analyzes the optimality of our Stereo HDR framework: (a) Optimal exposure sequence; (b) Relative ground truth sequence- 6 shots, exposure compensation of 1; (c) Test sequence- 3 shots, exposure compensation of 1; (d) Test sequence- 4 shots, exposure compensation of 1 (e) Test sequence- 5 shots, exposure compensation of 1; (f) Test sequence- 2 shots, exposure compensation of 2; (g) Test sequence- 3 shots, exposure compensation of 2 (h) Test sequence- 4 shots, exposure compensation of 2.	27
6.4	This figure describes the accuracy of the ICRF estimation algorithm for the main and secondary cameras, through comparison with ground truth responses. It also shows the main and secondary camera ICRFs on the same plot for comparison.	28

6.5	This figure describes the effectiveness of the proposed jointly iterative ICRF and disparity estimation framework, in terms of disparity error and cross camera log radiance estimation error.	29
6.6	This figure describes the performance of the Simulated Saturation step, by comparing the disparity estimation performance with and without simulated saturation for two scenes.	30
6.7	This figure describes the performance of the refocusing, performed using the HDR and depth estimates.	30

ABBREVIATIONS

HDR	High Dynamic Range
LDR	Low Dynamic Range
CRF	Camera Response Function
ICRF	Inverse Camera Response Function
SNR	Signal to Noise Ratio

NOTATION

t	Exposure time
$g_j(\cdot)$	log Inverse Camera Response Function

CHAPTER 1

INTRODUCTION

1.1 Background

Most commercially available cameras are limited by the range of scene radiance they can capture (without saturation). In such cases, scenes with a large radiance range, such as indoor-outdoor scenes or scenes with light sources, cannot be satisfactorily captured in a single image.

High Dynamic Range imaging, through appropriate fusion of several Low Dynamic Range images, has consequently been a well-researched solution to this problem. A stack of images, captured at varying exposures, are used to estimate the radiance of each scene point. These are then fused into one HDR image.

In recent years, however, cell-phone cameras have become the primary means of imaging and photography. The most recent disruption in the cell-phone camera sphere has been the proliferation of multi-camera setups. Two different cameras (wide and ultra-wide, wide and telephoto etc.) are increasingly being seen in most cell-phones. These setups are used either to obtain depth cues or for diversity in the types of image capture possible.

1.2 Objective

In this work, we develop a framework to use such stereo camera setups in cell-phones for High Dynamic Range scene representations and depth information acquisition. Both HDR imaging and depth estimation involve a common processing step: disparity estimation. However, as we shall show, both these tasks have different notions of optimality. Hence, we develop a joint framework of optimality for simultaneous estimation of both.

The major contributions of this work are as follows:

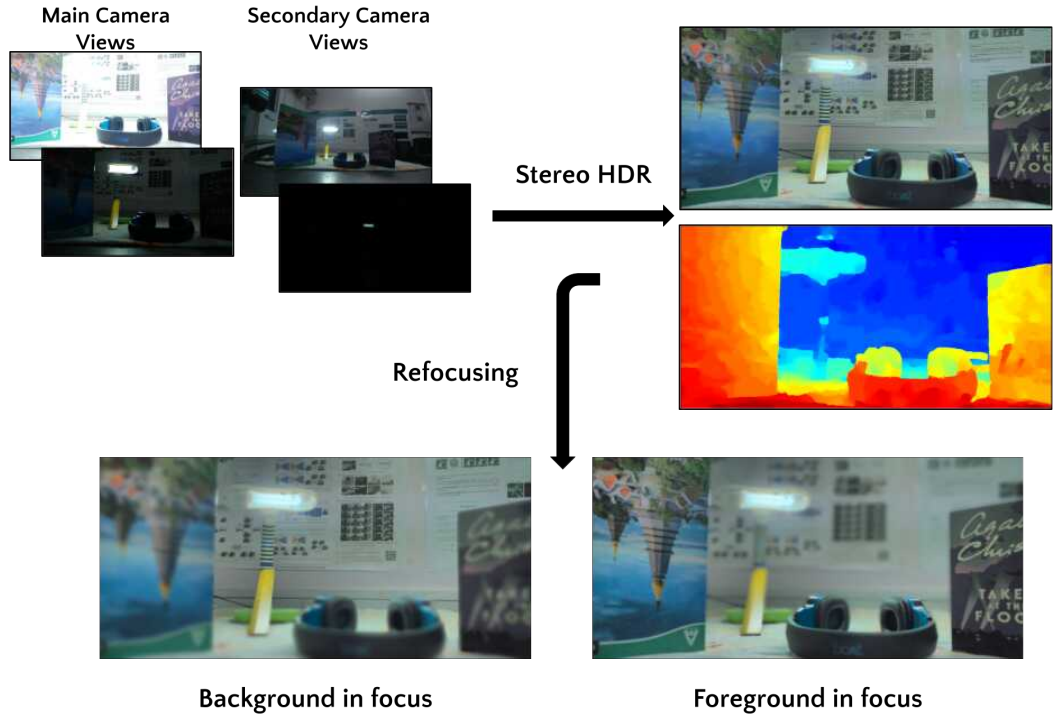


Figure 1.1: An overview of the StereoHDR algorithm, along with a downstream application.

- Framework for simultaneous HDR scene capture and depth estimation
- Pipeline for a general multi-image Stereo HDR setup
- Presenting results and viability of Stereo HDR for real world cell-phone data
- Characterizing time speed-up for HDR scene capture using a stereo setup
- Dataset for multi-image stereo HDR applications

1.3 Structure of Thesis

This thesis begins with a brief literature review of relevant research carried out in Chapter 2. Chapter 3 establishes a few background concepts relevant to this work. This is followed by a description of the formulation of the optimization framework and the regime for solving it, in Chapter 4. Chapter 5 addresses the remaining steps of the optimization pipeline. Chapter 6 rounds up the Results and Chapter 7 looks to highlight the Conclusions, along with future scope.

CHAPTER 2

LITERATURE REVIEW

2.1 HDR Imaging

High Dynamic Range imaging, through fusion of multiple Low Dynamic Range images is a well-researched topic in the domain of photography. Several previous works have looked into various aspects contained within HDR imaging, namely identifying required exposures, as in Pourreza and Kehtarnavaz (2015), conversion of images to radiance space by identifying and inverting the camera response function, as in Debevec and Malik (1997), and fusion of scene radiance information from multiple LDR images to obtain a single HDR image, which is addressed in Robertson *et al.* (1999). Additionally, deghosting, or the removal of motion induced artifacts, is also well explored, in works such as Hu *et al.* (2013). More recently, several deep learning based HDR frameworks, both for multi-image HDR and single-image HDR have been successful in obtaining state of the art results towards the same. Some promising approaches include Eilertsen *et al.* (2017) and Kalantari and Ramamoorthi (2017).

2.2 Disparity Estimation

Disparity, and consequently depth estimation using stereo cameras is also widely researched. The Middlebury Stereo Evaluation framework, developed over the course of Hirschmüller (2007), provides a comprehensive performance analysis for various disparity estimation algorithms. In this work, we do not focus on finding the optimal disparity estimation algorithm for our purposes. Instead, we make the framework and optimality criteria flexible with respect to the disparity algorithms. However, for the purpose of experiments, we use the Disparity estimation algorithm proposed by Mozev and Weijer (2015).

2.3 Stereo Camera based HDR

Using stereo camera setups for HDR imaging has been explored in past works. Lin and Chang (2009) established an early pipeline for HDR imaging using a stereo camera setup. Bätz *et al.* (2014) extends the formulation to account for stereo HDR for videos. Park *et al.* (2017) looks at improving the pipeline for Stereo HDR over the previously cited works. However, to the best of our knowledge, there has been no research towards jointly optimal HDR and depth capture, by designing a framework for said optimality.

2.4 Other Related Work

The idea of optimally selecting exposures and ISOs for imaging was previously proposed in Hasinoff *et al.* (2010). Here, the optimality was established based on the notion of maximizing SNR and/or minimizing capture time.

CHAPTER 3

BACKGROUND CONCEPTS

3.1 Camera Response Functions

Mathematically, a camera works by mapping the scene radiance to a value between 0 to $2^n - 1$, where n is the bit-depth for the camera. That is, for a given sensor gain (ISO)

$$I = f(\phi t), \quad \tau_l < \phi t < \tau_u, \quad (3.1)$$

where I is the image pixel value, $f(\cdot)$ is the monotonic Camera Response Function (CRF), ϕ is the radiance of the scene point, t is the exposure duration and τ_l and τ_u are the lower and upper pixel value thresholds, beyond which the pixel is noisy or over saturated respectively.

