Omnidirectional High Dynamic Range Imaging with a Moving Camera

by

Fanping Zhou

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the M.A.Sc. Degree in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Common cameras with a dynamic range of two orders cannot reproduce typical outdoor scenes with a radiance range of over five orders. Most high dynamic range (HDR) imaging techniques reconstruct the whole dynamic range from exposure bracketed low dynamic range (LDR) images. But the camera must be kept steady with no or small motion, which is not practical in many cases. Thus, we develop a more efficient framework for omnidirectional HDR imaging with a moving camera.

The proposed framework is composed of three major stages: geometric calibration and rotational alignment, multi-view stereo correspondence and HDR composition. First, camera poses are determined and omnidirectional images are rotationally aligned. Second, the aligned images are fed into a spherical vision toolkit to find disparity maps. Third, enhanced disparity maps are used to warp differently exposed neighboring images to a target view and an HDR radiance map is obtained by fusing the registered images in radiance. We develop disparity-based forward and backward image warping algorithms for spherical stereo vision and implement them in GPU. We also explore some techniques for disparity map enhancement including a superpixel technique and a color model for outdoor scenes.

We examine different factors such as exposure increment step size, sequence ordering, and the baseline between views. We demonstrate the success with indoor and outdoor scenes and compare our results with two state-of-the-art HDR imaging methods. The proposed HDR framework allows us to capture HDR radiance maps, disparity maps and an omnidirectional field of view, which has many applications such as HDR view synthesis and virtual navigation.

Acknowledgements

First, I would like to express my sincerest gratitude to my supervisor, Dr. Jochen Lang for his time, advices and encouragement during my research, and invaluable feedbacks and suggestions in writing the thesis.

Second, I would like to thank Dr. Alan Brunton for his generosity in sharing with me his spherical stereo vision framework and his expertise.

Third, I would like to extend my appreciation to Mr. Dave O'Neil, my boss at CBN for his support.

Finally, I would thank my wife Helen and my son Manuel for their support and they are the motivation for pursuing this study.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Notations

2D	Two dimensional
3D	Three dimensional
dB	Decibel, a logarithmic quantity that measures the ratio of a physical quantity to a reference level
CUDA	Compute unified device architecture
E	Irradiance
GPU	Graphics processing unit
HDR	High dynamic range
LDR	Low dynamic range
SDK	Software development kit
SVT	Spherical vision toolkit
X	Exposure: product of irradiance and exposure time
Δt	Exposure time

Chapter 1

Introduction

In the era of digital cameras, it becomes simple to take sharp and well-exposed photographs, with the aid of auto-exposure and auto-focus features commonly available even in most low-end cameras and camera phones. It is still impossible, however, to capture photographs of high contrast scenes free of over- or under-exposed regions, no matter what exposure settings may be selected. Most outdoor scenes have a much wider dynamic range in luminance than what a common camera sensor is capable to cover. As a result, the scene regions with luminance values outside the range of the sensor at a given exposure setting will be clipped out and appear as saturated regions in photographs.

As a branch of computational photography [1], High Dynamic Range Imaging (HDRI) [2] is devoted to develop techniques and algorithms for enhancing the dynamic range of conventional digital photographs. In computer graphics, HDR radiance maps recovered from photographs [3] can help produce a more realistic scene rendering. In addition, HDR radiance maps can be used to recover the reflectance properties of a scene if the geometric model of the scene is known or reconstructed, and the illumination is known [4]. Furthermore, HDR imaging (or video) has also found different applications in other fields such as driver assistance [5] and outdoor navigation of mobile robots [6].

The term ‘dynamic range’ used in photography is referred to as a ratio between the highest and the lowest luminance values [7]. The luminance of the physical world can vary tremendously from below $10^{-4}\ cd/m^2$ under starlight, to above $10^{6}\ cd/m^2$ under sunlight [8] as illustrated in Figure 1.1.

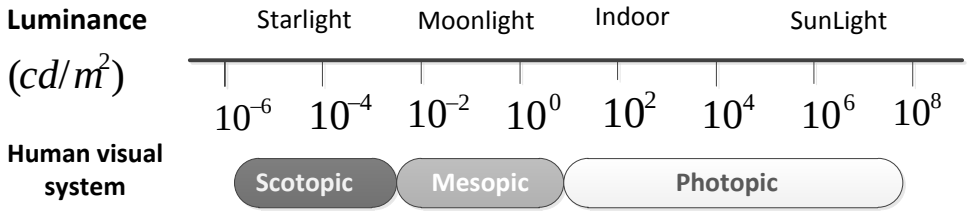


Figure 1.1 Range of luminance in the physical world and the human visual system (HVS) (adapted from [8]).

Although it can operate over only one fraction (~10^5 cd/m^2) of the range at any single time, the human visual system (HVS) is capable of handling almost the whole dynamic range of luminance in the physical world, thanks to its light adaptation.

Compared to HVS, most image capture and display devices have considerably lower dynamic range. The typical dynamic range of common camera sensors is lower than that of traditional films [5] in the range of 2-3 orders. The direct approach for overcoming the limitation in dynamic range is to develop special HDR sensors [5, 9]. HDR sensors, however, have not been widely used in consumer cameras possibly due to high costs.

The most common way of acquiring HDR images is to use cameras with common low dynamic range (LDR) sensors to capture multiple images with different exposures [3]. Using multiple exposures to extend dynamic range can be traced a long way back to the early years of analog photography. In 1850s, Le Gray, a French photographer, created a photograph of the sea scene by combining two negatives of the bright sky and the cloud shaded sea surface taken with different exposures [10].

The pioneering work on HDR imaging for computer graphics may be attributed to Debevec et al. [3] in the late 1990s. They proposed the classic method for deriving the inverse camera response function (radiance to pixel value relationship) and generating HDR images from LDR photographs with different exposures. After a rapid development for a decade or so, HDR imaging techniques have been utilized in some commercial cameras (e.g. Canon Rebel T4i camera [11]). Some high-end HDR cameras combine HDR and panoramic imaging techniques to generate full-view HDR videos [12].

As an HDR image is fused pixel-wise from multiple LDR images, its quality is directly related to the accuracy of pixel registration across different images. Most of HDR imaging techniques are applicable only if the LDR images are captured from a single view, and with no or small camera motion [13]. Very limited research has been conducted on the recovery of HDR from multi-exposed LDR images captured from multiple views.

The classic HDR imaging requires the use of stationary cameras and is neither practical nor efficient. It would be more convenient to use a moving camera to capture LDR images and generate HDR images in applications such as Google Street View [14]. If we capture M images with different exposure times from a single view, we can generate only one HDR image. However, if we capture M images from different views and with different exposure times, it may be possible to generate M multiple HDR images - one for each view.

This thesis explores the feasibility of reconstructing HDR images from LDR images captured from different views using an omnidirectional camera. In the proposed HDR imaging approach, we can obtain wide view (>75% of full view) and depth information (3D information) on top of an extended dynamic range which allows for more applications such as view synthesis, virtual navigation etc.

1.1 Problem Definition

The objective of the thesis is to conduct research into omnidirectional HDR imaging using multi-exposed LDR images captured from a moving camera. The essential problem is the determination of dense stereo correspondences among multi-view and multi-exposed images for image registration before fusing those LDR images into HDR images. Two assumptions are made in our research: global illumination remains constant and no blurring occurs. Constant illumination does hold in most controllable indoor environment but may not be true outdoors. However, it is approximately constant if the capture duration is short. The second assumption is reasonable as the maximum shutter time is set to be less than 1/10 s.

Although a large amount of research has been conducted on stereo matching and well over 100 different algorithms [15, 16] have been proposed, dense stereo correspondence is still considered challenging, due to half-occlusion, poorly textured areas etc. When images are taken with very different exposures, under- or over-exposed regions present in images and brightness differences between image pairs pose additional challenges to most stereo matching algorithms that assume brightness constancy.

Unless image pairs are captured using a stereo camera, rectification is required to align each image pair to the same plane to facilitate stereo matching. If multi-view images are captured with a moving camera, geometric calibration can be carried out to determine camera poses: rotations and translations (up to a scale) which are required in image alignment and stereo matching. Then stereo matching algorithms can be applied to the aligned or rectified images to obtain disparity maps which are used for the registration of multi-view images before fusing them in radiance space to obtain HDR radiance maps.

First, large exposure differences can make pose estimation less accurate. When exposure differences increase, clipped (under- or over-exposed) regions may become so large that both the quantity and the quality of feature points matched between image pairs are affected. Compared to planar images, omnidirectional images with their wider view angles can alleviate this problem.

Second, different exposures and saturated regions can have a great impact on stereo matching. We need to explore possible measures to find good stereo correspondence in the presence of exposure differences and saturated regions. Again, omnidirectional images can facilitate the stereo matching with its full view. However, omnidirectional images limit our choices of stereo matching algorithms as most of those algorithms are developed only for planar images and cannot be applied directly to omnidirectional images.

Third, saturated regions in images with very low and high exposures can result in low quality disparity maps. Since we employ disparity maps for image registration, the quality of the HDR images is dependent on disparity maps. To this end, we need to explore different ways of improving disparity maps, e.g. removing invalid values and filling holes using the prior information about scenes and superpixel segmentation in color images.

Lastly, different disparity-based warping algorithms can be used for the registration of multi-view images in HDR fusion which we assess. Once disparity maps are found and enhanced, two disparity-based warping algorithms: forward and backward warping, can be used to warp images from neighboring views to the target view before generating an HDR image for the target view. We need to explore the impact of forward and backward warping methods on the quality of warped images and fused HDR images.

1.2 Background

The classic HDR imaging technique proposed by Debevec et al. [3] recovers high dynamic range of scene radiance from multi-exposed LDR images. According to Reinhard et al. [2], radiance is radiant power (light) leaving (or arriving) at a point in a particular direction whereas irradiance is radiant power arriving from all possible directions at a point on a surface. Luminance and illuminance are photometrically weighted radiance and irradiance, respectively. In the HDR community, radiance and irradiance are used more often than luminance and illuminance. Since a camera sensor responds linearly to irradiance, an image of irradiance can be recovered from pixel values if camera response curves and the exposure times are known. As the irradiance value recording in each sensor cell is related to the radiance of the corresponding scene point, the image of irradiance is often called radiance map.

Common classic HDR imaging techniques consists of radiance map conversion, HDR fusion and tone mapping. To generate the full dynamic range of radiance in a scene, different exposure times can be used to capture a sequence of LDR images each recording a different band of dynamic range. Those LDR images are converted into LDR radiance maps which can be fused into an HDR radiance map. As the radiance values can have a considerably wider range than what common display devices can display, an HDR radiance map needs to be compressed into a range of 0-255. This compression is called tone-mapping and many different tone mapping operators have been developed [2].