Rewriting the limits of (1) provides us with the following range for the captured radiances:

$$\log(\tau_l) - \log(t) < \log(\phi) < \log(\tau_u) - \log(t) \quad (3.2)$$

An image captured at a particular exposure can therefore be interpreted as capturing a particular range of radiances on the log radiance scale (subsequently referred to as *log radiance interval* in this work). The standard HDR problem, in such a framework, reduces to identifying a set of exposures, so as to span the dynamic range of interest.

Figure 3.1 shows a pair of inverse Camera Response Functions, which map from the pixel intensity space to the scene radiance space. One may note that CRF estimation is accurate up to a scale factor, or in the log scale, up to a constant offset. However, for applications such as HDR, this scale uncertainty is not of concern, since most HDR images, while being displayed, undergo a tone-mapping process which only looks at relative radiance values.

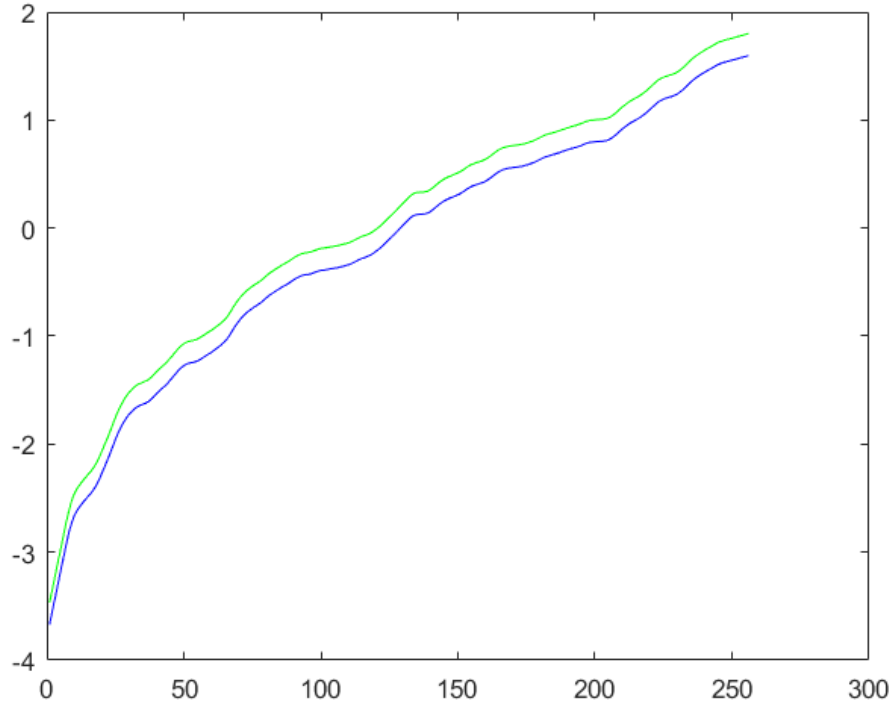


Figure 3.1: A pair of inverse Camera Response Functions. Note that the x-axis corresponds to the pixel value, while the y-axis corresponds to the log of the scene Radiance.

3.2 Exposure Compensation

A common method of choosing images for single camera HDR is to use exposure bracketing. Here, the effective exposure (obtained by varying both the ISO and shutter speed) between successive captured images are related by a constant multiplying factor. That is,

$$t_{exp}^{i+1} = t_{exp}^i \times 2^C \quad (3.3)$$

Where t_{exp}^i is the effective exposure of the i^{th} image from a particular camera, and C is the exposure compensation factor.

Commercial cameras often allow exposure compensations anywhere from -2 to +2, with the integer compensations being common choices for HDR photography.

3.3 Noise in Imaging and the Signal to Noise Ratio

The imaging process is inherently affected by noise. Noise occurs in various forms, ranging from thermal noise, due to sensor electronics, to quantization noise arising out of the Analog to Digital conversion step. Boyat and Joshi (2015) explains a few of these noise sources, and their corresponding models, in some detail.

Based on our understanding of these noise sources, the Signal to Noise Ratio can be identified and defined as a metric for image quality. From ?, we use the following definition for SNR:

$$SNR^j = \frac{\phi^2 t_i^{j2}}{\phi t_i^j + \sigma_r^2 + \sigma_q^2 g_i^{j2}} \quad (3.4)$$

Where ϕ is the scene point radiance, t_i^j is the exposure duration of the i^{th} image in the j^{th} camera, g_i^j is the sensor gain of the i^{th} image in the j^{th} camera, σ_r is the read noise and σ_q is the quantization noise. Note that the ISO setting and the sensor gain are related as $ISO = \frac{K}{g}$, where K is a camera-dependent constant. Additionally, the imaging model used here assumes a linear Camera Response Function, for simplicity in computation. Hence, in non-saturated regions, the image pixel value $I = \frac{\phi t_i^j}{g_i^j} + I_{bias}$.

CHAPTER 4

OPTIMAL FRAMEWORK FOR HDR AND DEPTH

Our goal in this work is to recover an HDR image and depth map of the scene using optimal exposure sequences captured using cell-phone stereo cameras. Since the cameras can be accessed independently, we want to reduce the capture time by splitting exposure sequences among the left and right camera. Note that estimating depth in this case is no longer similar to the traditional depth from stereo, since the intensity values differ across the stereo pairs due to the exposure differences, violating the brightness constancy assumption. Ideally, without capture time constraints, a naive solution would be to span the entire exposure stack on both cameras, so as to capture the entire radiance range in each. This provides the best case for disparity estimation among the stereo pair with intensity values remaining close due to exposure overlap. However, this is not an optimal solution in terms of the total acquisition time.

As a result of non-identical exposures on the two cameras, the disparity estimation step would require transformation to radiance space followed by tonemapping. Consequently, the resulting HDR and depth are not independent, thereby leading to iterative refinement for improvements.

Fig. 4.1 describes our proposed pipeline for jointly optimal HDR and depth estimation. We begin by looking at the optimality framework and the domains of optimization for joint stereo depth and scene HDR data acquisition. Following this, we shall describe the pipeline to acquire the said depth and HDR information from a given set of images from both cameras.

4.1 HDR and Depth estimation

Any scene to be captured can be represented on the basis of its radiance distribution, a histogram of the various radiance values in the scene. In order to estimate the HDR image, the entire dynamic range of interest must spanned, such that each radiance is captured, without saturation, by at least one camera.

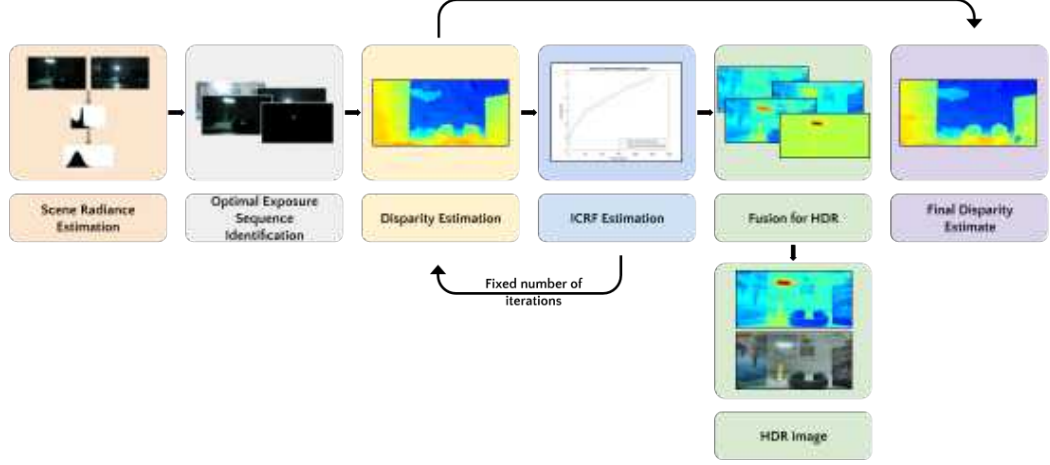


Figure 4.1: Description of the StereoHDR Algorithm Flow, highlighting various steps in the pipeline and the corresponding results at those steps.

Additionally, since we are dealing with a stereo camera setup, the second aspect relates to disparity estimation. This is required both for depth estimation as well as for HDR representation. The performance of disparity estimation algorithms can be understood in terms of (a) Radiance coverage histograms, and (b) Spatial correlations. If a particular scene radiance is captured without saturation in both cameras, disparity can be estimated for that radiance. However, even if certain radiance values are not captured unsaturated in both cameras, accurate disparity may still be estimated due to spatial correlation and smoothness constraints that most disparity estimation algorithms espouse.

In this work, we look solely at the scene radiance histogram representation in order to set up the optimization for HDR and disparity. The effects of spatial correlation on disparity estimation can be considered to be algorithm-dependent and can be abstracted to a disparity error-allowance in the optimization framework, as will be shown later.