The classic high dynamic range (HDR) imaging techniques may be referred to as radiance space fusion [3] approach. Another simpler approach is direct exposure fusion [17, 18]. The direct exposure fusion approach skips the radiance conversion and tone mapping, and combines HDR contents for each block or pixel from multi-exposed LDR images.

Although the direct exposure fusion approach is simpler and faster, no true HDR radiance map is obtained. In addition, the brightness contrast between different regions in the final fused image may not reflect the true radiance relation of the scene. In other words, brightness inversion may occur. Thus, the radiance space fusion approach has been selected in this thesis as radiance maps would allow for a wide range of applications in computer graphics and for application-specific post-processing.

HDR imaging studies have shifted from classic topics such as radiance recovery and tone mapping [2], to HDR quality improvement, e.g. detection and removal of artifacts resulting from the imperfect registration caused by object or camera movement [13]. If a stationary camera is used to capture the radiance map of a static scene, it is straightforward to create HDR images of high quality using any radiance based algorithm [3]. In practice, we need to deal with both object and camera movements when generating HDR images.

When a stationary camera is used to capture a dynamic scene, the detection and removal of moving objects become very similar to the classic background model and change detection in computer vision [19, 20], although exposure difference and saturation in LDR images may need special attention. A great number of research papers, e.g. [21-37], have been published on the detection and removal of artifacts caused by moving objects and a good review can be found in [13]. If the images of multiple exposures are registered correctly, moving objects can be detected based on the changes in exposure-invariant properties, e.g. entropy [31], or intensity in exposure-normalized images [26] or in radiance maps [31]. Optical flow [28], or statistical model based background modeling [22] may be used to find moving objects. Once the pixels associated with moving objects are identified, the removal of moving objects can be achieved by eliminating those pixels

completely, keeping only one instance of each moving object, or using weight reduction in the final HDR fusion.

In the presence of camera movement, the scene background has to be registered as accurately as possible across the images captured at different camera positions before performing HDR fusion. Optical flow can be used to find dense correspondence to register images if the movement is small. As optical flow requires brightness constancy, multi-exposed images have to be converted into gradient images [28] or exposure-normalized before applying optical flow.

The simplest camera movement may be translations in image plane. In the work of Ward [29], images are converted into bitmap images using the median threshold bitmap (MTB) method and the XOR (eXclusive OR) operator is applied to binary image pairs. XOR differences are computed for all possible row and column translations and the optimum solution is the row and column translations that produce the minimal XOR difference. This technique is very fast and robust against exposure difference. Grosch [30] has extended Ward's MTB approach to deal with both translations and rotation in image plane.

Another simple camera movement is pure rotation about the optical centre and a homography matrix can be used to transform images from one view to another. A homography is also applicable to a planar scene. Tomaszewska and Mantiuk [32] applied homography transformation to the alignment of images captured from different views. A homography may be approximately applicable as the scene captured is far away from the camera.

When multi-exposed LDR images are captured with a freely moving camera, we can resort to stereo matching to find dense correspondence between images. Although both stereo matching and HDR imaging have been very active research areas, there is very limited work on the generation of HDR images using stereo correspondence. Troccoli et al. [33] may be one of the earliest researchers working on the reconstruction of HDR images from multi-view images using stereo matching to find image correspondence. Normalized cross-correlation is used to find the initial matches to determine the transfer function which

is used to convert LDR images into radiance space. Final disparity maps are evaluated from stereo matching in radiance space. Recently, Sun et al. [35] also proposed to find dense correspondence between images using stereo matching. The initial stereo matching is conducted on LDR images using the Adaptive Normalized Cross-Correlation (ANCC) as a cost function. Then the camera response curves determined are used to convert LDR images into radiance maps. The final stereo matching is performed in radiance space on census-filtered images. The images used in their work are mildly different in exposures with small spotty saturation. In the thesis [36] of Rufenacht, stereo matching is performed on the exposure-normalized images to find disparity maps for image registration. Because of large saturated regions and big differences in exposure, the disparity maps found are very noisy. The holes in the disparity maps are filled with iterative propagation of valid disparity values from adjacent pixels. Two ways of HDR imaging are compared: temporal (multi-exposed images captured at different time instances but from the same view), and spatial (multi-exposed images captured from different views simultaneously).

1.3 Thesis Statement

The classic HDR imaging approach requires exposure bracketing of images captured with no or small camera motion but practical HDR applications such as virtual tours and mobile robot navigation necessitate omnidirectional HDR imaging with a moving camera. The proposed disparity-based spatiotemporal HDR framework is more efficient than the classic approach but is still capable of recovering similar dynamic range of radiance from multi-view multi-exposure LDR images. Although the proposed framework seems to produce omnidirectional HDR images of lower quality than does the classic approach, over-segmentation and color information can be used to improve disparity maps and HDR images. The framework can yield more consistent HDR images across different viewpoints for both indoor and outdoor scenarios than do two state-of-the-art HDR imaging methods.

1.4 Thesis Overview

The goal of this thesis is to develop a framework for reconstructing omnidirectional HDR from multi-view LDR images. The Ladybug 2 camera system by Point Grey [38] was used to capture omnidirectional images. Omnidirectional images can be represented in the formats of, e.g. cubic, cylindrical, spherical panorama, or RD-map (rhombic dodecahedron map [39]). The alignment and registration of images captured from different views can be affected by exposure differences and saturated regions. An increasing field of view can offset this impact as more well-exposed pixels are available in omnidirectional images than planar images in most cases.