Fig. 4.2 shows an example scene radiance distribution and the radiance coverage of the images from the cameras. Through this, we can identify the criteria that the images must satisfy in order to capture an optimal representation:

- The capture time for the images should be minimum (under the assumption that simultaneous capture is possible from both cameras)
- Every radiance value within the dynamic range of interest must be unsaturated in at least one image
- The fraction of scene pixels without unsaturated images from both cameras should be less than a threshold value

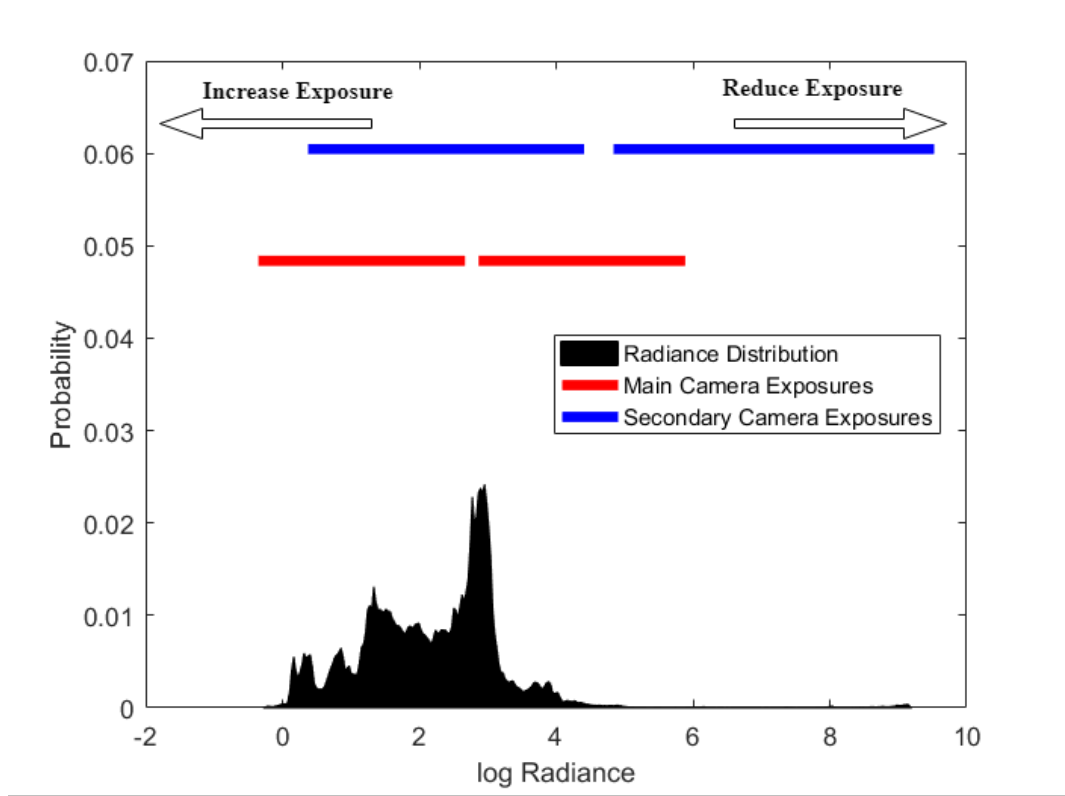


Figure 4.2: Scene Description and Camera Performance Analysis Using Radiance Maps.

We shall now set up the optimization framework for a stereo camera setup where both the cameras can be independently accessed, configured and utilized.

4.1.1 Capture Time

Let the various image exposures be represented as t_j^i , for the j^{th} image in the i^{th} camera. Let m images be captured from the main camera, and n images be captured using the secondary camera. Then, the overall capture time can be given as,

$$t_{cap} = \max(\sum_{j=1}^m t_j^1, \sum_{j=1}^n t_j^2) \quad (4.1)$$

4.1.2 Dynamic Range Coverage

Let K_j^i represent the log radiance interval captured by the j^{th} image from the i^{th} camera. Additionally, let R represent the radiance range of interest, which we wish to capture

using our imaging setup. Then, the various log radiance ranges must satisfy,

$$(\cup_{j=1}^m K_j^1) \cup (\cup_{j=1}^n K_j^2) \supseteq R \quad (4.2)$$

4.1.3 Disparity Error

Let $h(\cdot)$ represent the probability distribution function corresponding to the log radiance histogram for the scene, and let τ_{err} be the allowed error in disparity. Then, the disparity error criterion reduces to,

$$\int_O h(x)dx \geq 1 - \tau_{err}, \quad O = (\cup_{j=1}^m K_j^1) \cap (\cup_{j=1}^n K_j^2) \quad (4.3)$$

4.1.4 Joint Optimization

Based on the above description of the various factors involved, we now set up the optimization framework. It is as follows:

$$\begin{aligned} & \text{Minimize } t_{cap} \\ & \text{Subject to} \\ & (\cup_{j=1}^m K_j^1) \cup (\cup_{j=1}^n K_j^2) \supseteq R \\ & \int_O h(x)dx \geq 1 - \tau_{err}, \quad O = (\cup_{j=1}^m K_j^1) \cap (\cup_{j=1}^n K_j^2) \end{aligned} \quad (4.4)$$

The set of exposures for each camera that arise out of this optimization characterize the optimal exposure capture sequence.

4.2 ISO Control

We now explore an additional domain of control towards optimal exposure sequences. Increasing ISO allows for capturing a certain radiance range with a lower exposure time. As a trade-off, the noise characteristics of the image change with changing ISO. In general, increasing the ISO leads to a degradation in the SNR for an image, as described in Chapter 3.

For a given log radiance interval to be captured, the sensor gain and exposure time must satisfy $\frac{t}{g} = \text{constant}$. This characterizes the relation to be satisfied, between the ISO and shutter speed (exposure time).

This aspect can now be included in the optimization framework to further allow for freedom to reduce capture time. We propose a lower threshold on the worst-case (minimum) SNR (similar to ?) for each of the captured images. Using this, the optimized framework can be rewritten as:

$$\begin{aligned}
& \text{Minimize } t_{cap} \\
& \text{Subject to} \\
& (\cup_{j=1}^m K_j^1) \cup (\cup_{j=1}^n K_j^2) \supseteq R \\
& \int_O h(x) dx \geq 1 - \tau_{err}, \quad O = (\cup_{j=1}^m K_j^1) \cap (\cup_{j=1}^n K_j^2) \\
& \min_{\phi \in K_i^j} SNR^j(\phi, t_i^j, g_i^j) \geq SNR_{min}, \quad i \in \{1, 2, \dots, m\}, \\
& j \in \{1, 2\}
\end{aligned} \tag{4.5}$$

4.3 Implementing the Optimization

The optimization framework previously established has a non-convex nature. In addition, closed-form solutions are not easily obtained. In order to make the problem more tractable, we split it into two parts: first, an initial exposure and ISO estimate that satisfies the required constraints, followed by a final optimization over exposure times, while keeping the ISOs constant.

By virtue of the nature of the formulation, the choice of initial estimate is essential in order to obtain a sufficiently optimal solution. The initial estimate has two aspects: exposures and ISOs. For a given ISO configuration, the exposures are chosen such that the entire dynamic range is covered. In addition, since our objective is to minimize the capture time, we choose exposures such that most of the disparity error is towards the lower radiance parts of the scene histogram. As a result, the exposures are shifted to lower values. In order to optimize over ISO, each image is considered independently, and the ISO for that image which minimizes the overall capture time, while keeping

other ISOs constant, is chosen.

This initial estimate is optimized by an iterative objective minimization approach using the Levenberg-Marquardt algorithm. This approach provides a locally optimal exposure sequence. The experiments and results will further analyze the nature and robustness of this optimality.

CHAPTER 5

PIPELINE STEPS

Having looked at the optimization scheme, we now describe the remaining pipeline steps from Figure 4.1 to facilitate the estimation of HDR representation and depth estimation, for a High Dynamic Range scene.

5.1 Scene Radiance Distribution Estimation

In order to utilize the optimality framework described earlier, the radiance distribution for the scene must be estimated. We look at two methods for the same:

5.1.1 Accurate Radiance Distribution Estimation

In order to identify accurate radiance distribution for a scene, multiple images can be used from both cameras, as a pre-processing step, so as to span the radiance range of interest. Since we deal with small baseline camera setups, the radiance distributions seen from both the cameras can be assumed to be similar. Hence, the exposures from the two cameras can be interleaved, to capture alternating subsets of the range of interest. By fusing information from each of these images, and with a prior Camera Response Function estimate, the radiance distribution can be estimated.

Figure 5.1 describes this process. This procedure involves a relatively time-consuming capture of a stack of images. However, this step, being a pre-processing step, does not play a role in the actual image captured. Hence, the capture time for the images used for HDR representation will be minimal, thereby allowing for the benefits of the same.