The pipeline of our omnidirectional HDR imaging framework is shown in Figure 1.2. The main stages are as follows:

- Camera geometric calibration and generation of rotationally aligned omnidirectional LDR images.
- Stereo matching and disparity map creation using the aligned omnidirectional LDR images as input.
- Omnidirectional HDR composition from multi-view and multi-exposed LDR images based on disparity maps.

Figure 1.2 Pipeline of the multi-view omnidirectional HDR imaging

In the first stage, the image data captured using the Ladybug 2 omnidirectional camera system will be used to create cylindrical panoramas. The cylindrical panoramas are used as input to conduct geometric calibration to determine camera poses (rotations and translations). The rotation information is thus used to align images from different views to facilitate spherical stereo matching.

In the second stage, spherical stereo matching is used to determine disparity maps. Spherical Vision Toolkit (SVT) [40] is used to find dense correspondence between the images captured from different views. Different cost functions are compared in term of exposure invariance and different post-processing techniques are used to improve the disparity maps.

In the last stage, we use the disparity-based forward or backward warping algorithms to warp multi-exposed images from neighbouring views to a target view. Then we can use the standard radiance based approach to create HDR images from the warped (registered) images with the original images in radiance space. If disparity maps are of poor quality, especially for outdoor scenes with large poorly textured regions, we explore different disparity enhancement techniques, e.g. removing invalid regions and filling holes using superpixels and color information.

1.5 Contributions

The major contributions are summarized as follows:

- An efficient spatiotemporal HDR imaging framework that can generate omnidirectional HDR images from LDR images captured with different exposures at multiple positions.
- Disparity-based forward and backward image warping algorithms for spherical stereo vision implemented in GPU.
- Evaluation of the temporal and spatiotemporal HDR imaging methods and examination of the different factors such as indoor and outdoor scenes, exposure increment steps, baseline and exposure sequence ordering.
- Use of a superpixel method and color prior information to improve disparity maps for the outdoor scenes with large sky regions.

1.6 Organization of Thesis

The rest of the thesis is organized in the following chapters:

- Chapter 2 presents a review of related work on high dynamic imaging and artifact detection and removal in the presence of object and camera movements, and the recent development in reconstructing HDR from multi-view LDR images with multiple exposures.
- Chapter 3 describes the hardware and software for omnidirectional image data capture, the experimental design and photometric calibration for the determination of inverse camera response curves.
- Chapter 4 presents the multi-scale spherical stereo vision framework and Spherical Vision Toolkit (SVT) developed by Brunton et al [41]. SVT is used to find stereo correspondence between omnidirectional images with different exposures. An evaluation of SVT with differently exposed images is presented to show different stages of the SVT pipeline.
- Chapter 5 describes the framework for generating omnidirectional HDR images from multi-view and multi-exposure images. Both the temporal and the spatiotemporal HDR imaging pipelines are described. Some techniques for disparity map enhancement are discussed. Two image quality metrics are also described.
- Chapter 6 presents an evaluation of the proposed spatiotemporal omnidirectional HDR imaging framework. The different factors are examined including exposure increment step size, baseline, exposure sequence ordering and scene type (indoor and outdoor scenes). A comparison is made between the proposed HDR imaging framework and the state-of-the-art HDR methods proposed by Sen et al. [42] and by Hu et al. [43].
- Chapter 7 summarizes the thesis, draws the conclusions, discusses the limitations of the work in this thesis and makes some recommendations for future work.

Chapter 2

Related Work

Although the research in this thesis involves both high dynamic range (HDR) imaging and stereo matching, the focus is placed on the development of HDR imaging techniques more than stereo matching. Thus, the review will cover mostly high dynamic range (HDR) imaging and touch on stereo matching.

High dynamic range imaging is one of the sub-fields in computational photography which aims to overcome the limitations and expand the capacities of digital cameras. After an outline of computational photography in Section 2.1, different methods for generating HDR images will be introduced in Section 2.2.

When reconstructing HDR images from low dynamic range (LDR) images in the presence of camera or object movement, the challenging problems are registration of LDR images with different exposures and detection and removal of possible artifacts. Section 2.3 will give a review on the research conducted on the generation of artifact-free HDR images in the presence of object and camera movements. Section 2.4 will cover the reconstruction of HDR images from multi-view and multi-exposed LDR images based on stereo matching.

Finally, Section 2.5 will discuss the research on omnidirectional HDR imaging, and Section 2.6 will give a short summary of this chapter.

2.1 Computational Photography

Since the advent of digital cameras, more and more image processing algorithms have been incorporated in camera processing pipelines. Computational photography has evolved into a multi-disciplinary research field that aims to overcome the limitations inherent in consumer cameras such as narrow field of view, limited depth of field and low dynamic range. HDR imaging is one such sub-field in computational photography developed to overcome the limitation in dynamic range.

Although computational photography has been offered as a graduate level course in many world-class academic units and an international conference on computational photography (ICCP) has been taking place annually since 2012, a precise definition of computational photography has not been agreed upon yet. According to Raskar [44], computational photography can be considered as a field that "*combines plentiful computing, digital sensors, modern optics, actuators, probes and smart lights to escape the limitations of traditional film cameras and enables novel imaging applications*."

Enhancing lens, sensor, processor and illumination (e.g. flash) capabilities can find many practical applications. For instance, optics can be modified to capture 4D light fields so that the images captured can be refocused to different depths of field [45]. Field of view can be expanded by stitching and mosaicking images captured by panning a camera [46]. Dynamic range can be extended by merging differently exposed photographs [3]. The interested readers can find a good coverage of computational photography in [1], and the most updated ongoing research in the conference proceedings of the ICCP (International Conference on Computational Photography).