5.1.2 Approximate Radiance Distribution Estimation

Here, we capture a single image from each of the cameras. Using these, limited information about the radiance distribution of a scene can be obtained. This information can

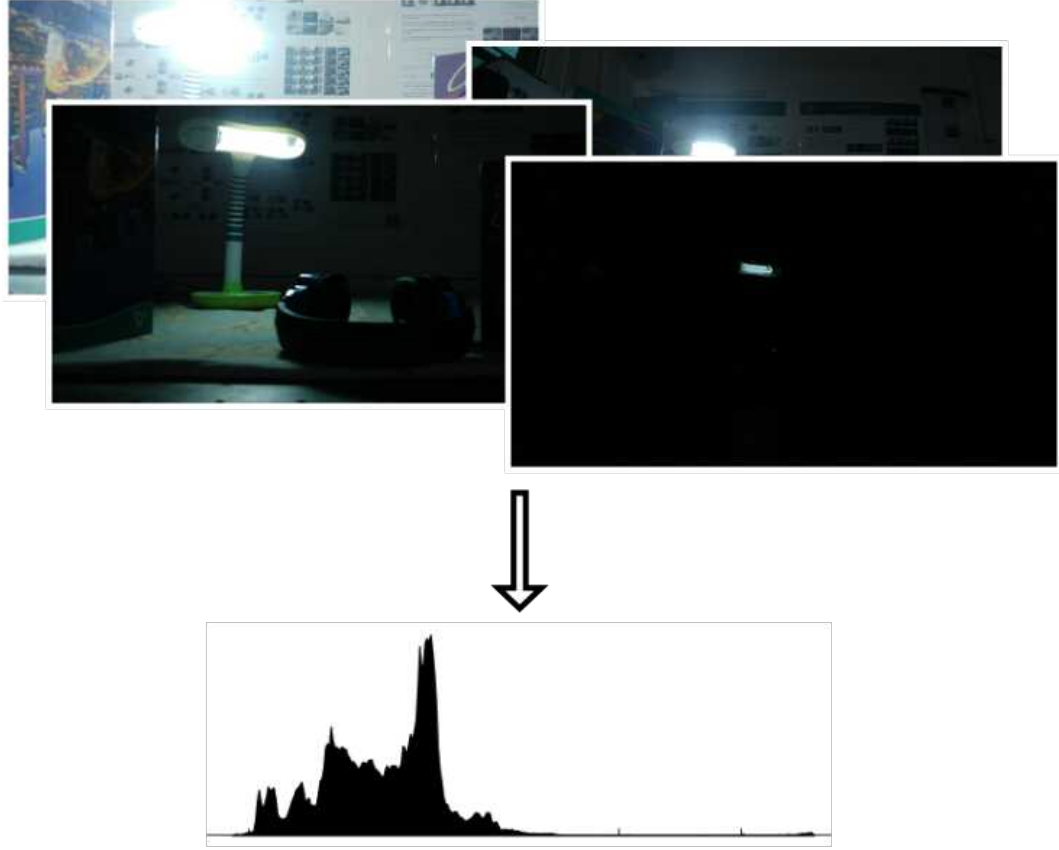


Figure 5.1: Figure description of the Accurate Scene Radiance Distribution Estimation. Interleaved images are captured from both cameras to cover the entire radiance range of interest.

be used to map the scene radiance distribution to the closest among a set of prior distributions, for different scene types. While this scene radiance estimation process is much quicker, it will only provide an approximate result, thereby allowing for a sub-optimal exposure sequence. Figure 5.2 describes this method of estimation.

Several different regimes may be adopted to identify the nearest radiance distribution. One method may be to compute a prior set of radiance distributions, among which the nearest distribution may be identified using an appropriate metric, for instance the Earth Movers' Distance. Alternatively, a parametric model may be identified, to fit best to the known distribution, given the radiance range of interest.

In all further experiments and discussions in this work, we use the Accurate Scene Radiance Distribution. Identifying suitable algorithms for the Approximate Scene Radiance Distribution Estimation is left for a future work.

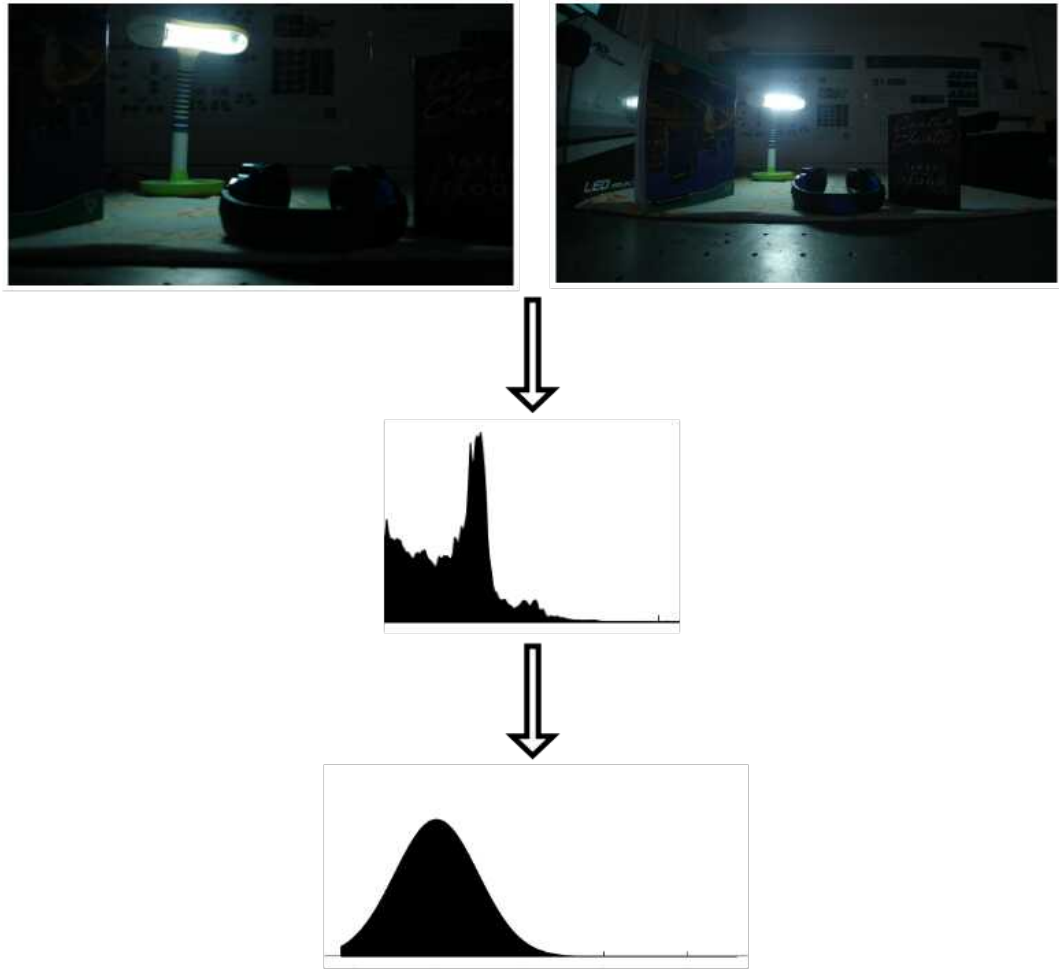


Figure 5.2: Figure description of the Approximate Scene Radiance Distribution Estimation. A subset of the distribution is estimated, and the closest distribution to it is identified from a set of pre-defined distributions.

5.2 Iterative Disparity and CRF Estimation

In order to estimate accurate disparity for a scene, we utilize an iterative approach towards estimating the disparity and the inverse Camera Response Function (ICRF). In each iteration, the ICRF estimate is refined using the point correspondences from the disparity estimate of the previous iteration. For the first iteration, prior known CRFs for both the cameras are used.

The estimated ICRFs are then used to transform the images to radiance space. Following processing, this is tonemapped back to the image domain in order to estimate disparity. The iterations are continued until both the disparity estimate and the CRF stop improving considerably. We now describe the two estimation processes in some detail.

5.2.1 CRF Estimation

The problem of estimating the CRF is reduced to a least squares estimation problem, as an extension to Debevec and Malik (1997). The cost function for this is given by:

$$C = C^1 + C^2 \quad (5.1)$$

$$C^j = \sum_i \sum_{p \in S_i^j} (w^d(Z_{i,p}^j) (g_j(Z_{i,p}^j) - E_p - \log(t_i^j)))^2 + \lambda_{sm} \sum_{l \in \{1, \dots, 254\}} (w_{sm}(l) g_j''(l))^2 \quad (5.2)$$

Where $g_1(\cdot)$ is the ISO normalized log inverse CRF for the main camera and $g_2(\cdot)$ is the ISO normalized log inverse CRF for the secondary camera. S_i^j is the set of points in the i^{th} image of the j^{th} camera, that are valid for estimated (that is, not saturated). E_p is the scene radiance for each of these points, and t_i^j is the corresponding exposure time. $Z_{i,p}^j$ is the pixel value for scene point p in the i^{th} image of the j^{th} camera.