2.2 HDR Imaging Methods

Many methods have been developed to extend the dynamic range of consumer cameras [1]. Those methods may be divided into two broad categories: hardware based and software based. In the former category, special HDR sensors are developed or conventional lenses

and sensors are modified to capture different bands of the dynamic range in one single shot. In the latter category, high dynamic range is recovered from multi-exposed LDR images captured using conventional cameras. Most of the HDR imaging methods fall into the latter category. Compared to hardware based methods, it is more affordable to generate HDR images from LDR images captured with exposure bracketing using consumer cameras. Most digital cameras offer exposure bracketing and some cameras even include limited HDR functions.

Depending on what image space is used, the software based HDR imaging techniques may be further divided two approaches: radiance fusion [7] and direct exposure fusion [18].

2.2.1 Hardware Approach

Since conventional camera sensors have a very limited dynamic range, it is natural to try to increase the dynamic range by developing specific HDR sensors. One type of HDR sensors is the logarithmic sensor. A typical camera sensor responds linearly to irradiance whereas a logarithmic sensor utilizes the exponential relationship between photo-current and voltage and exhibits a logarithmic response to irradiance [47]. Thus, a considerably wider dynamic range can be achieved with a logarithmic sensor than with a linear sensor. Recently, Martinez-Sanchez et al. [48] proposed a novel method to adjust the sensor response quickly between linear and logarithmic accordingly based on the dynamic range of scene radiance. The shortcomings of logarithmic sensors are possible loss of contrast and high noise.

Instead of designing HDR sensors, more researchers choose to modify conventional camera lenses or sensors so that different ranges of scene radiance can be captured in one single shot and assembled into a wide dynamic range in real-time. Here are some of those techniques developed: Multiple Image Detectors [49-51], Multiple Sensor Elements within a Pixel [52, 53], Spatially Varying Pixel Exposures [54] and Pixels with Adaptive Exposure [55].

Multiple Image Detectors. Beam splitters are used to produce multiple copies of scene radiance and each copy is cast onto a different sensor. By adjusting beam angle or using an optical attenuator, the images captured by different sensors with different exposures can be combined in real-time to produce HDR images. The Multiple Image Detectors technique has been applied to the generation of HDR still images by Aggarwal and Ahuja [50], and to HDR videos by Tocci et al. [51]. This approach is able to capture HDR images in real time but requires multiple image sensors and additional optical devices for a precise alignment of multiple images.

Multiple Sensor Elements within a Pixel. In this approach, each sensor cell includes multiple elements so that each scene point recorded by a cell can have different exposures. Multiple exposures are combined on chip to produce HDR images. Wen [52] and Street [53] have used this approach with two sensitive elements in each cell to increase dynamic range. The shortcomings of this approach are an increased complexity in sensor fabrication and a reduction in spatial resolution

Spatially Varying Pixel Exposures. Nayar and Mitsunaga [54] proposed to place an optical mask with spatially varying transmittance onto the image sensor so that pixels are exposed spatially differently. In doing so, an over-exposed pixel can interpolate from the adjacent pixels with lower exposures and an under-exposed pixel can find the radiance information from its adjacent pixels with higher exposures. It is likely that spatial resolution will be affected in this approach.

Pixels with Adaptive Exposure. In the approach of Nayar and Branzoi [55], a real-time control algorithm is employed to adjust the light modulator automatically so that the exposure of each pixel can adapt to the radiance value of the corresponding scene point. The captured image and the corresponding transmittance function are used to generate the final HDR image. The challenges to keep the control algorithm stable against scene changes are lack of a suitable model for scene dynamics and saturation [55].

2.2.2 Radiance Fusion Approach

The radiance fusion approach was proposed by Debevec et al. [3]. The common radiance fusion approach may include the stages of radiance conversion, alignment and registration, HDR composition and tone mapping as shown in Figure 2.1. First, LDP images are converted into radiance maps based on pre-determined inverse camera response curves. Second, if camera or object movement is present, image alignment and registration are performed. Third, an HDR radiance map can be reconstructed as a weighted mean of partial radiance maps from different exposures. Last, a tone mapping operation is used to produce tone mapped images for display because HDR radiance maps cannot be directly displayed.

In the following, a brief review will be given on the common techniques for photometric calibration and camera response curve determination, HDR composition and tone mapping. Image alignment and registration will be covered in some sections later.

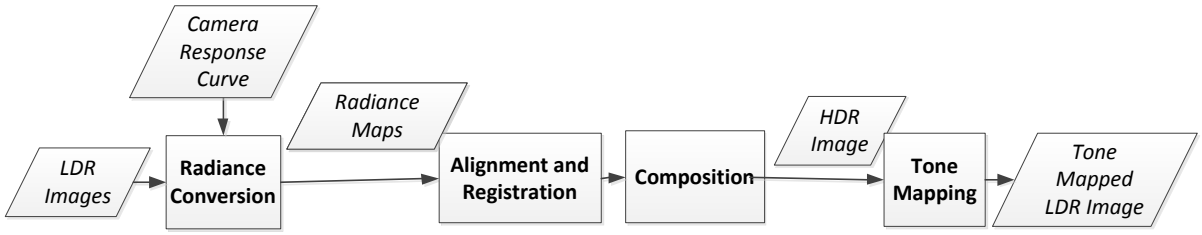


Figure 2.1 HDR imaging pipeline for radiance fusion approach.