The optimal main and secondary Camera inverse CRFs are therefore identified by optimizing this objective, using Least Squares. However, in practice, it is observed that estimation of 512 variables (for an 8 bit image) leads to an under-determined system. This leads to inaccurate ICRFs being estimated.

In this work, we apply a further relaxation step in order to make the ICRF estimation step more tractable. We assume that the two ICRFs to be estimated differ only by a constant offset factor. That is,

$$g_2(Z) = g_1(Z) + c \quad (5.3)$$

Where $g_2(Z)$ is the ICRF for the secondary camera, $g_1(Z)$ is the ICRF for the main camera and c is the offset factor. Z is the pixel intensity value, ranging from 0 to 255 in our case.

The performance of the ICRF estimation algorithm will be addressed in Chapter 6.

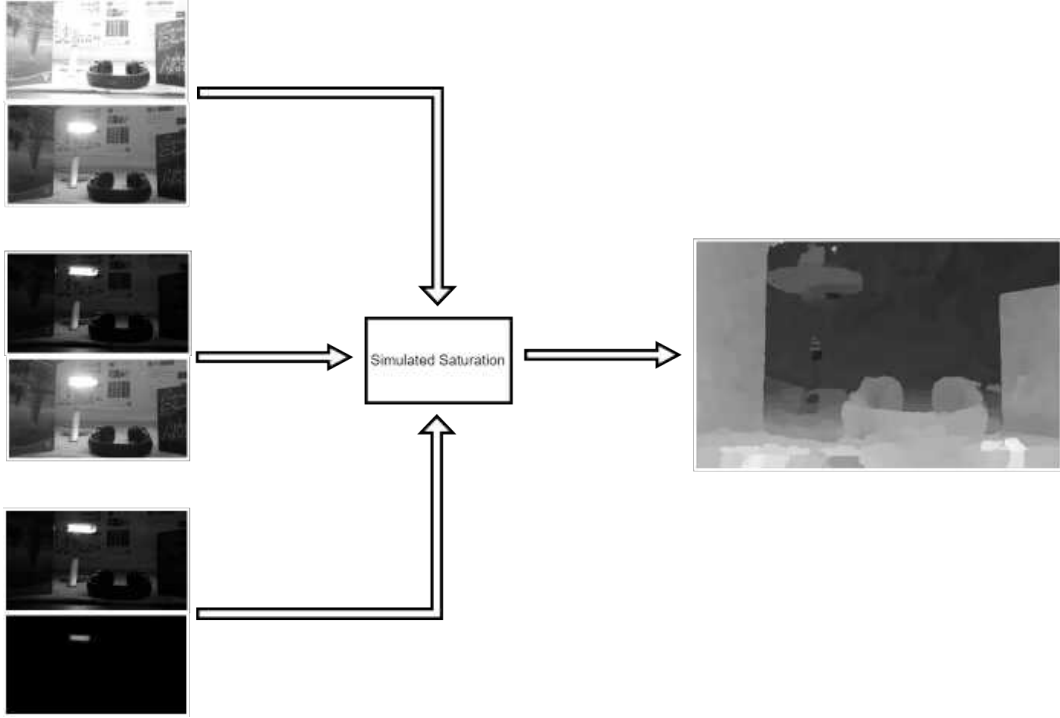


Figure 5.3: Figure description of the disparity estimation algorithm. While the core algorithm for disparity estimation is the one proposed by Mozerov and Weijer (2015), we add appropriate steps to accommodate the HDR scene requirements.

5.2.2 Disparity Estimation

The estimated CRFs are used to transform the images into radiance space. Following this, we estimate the HDR image for each of the cameras, by fusing various images for that camera. Disparity is estimated between these two tone-mapped HDR images. In order to improve the depth estimation, we introduce the notion of *Simulated Saturation*. Each picture of the pair is appropriately thresholded (in radiance space) so that both images occupy the same radiance range. In essence, the lower exposed image is thresholded for high pixel values and higher exposed image is thresholded for lower pixel values. Through this, the tone-mapping process, which is relative to the radiance range of the image being tone-mapped, would be similar for both the images even though their original radiance ranges are different. This results in more accurate disparity estimation. This induced accuracy in tone-mapping is at the cost of saturating out the radiances which are not present in both the images. However, it is experimentally observed that the loss of radiance information is sufficiently compensated by the tone-mapping process resulting to similar looking images, for disparity estimation. Chapter

6 analyzes the benefit of using simulated saturation in our case. Figure 5.3 describes this the disparity estimation pipeline.

Note that in our work, we do not look at optimizing for specific disparity estimation algorithms. Instead, we aim to create a general, modular optimization framework. This is in an effort to allow for provisions by which improvements in disparity estimation algorithms may be integrated into the framework. As mentioned earlier, for the purpose of experiments and results, we use the disparity estimation algorithm proposed by Mozerov and Weijer (2015).

5.3 Image Fusion

The disparity estimates and the fine-tuned inverse Camera Response Functions can be used to fuse the radiance information into the consolidated HDR image. The disparity estimates are used to warp the secondary view images into the main view, while the ICRFs are used to transform all the captured images into the radiance space.

The weights for image fusion are chosen based on the pixel intensity values at each point. The radiance for a scene point may only be estimated from an image where the pixel is neither saturated nor noisy (that is, the pixel values are within the thresholds defined in Chapter 3). Additionally, higher weightage is given to the information from the main view, in order to minimize stray occlusion-induced artifacts from the warped secondary view image.

In order to render the HDR image for display on traditional devices, a final tone-mapping is applied to the fused HDR image. This may be followed by a color correction step in order to transfer color information from the LDR images, in case the tone-mapping does not allow for related controls.

CHAPTER 6

EXPERIMENTS, RESULTS AND CONCLUSION

The setup used for testing this work included images captured using an LG G5 cell-phone camera. The camera setup consists of a wide and an ultra-wide angle camera. Before running our pipeline, the images from the cameras are appropriately rectified, in order to account for the differing fields of view. Additionally, effects of lens non-idealities such as radial and tangential distortion are appropriately accounted for and corrected.

Figure 6.1 highlights the various scenes captured for experiments. The figure displays, for each scene, the narrow view at high exposure, the wide view at low exposure and the radiance distribution. As can be seen, all the chosen scenes have radiance distributions with differing nature (unimodal, bimodal etc.) as well as differing radiance ranges. This allows us to establish the versatility and robustness of our optimization framework and the subsequent processing pipeline.

6.1 Validating the Framework

We first look to validate the notion of optimality established previously and to understand the various trade-offs involved. We try to establish empirical evidence towards the effect and understand the extent of the optimality involved.

6.1.1 HDR Performance

In order to quantify the performance of the StereoHDR algorithm, in terms of capturing HDR information, we compare it to three other results. First, we look at the HDR image obtained from fusion of three LDR images, with an exposure compensation of 2 between successive images. We refer to this setup as *Three Shot*. Second, we look at the image generated by the proprietary HDR algorithm present on the cell-phone in use (the LG G5 in our case). Finally, we look at the full Stack HDR image, obtained by taking

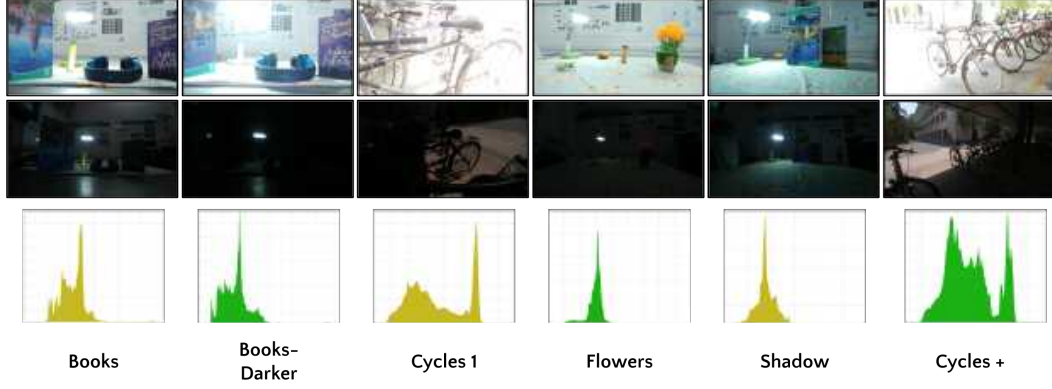


Figure 6.1: Description of the various scenes captured for experiments.

images with an exposure compensation of 2, with as many images taken as required to cover the dynamic range of interest. For instance, under this setting, 5 images were captured for the *Shadow* scene, 4 for the *Flowers* scene and 3 images for the *Cycles +* scene. The images obtained using full stack HDR are considered as ground truth for the purpose of this experiment. Such a setting, using an exposure compensation of 2, was chosen since exposure compensation based image stack capture is common for commercial HDR capture.