Photometric Calibration

To recover the entire dynamic range of scene radiance from LDR images, we need to know the inverse camera response curve (or function) which describes exposure (product of irradiance and exposure time) as a function of pixel value. Although most camera sensors are linear to exposure, image processing algorithms used in the pipeline of a camera may introduce nonlinearity. Thus, pixel values in an image cannot be assumed linear to scene

radiance unless the raw data are captured using a linear sensor and are extracted before going through the camera pipeline.

Different methods have been developed for the determination of inverse camera response curves and may be divided into two groups: parametric and non-parametric. In the former group, the camera response curve (pixel value versus exposure) is assumed to be expressed in some mathematical model, e.g. a polynomial relationship as in the work of Mitsunaga and Nayar [56], and Mann and Picard [57]. In the latter group, the camera response function is not described explicitly in a mathematical expression but is assumed monotonic and smooth, e.g. the methods proposed by Debevec and Malik [3], and by Robertson et al. [58]. If the response curve is monotonic and smooth, the inverse response curve exists and can be derived using some optimization methods described later.

The parametric approach may be too restrictive as certain forms of the response curves are assumed. Therefore, the non-parametric approach is more flexible. In the non-parametric approach, the basic idea behind the methods of Debevec and Malik [3] and Robertson et al. [58] is very similar. Robertson et al. [58] use a Maximum-Likelihood approach and iterative Gauss-Seidel relaxation to minimize the object cost function and to determine the optimum solution for the camera response function. It is claimed that the approach is able to handle low and high light areas properly. However, this approach may not be robust in some lighting conditions as reported by Silk [27].

Taking into account complexity and robustness, we opt to use the simple method of Debevec and Malik [3] which seems able to produce smooth response curves in different lighting conditions. Thus, we will focus on the non-parametric method proposed by Debevec and Malik [3].

In the above methods for the recovery of camera response curves, the reciprocity law is assumed to hold. In other words, the pixel value will be the same if doubling irradiance and halving exposure time. The reciprocity law can be expressed as

$$Z_{ij} = f(E_i * \Delta t_j) \tag{2.1}$$

Where Z_{ij} is the pixel value of pixel *i* with exposure index *j*. E_i is the irradiance falling on pixel *i*, and Δt_j the exposure time associated with exposure index *j*. Since the response curve is assumed smooth and monotonically increasing, the inverse function of the response curve should exist:

$$f^{-1}(Z_{ij)} = E_i * \Delta t_j \tag{2.2}$$

Taking the logarithm of the above equation, we obtain

$$g(Z_{ij}) \equiv lnf^{-1}(Z_{ij}) = lnE_i + ln\Delta t_j \tag{2.3}$$

Given a sequence of LDR images captured with different exposure times and registered perfectly, the inverse response curve can be evaluated using a least-square optimization technique [3]. When collecting sample pixels for the evaluation of the inverse response curve, those pixels with values close to 0 or 255 should not be used. For any given exposure setting, there might be an upper and a lower limit of radiance a camera can capture. Any radiance values under the lower limit will be clipped and mapped to pixel value of 0. Similarly, any radiance values over the upper limit will be mapped to a pixel of 255. Because of noise, pixel values close to 0 or 255 are less reliable for recovering radiance values.

HDR Composition

Once the inverse response curve is known, we can reconstruct the entire dynamic range of scene radiance by converting multi-exposed LDR images to LDR radiance maps which are in turn fused into an HDR radiance map. The radiance of each scene point corresponding to pixel *i* can be evaluated as the weighted mean of all values from different LDR radiance maps for the same pixel:

$$lnE_i = \frac{\sum_{j=1}^{n} w(Z_{ij})(g(Z_{ij}) - ln\Delta t_j)}{\sum_{j=1}^{n} w(Z_{ij})}. \tag{2.4}$$

$w(Z_{ij})$ is the weighting function expressed as:

$$w(z) = \begin{cases} z - Z_{min} & for\ z \leq (Z_{max} + Z_{min})/2 \\ Z_{max} - z & otherwise \end{cases} \quad (2.5)$$

Where Z_{min} and Z_{max} are the minimum and the maximum pixel values. This weighting function has a simple triangular hat shape but other weighting functions may be used, e.g. the Gaussian-like weighting function of Robertson et al. [58].

Tone Mapping

Since an HDR radiance map cannot be displayed directly on an LDR display device, tone mapping operators are needed to compress the HDR radiance map. To produce a pleasant tone mapped image preserving the appearance of HDR images, a large number of tone mapping operators have been proposed [2].

Tone mapping operators can be divided into four categories: global operators (e.g. [59]), local operators (e.g. [59]), gradient domain operators (e.g. gradient domain compression operator of Fattal et al. [60]), and frequency-based operators (e.g. bilateral filtering operator of Durand et al. [61]).

A global operator maps radiance values to pixel values uniquely without considering its neighboring spatial information. On the other hand, a local operator takes into account the local spatial information in addition to some global characteristics. Thus, the same radiance value may be mapped to a different pixel value depending on the pixel location.

Global operators are simple and fast to implement. But they tend to produce tone mapped images with lower local contrast and preserve less local details. Global operators are less effective in compressing dynamic range. As a result, they are suitable only for the scenes with low or medium range of radiance. On the other hand, local operators are more effective in compressing wide dynamic range and still produce pleasing results to human eyes. The major shortcomings with local operators are halo artifacts around the edges of large brightness difference and slow processing (local operators can be 10 times slower than global operators [2]).

The gradient domain compression algorithm [60] compresses high dynamic range by attenuating progressively the magnitudes of gradient field in luminance while maintaining the directions and reconstructing a compressed LDR image from the attenuated gradient field. Attenuating the magnitudes of large gradients seems to compress high dynamic range without the problems found in some tone mapping operators, such as loss of details, halo artifacts and gradient reversals. The bilateral filtering algorithm of Durand et al. [61] decomposes an image into a contrast encoding base layer and a detail layer. The contrast reduction is applied only to the base layer which is extracted using the bilateral filter. Thus, the halo artifacts can be reduced and the operation can be sped up by 2~3 orders [61].

2.2.3 Exposure Fusion Approach

An alternative approach for generating HDR-like images is to combine directly the best parts of multi-exposed LDR images. Two such methods based on direct exposure fusion have been proposed by Goshtasby [17] and Mertens et al. [18]. The exposure fusion approach skips the radiance conversion and the tone mapping stages and fuses LDR images directly into HDR content images. The pipeline of the common exposure fusion HDR approach is illustrated in Figure 2.2.

The exposure fusion approach is simple and fast to produce plausible HDR content images, and thus has gained more popularity recently. The exposure fusion method of Mertens et al. [18] has been applied in the work of Peci and Kautz [23], Zhang and Cham [24], Hu et al. [43]. However, the direct exposure fusion approach may not be suitable for image based rendering applications in computer graphics because the full range of scene

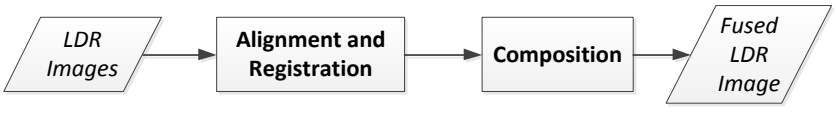


Figure 2.2 HDR imaging pipeline for direct exposure fusion approach.

radiance is required for a realistic rendering. Furthermore, this approach does not allow to post-process radiance maps to fit display devices with different dynamic ranges. Therefore, we prefer to use the radiance space fusion approach in spite of its complexity.

The exposure fusion method of Mertens et al. [18] measures the quality of each pixel in terms of contrast, saturation (standard deviation within each of RGB channels) and well-exposedness. The weight map function is expressed as a product of the quality measures as:

$$W_{i,j} = C_{i,j}^{\omega_c} S_{i,j}^{\omega_s} E_{i,j}^{\omega_E} \tag{2.6}$$

where *C*, *S* and *E* represent the contrast, the saturation and the well-exposedness terms, respectively. ω_c, ω_S and ω_E are the corresponding exponents for the three measures. $W_{i,j}$ is the weight assigned to pixel *i* in the image with exposure index *j*. The contrast is the absolute magnitude of the filtered response after applying a Laplacian filter to the grey image of each exposure. The saturation is calculated as the standard deviation within each RGB channel. The well-exposedness is measured as the closeness to the middle pixel value expressed in a Gaussian function.

The final HDR content image is fused using a Laplacian decomposition of LDR images and a Gaussian pyramid of the weight maps. Although the fused images with HDR contents are comparable to the tone mapped images derived from radiance maps, there is no guarantee that the global contrast of brightness will be maintained.

One sample image using the exposure fusion method [18] is shown in Figure 2.3 and the fused image catches the information from each exposed image of the scene with a great radiance range.

(a) Very low (b) Low

(c) Medium (d) High

(e) Exposure fused HDR image

Figure 2.3 Multiple images from very low to high exposures (a, b, c and d) and HDR image (e) created using the exposure fusion method [18].

2.2.4 HDR Imaging Challenges

When deriving HDR images from an exposure bracketed LDR image sequence, we can face four capture scenarios as shown in Table 2.1 depending on whether the camera is kept stationary or moving, and the scene is static or dynamic. In the ideal case with a stationary camera and a static scene, we can readily assemble an HDR image from multi-exposed images because all pixels across the images can be assumed to register perfectly.

If the camera is kept stationary but the scene is filled with moving objects such as people walking or cars moving, the pixels corresponding to the static background are registered properly but the pixels associated with moving objects are not. The incorrectly registered pixels may appear as ghosting artifacts in the final HDR image. In that case, ghost detection and removal are necessary to produce artifact-free HDR images. Most of the research on the HDR imaging for dynamic scenes falls in this category.

If camera movement also comes into play, we need to perform the registration of both background and moving objects across different images. It becomes very challenging to reconstruct artifact-free HDR images from an image sequence with camera movement.

Different methods have been introduced to generate artifact-free HDR images in the presence of object and camera movements and will be reviewed in Section 2.3. Section 2.4 will review the research on multi-view HDR imaging based on stereo matching which can be used to deal with a large and general camera movement.

Table 2.1 LDR image capture scenarios and problems for HDR reconstruction.

Camera	Scene	Movement	Problem
Stationary	Static	None	None
	Dynamic	Object	Ghosts
Moving	Static	Camera	Misalignment
	Dynamic	Camera & object	Misalignment & ghosts

2.3 HDR Imaging for Dynamic Scenes

As HDR imaging combines the color information from multiple images for each pixel, perfect pixel registration across all images is critical to the quality of fused HDR images. In an ideal case where images are captured of a static scene using a stationary camera, pixel registration is not an issue. In practice, most outdoor scenes are dynamic, and camera shaking or movement can occur easily when capturing an image sequence if the camera is not placed on a steady tripod.