For all the scenes under consideration, the Three Shot HDR images are unable to capture the dynamic range of interest. In addition, as can be seen for the *Cycles +* scene, the Three Shot HDR image is also unable to appropriately reproduce color information. This arises due to saturation of the corresponding scene points, leading to inaccurate color information being sensed.

Similarly, the in-built HDR framework is also unable to capture the entire dynamic range, in the scenes under consideration. However, it performs better than the Three Shot setup, and is able to reproduce scene color information accurately.

Our proposed framework shows performance close to ground truth for all the scenes. It is able to appropriately capture the dynamic range of interest, while also being able to reproduce color accurately. The color reproduction can be seen easily in the results for the *Cycles +* scene. One may however notice the glare artifact visible in the results for the *Shadow* scene. This is a result of lens glare, augmented by the high dynamic range of radiances present in the scene.

Using a stereo setup for HDR also allows for capture-time speedup, in compari-













	Shadow	Flowers	Cycles +
3 shots; Compensation of 2			
	Capture time= 0.6583s	Capture time= 0.1313s	Capture time= 0.0108s
In-Built HDR			
	Capture time= N.A.	Capture time= N.A.	Capture time= N.A.
Ours			
	Capture time= 0.5125s	Capture time= 0.1006s	Capture time= 0.0086s
Optimal (GT); Exposure Compensation of 2			
	Capture time= 0.6688s	Capture time= 0.1332s	Capture time= 0.0111s

Figure 6.2: Comparison of the StereoHDR algorithm against other single camera HDR algorithms.

son with single camera HDR. This is because the exposures can be split between the two cameras. Additionally, unlike simultaneous HDR and depth estimation, this application requires disparity estimation for only the saturated regions in the main-view images. Therefore, the optimization framework may be run with a higher disparity error allowance. As shown in Fig. 6.2, our algorithm consistently provides the best capture time over all other image capture regime. This is because the optimization minimizes radiance redundancy during capture, while keeping SNR constraints under check. In the scenes considered, our algorithm is able to achieve a time gain of 22.15%, 23.38% and 20.37% over the Three Shots HDR image, for the Shadow, Flowers and Cycles + scenes respectively, resulting in an average time speedup of 21.97%.

6.1.2 Optimal HDR+Depth

We now look to analyze the optimality of the stereo HDR framework, through comparison with other image selection regimes. To establish optimality, we consider the disparity error and capture time as relevant metrics.

Since to the best of our knowledge, no established framework for selecting images for stereo HDR exists, we use a variant of exposure bracketing on both images. The exposure selection rule involves capturing an equal number of images in the narrow and wide views, with identical exposure compensations for each image stack. The narrow view exposures start from the image with the lowest possible exposure time, that captures the lowest radiance of interest. Each subsequent image exposure time is obtained using Equation 3.3, thereby creating a stack of images with successively decreasing exposures. Similarly, for the wide view, the exposures start from the image with the highest possible exposure time, that captures the highest radiance of interest, leading to a stack of images with successively increasing exposures.

We present results for exposure compensation factors of 1 and 2 (which are commonly used for commercial single camera HDR), for cases with 2, 3, 4 and 5 images per camera. Figure 6.3 shows these results. Note that only those exposure sequences are considered which cover the entire radiance range of interest, since the search space for the optimization is constrained similarly. As a result, the analysis of stereo HDR optimality reduces to analyzing disparity estimation optimality, under capture time considerations.

The accuracy of the disparity estimated depends on the range of radiances covered in both the cameras. Hence, the disparity estimate improves with an increase in the number of images captured, in our test regime. We therefore use an exposure sequence which captures the entire dynamic range in both cameras to identify a relative ground truth disparity estimate.

The optimal sequence is found to have the least capture time when compared to the test sequences. It is also found to have a better disparity estimation performance than most of the sequences, which establishes the trade-off between capture time and disparity error that our framework looks to optimize over. The allowed error in disparity, which is an input to the optimization scheme, establishes the tolerance towards disparity

error, depending on the disparity estimation algorithm and application. The analysis shows that the optimization scheme indeed provides us with a scene capture regime that is more optimal (with regards to capture time and disparity error) as compared to various naively chosen exposure sequences.

6.2 Validating Pipeline Steps

In this set of results, we analyze various pipeline steps in order to qualify and quantify their performance. We provide empirical evidence of their effectiveness towards accurate estimation of HDR representation for scenes.

6.2.1 ICRF Estimation

Here, we establish the accuracy of the ICRF estimation process highlighted in Chapter 5. For the purpose of comparison, therefore, we evaluate the ground truth ICRFs for both the cameras, by capturing single camera exposure stacks and estimating the ICRF using the method proposed in Debevec and Malik (1997). Additionally, to minimize the effect of false point correspondences, we use the disparity estimate obtained from the full range exposure compensation setup, which was also used to identify the ground truth disparity in Section 6.1.2. Additionally, as mentioned earlier, the ICRF may only be estimated up to a scale factor from such setups. Therefore, while comparing with ground truth ICRFs, we remove any average offset that may be present between the two responses.

Figure 6.4 shows the relevant results. Both the main and secondary camera ICRFs are seen to be estimated accurately, and they can be seen to closely resemble the corresponding ground truth ICRFs. Additionally, we can also see the estimated main and secondary camera responses on the same radiance scale. This shows the presence of an offset between the two cameras.

Based on these results, we are able to establish that for the present setup, the relaxation of assuming identical ICRFs separated by an offset is general enough to suitably account for system characteristics.

6.2.2 Iterative CRF and Disparity Fine-tuning

We now look to validate the jointly iterative ICRF and disparity estimation framework. For this purpose, we purposely choose corrupted initial estimated for the main and secondary ICRFs. This is achieved by introducing a large offset in one of the ICRFs, with respect to the other. This would have the effect of adversely affecting the cross camera radiance estimation error, the simulated saturation step and consequently the disparity estimation setup.

The results can be seen in Figure 6.5. The error in disparity is seen to decrease with the iterations. The initial error in disparity arises as a result of the corrupted ICRFs fed in as initial estimates. However, the disparity drops to the disparity error value of convergence within 1-2 iterations. However, the rate of convergence directly depends on the robustness of the disparity estimation algorithm in use. Therefore, the parameters controlling the number of iteration allows for adjustment, and must be adjusted depending on the disparity estimation algorithm in use.

The effectiveness of the iterations can also be seen in terms of the cross camera radiance estimation error. This metric is defined as the average error in log radiance estimated for a scene point by the two cameras. Since the initial ICRFs were corrupted, the initial error is very high. However, the error drops very rapidly to converge at a low error value. Again, the rate of convergence depends on the robustness of the disparity estimation algorithm, since the estimated disparity is the source of point correspondences for ICRF estimation. And persistent/large scale errors in disparity estimation would therefore degrade the ICRF estimation performance as well.

6.2.3 Simulated Saturation

The final pipeline step we look to validate is the Simulated Saturation step. The validation is carried out over two scenes, Books- Darker and Flowers. For the first scene, we choose the optimal exposure sequence. For the second scene, we choose a jointly exposure compensated sequence (non-optimal) between the two cameras. Using the optimal (ground truth) ICRFs, we perform the radiance space conversion, leading up to the simulated saturation. This was done so that the results observed were solely due to the effect of simulated saturation.

Figure 6.6 holds the results. For the Books- Darker scene, the lack of simulated saturation leads to spread out disparity estimation errors. The performance is degraded, which can be verified both visually, as well as in terms of the average absolute disparity error, with respect to the ground truth. For the Flowers scene, however, the disparity estimation specifically fails for the lamp region. This can be visually seen and is corroborated by the worse average absolute disparity error metric.

Over the two scenes, therefore, simulated saturation augmented disparity estimation is able to provide much better disparity estimates than simple disparity estimation, without simulated saturation.

6.3 Applications

In this section, we attempt to look at some possible downstream applications of the joint estimated HDR and depth. Primarily, we look at image refocusing as an application, in order to understand if the disparity estimates are good enough for the purpose.

6.3.1 Image Refocusing

Here, we display some of our results pertaining to HDR refocusing. This is achieved by segmenting the HDR image based on the disparity values, and applying segment-dependent blur kernels, depending on the location of the focal plane.

The two scenes used for this are the Cycles + and the Books scene. In both, the first image is focused on the background, while the second image is focused on the foreground. As can be seen, these results are perceptually good, since the difference between the foreground and the background can be quite easily identified and resolved.

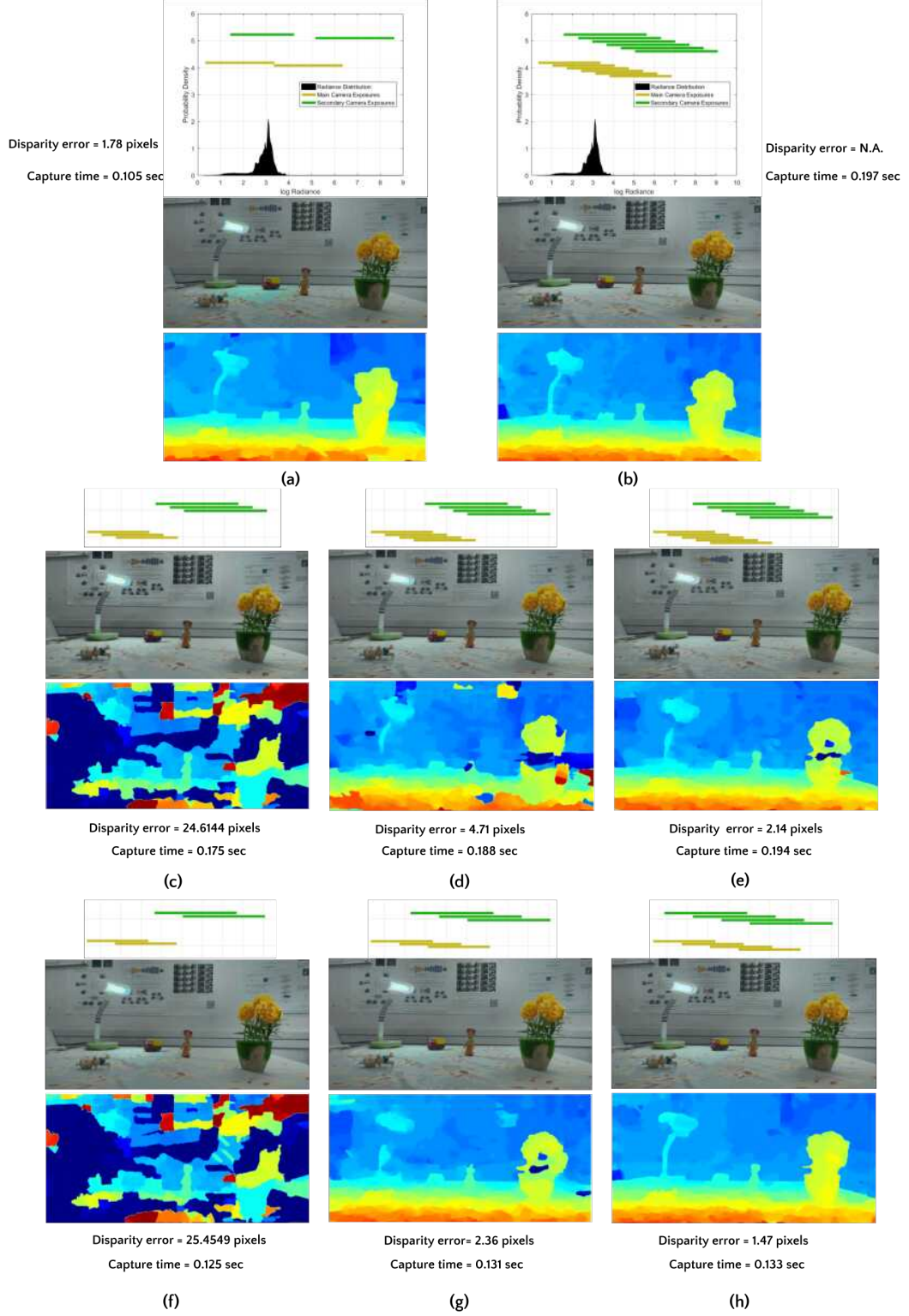


Figure 6.3: This figure analyzes the optimality of our Stereo HDR framework: (a) Optimal exposure sequence; (b) Relative ground truth sequence- 6 shots, exposure compensation of 1; (c) Test sequence- 3 shots, exposure compensation of 1; (d) Test sequence- 4 shots, exposure compensation of 1 (e) Test sequence- 5 shots, exposure compensation of 1; (f) Test sequence- 2 shots, exposure compensation of 2; (g) Test sequence- 3 shots, exposure compensation of 2 (h) Test sequence- 4 shots, exposure compensation of 2.

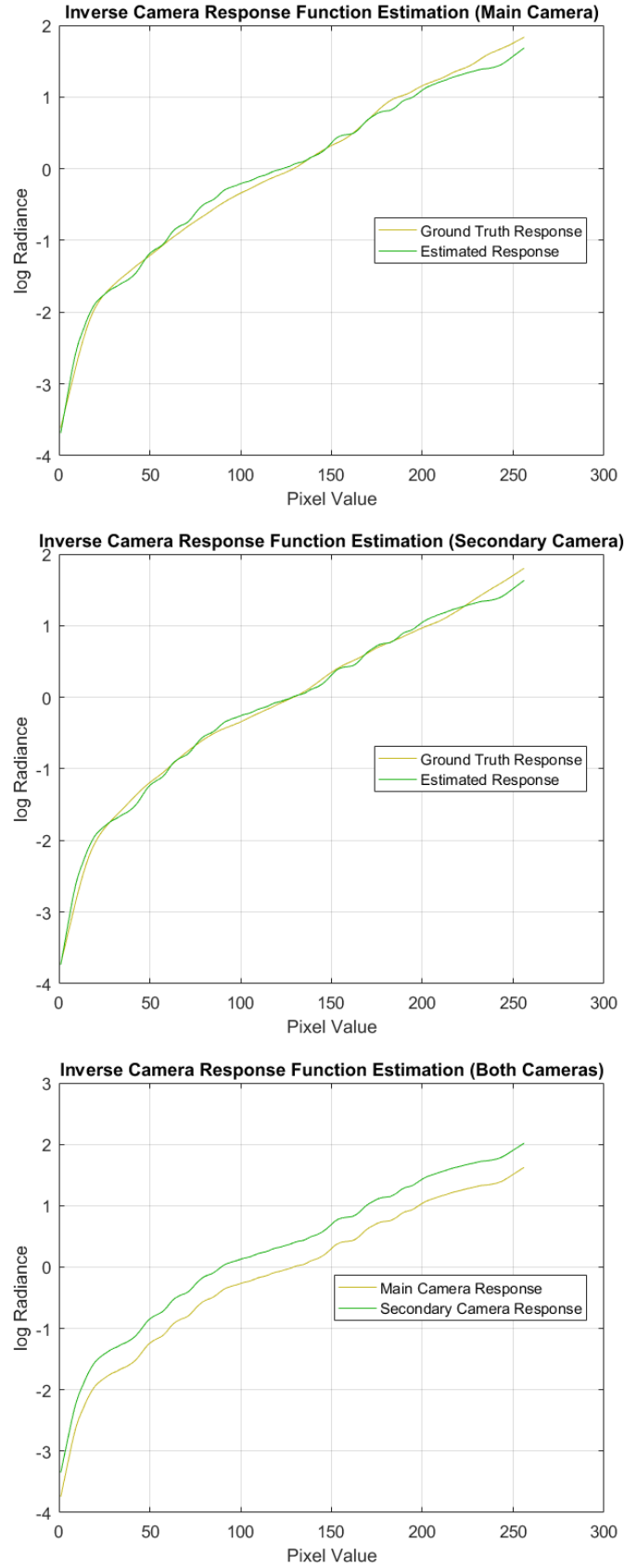


Figure 6.4: This figure describes the accuracy of the ICRF estimation algorithm for the main and secondary cameras, through comparison with ground truth responses. It also shows the main and secondary camera ICRFs on the same plot for comparison.

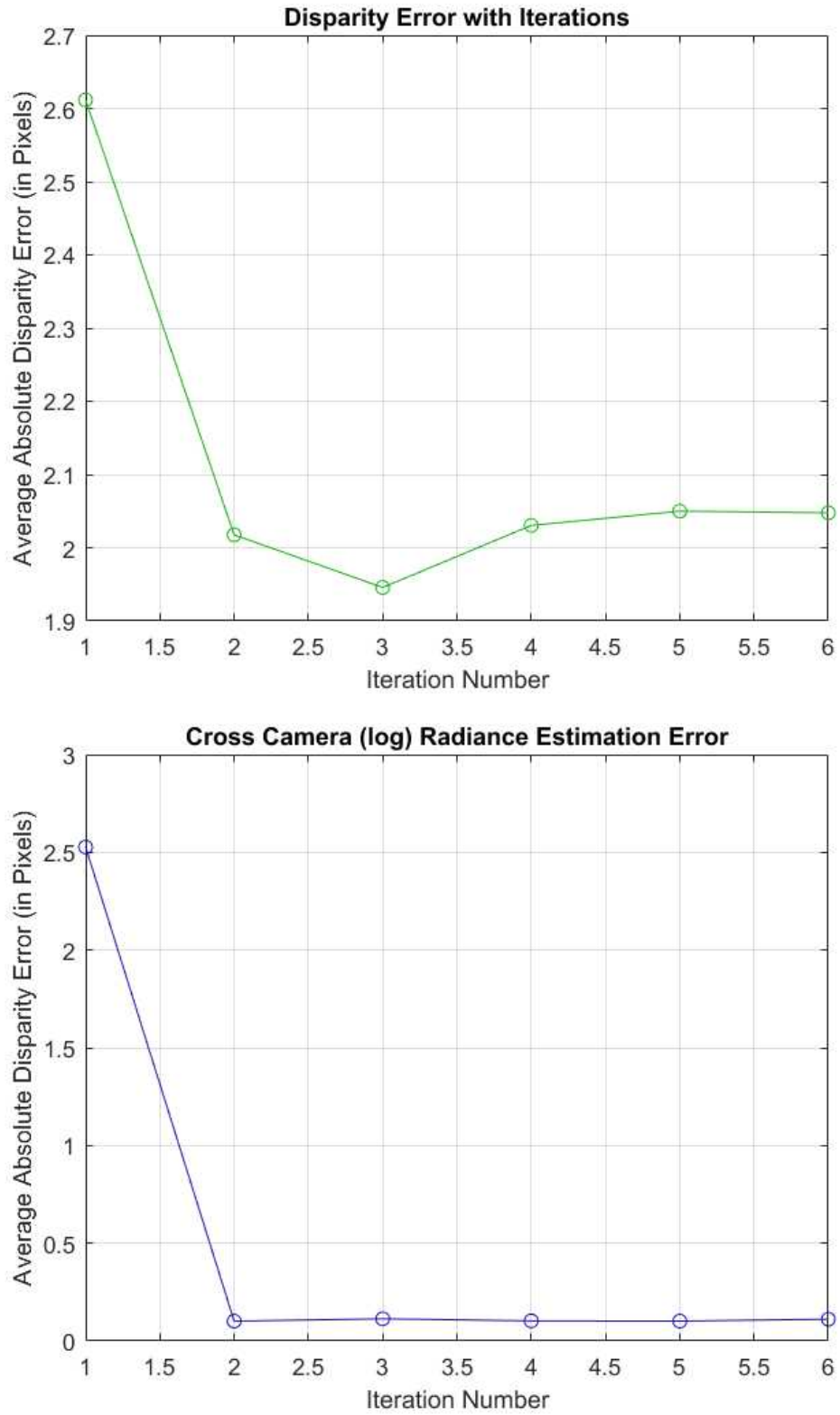


Figure 6.5: This figure describes the effectiveness of the proposed jointly iterative ICRF and disparity estimation framework, in terms of disparity error and cross camera log radiance estimation error.

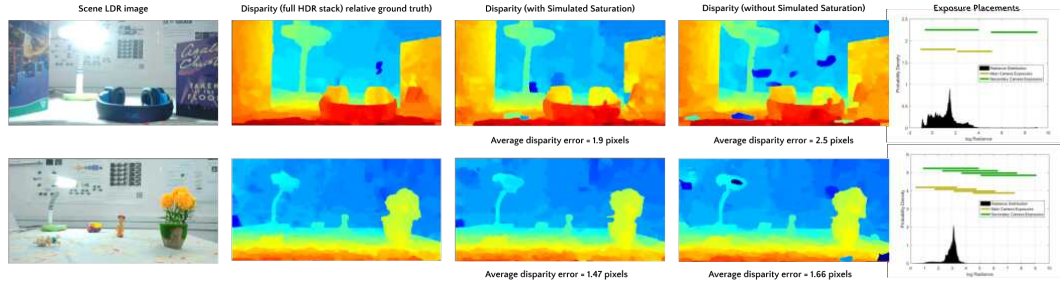


Figure 6.6: This figure describes the performance of the Simulated Saturation step, by comparing the disparity estimation performance with and without simulated saturation for two scenes.

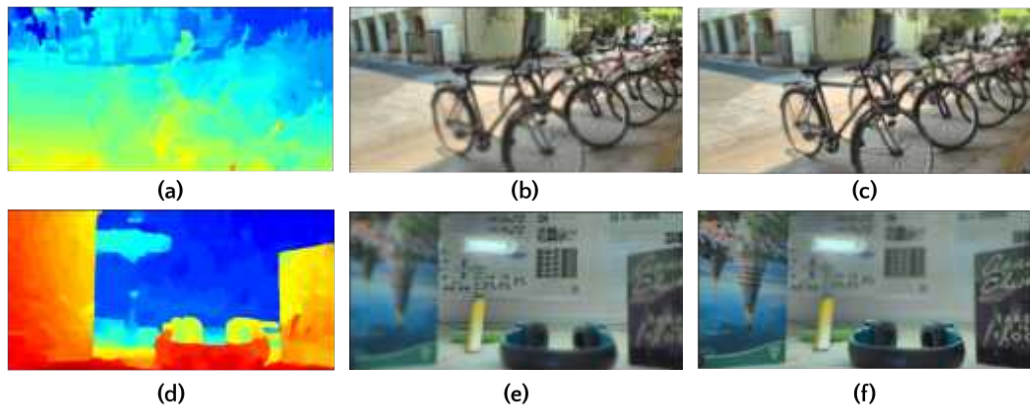


Figure 6.7: This figure describes the performance of the refocusing, performed using the HDR and depth estimates.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

This work establishes an optimal framework for the acquisition of HDR scene information using cell-phone stereo cameras. We propose a novel framework that is modular and flexible in order to accommodate a wide variety of state of the art disparity estimation algorithms, camera setups and imaging noise requirements. The optimality of the framework is established by comparison with other candidate exposure sequences, which shows that the optimal capture sequence leads to better or comparable results as compared to other capture sequences. It was also shown that stereo HDR is more than 20% quicker than Three Shot single camera HDR, which is often traditionally used for commercial HDR photography.

We also validate the various pipeline steps and motivate their necessity in the algorithm. These include the ICRF estimation and the simulated saturation, as well as the benefit of iteratively estimating the ICRF and disparity map for the scene. All these steps are novel proposals as part of this work, and have not been proposed or utilized as part of previous Stereo HDR works.

Finally, we also show some applications for Stereo HDR, such as image refocusing and other possibilities such as perspective shift. Again, to the best of our knowledge, no other Stereo HDR works look to propose and show downstream applications such as HDR refocusing.

Based on the work carried out and possible avenues for improvement, the following may be considered as future directions:

- General two camera ICRF estimation algorithm (without relaxations)
- Theoretical analysis of optimization framework
- Ghosting correction
- Deployment on cell-phone platform

REFERENCES

1. **Bätz, M., T. Richter, J.-U. Garbas, A. Papst, J. Seiler, and A. Kaup** (2014). High dynamic range video reconstruction from a stereo camera setup. *Image Commun.*, **29**(2), 191–202. ISSN 0923-5965. URL <http://dx.doi.org/10.1016/j.image.2013.08.016>.
2. **Boyat, A. and B. Joshi** (2015). A review paper: Noise models in digital image processing. *Signal Image Processing : An International Journal*, **6**.
3. **Debevec, P. and J. Malik**, Recovering high dynamic range radiance maps from photographs. 1997.
4. **Eilertsen, G., J. Kronander, G. Denes, R. Mantiuk, and J. Unger** (2017). Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, **36**(6).
5. **Hasinoff, S. W., F. Durand, and W. T. Freeman** (2010). Noise-optimal capture for high dynamic range photography. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 553–560.
6. **Hirschmüller, H.**, Evaluation of cost functions for stereo matching. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2007.
7. **Hu, J., O. Gallo, K. Pulli, and X. Sun** (2013). Hdr deghosting: How to deal with saturation? *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1063–1170.
8. **Kalantari, N. K. and R. Ramamoorthi** (2017). Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, **36**(4).
9. **Lin, H.-Y. and W.-Z. Chang**, High dynamic range imaging for stereoscopic scene representation. *In Proceedings of the 16th IEEE International Conference on Image Processing, ICIP’09*. IEEE Press, Piscataway, NJ, USA, 2009. ISBN 978-1-4244-5653-6. URL <http://dl.acm.org/citation.cfm?id=1819298.1819909>.
10. **Mozerov, M. and J. Weijer** (2015). Accurate stereo matching by two-step energy minimization. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, **24**.
11. **Park, W.-J., S.-W. Ji, S.-J. Kang, S.-W. Jung, and S.-J. Ko** (2017). Stereo vision-based high dynamic range imaging using differently-exposed image pair. *Sensors*, **17**, 1473.
12. **Pourreza, R. and N. Kehtarnavaz**, Exposure bracketing via automatic exposure selection. 2015.
13. **Robertson, M. A., S. Borman, and R. L. Stevenson**, Dynamic range improvement through multiple exposures. *In In Proc. of the Int. Conf. on Image Processing (ICIP’99)*. IEEE, 1999.