Numerous research papers have been published on the HDR imaging methods for dynamic scenes and an extensive review can be found in [13]. HDR imaging methods for dynamic scenes may be classified into the following three categories:

- *Camera movement only*: application of geometric transformation to align the static background across images. This approach is strictly applicable only to certain type of camera movement and can provide approximated solutions for the cases with small camera movement or remote scenes.
- *Object movement only*: the detection and removal of artifacts caused by moving objects assuming that the background is aligned perfectly;
- *Camera and object movements*: image registration through pixel motion estimation, e.g. optical flow. Both camera and object movements can be handled simultaneously.

2.3.1 Camera Movement Only

When capturing a sequence of images with a hand-held camera, small camera movement is unavoidable. Some researchers have applied simple geometric transformation to the alignment of images with small camera movement or simple movement such as pure rotation or in-plane translation. For a general (3D) camera movement, however, multi-view

geometry and stereo correspondence should be used to register images and the related HDR imaging work dealing with a general camera movement will be reviewed in Section 2.4.

To determine the shifts between images caused by small camera translations, Ward [29] suggested the Median Threshold Bitmap (MTB) method that seems to be robust against exposure differences. MTB is a binary image that can be derived from an input gray image with the following procedure [29]:

- Find the median pixel from the eight neighboring pixels in 3x3 window for each pixel in the input image and create a median image with the median pixels,
- Generate a binary MTB image with 1s for the pixels whose intensity in the input image is greater than that in the median image; 0s otherwise.

To align two images, every possible offset along row and column directions can be tested by evaluating the XOR difference between the corresponding two MTB images. The coordinates that align up two images can be determined from the offset pair that minimizes the XOR difference. An MTB pyramid can be utilized to speed up the process. In addition, the MTB approach seems to be more robust than the gradient field, as it uses relative ranks.

MTB has also been extended to handle small image rotations by Grosch [30], and Jacobs et al. [31]. In Grosch's work [30], the optimum translations and rotation between two images are determined in two steps. First, perform a translational alignment only to get the initial estimation of an optimum match location. Second, try both translational and rotational alignment to find both the translations and orientation for the alignment of two images by minimizing the XOR difference. Although this approach may get trapped in a local minimum in the first translational alignment step, the alignment of sample images seems to produce good results. MTB may still work well in the presence of small moving objects. But it may fail if moving objects affect large image regions.

Tomaszewska and Mantiuk [32] proposed the use of a homography transformation for image alignment. The homography matrix is computed using the direct linear transform (DLT) [62]. A modified SIFT (Scale Invariant Feature Transform) algorithm [64], one of

the most robust feature descriptors, is used to extract matching key points. Furthermore, RANSAC (RANdom SAmple Consensus [65]) is used to find valid matching points from noise feature points owing to different exposures and view differences. The proposed algorithm seems to align the sample images from different views well. Strictly speaking, a homography can be applicable only to the images captured with a camera rotated about its nodal point, or to images of a planar scene [62]. Perhaps the handheld camera only rotates or the scene is far away from the camera.

In theory, a single geometric transformation is applicable only if the camera movement is a pure rotation about its nodal point or the scene is a plane. In those two cases, a single homography [63] can be found between an image pair. In most practical cases where camera movement is very small or the scene is sufficiently far away, a homography may be approximately applicable. As we are interested in HDR imaging with a moving camera, neither homography nor MTB is applicable.

2.3.2 Object Movement Only

When a sequence of images are captured of a scene with moving objects, the objects may be present in different images at different locations. If assembling the images directly into an HDR image as a weighted mean of the images, the moving objects will appear at multiple locations in the fused HDR image in a faded version, which is referred to as ‘ghosting artifacts’ or simply ‘ghosts’. Ghosting artifacts have a great adverse impact on the quality of HDR images and many techniques have been developed for the detection and removal of ghosts in HDR images.

Ghosts may be further divided into discrete and fluid as suggested by Silk and Lang [26]. The discrete ghosts are related to large motion caused by people or cars moving whereas fluid ghosts are associated with small motion or fluctuation such as tree leaves weaving.

The removal of ghosts in HDR composition may be done in one of the following three ways:

- Disregard all the pixels associated a moving object,
- Keep one single occurrence of each moving object, or
- Reduce weights for the pixels associated a moving object.

Keeping one single instance of a moving object may be preferred when capturing holiday photographs whereas eliminating all the pixels linked to people or cars may be required for street viewing and virtual navigation.

The approaches for resolving the problems with moving objects may be loosely classified into the two groups:

- *Scene difference based*: detect moving regions by measuring the differences in some properties between image sequences, e.g. change in radiance values proposed by Jacobs et al. [31].
- *Background modelling based*: use some statistical method to model background, e.g. non-parameter model proposed by Khan [22].

Scene Difference Based

The detection of moving objects has been subjected to a lot of research in computer vision applications such as traffic monitoring [19]. If the background is assumed static, moving objects can be detected by assessing the changes in intensity or other properties. Similar approaches for detecting moving objects have been applied to the HDR imaging for dynamic scenes although exposure differences and saturated regions require special attention. One option is to evaluate the changes in radiance maps or intensity in exposure-normalized images. Another option is to use an exposure invariant property such as entropy [31], gradient orientation [24] , and median threshold bitmap [23].

In the Variance Image approach proposed by Jacobs et al. [31], the variance is evaluated for each pixel from the radiance maps converted from the corresponding LDR images. The Variance Image (VI) is defined as

$$VI(i) = \frac{\sum_{j=0}^{N-1} W(i,j)E(i,j) \,/\, \sum_{j=0}^{N-1} W(i,j)}{\left(\sum_{j=0}^{N-1} W(i,j)E(i,j)\right)^2 / \left(\sum_{j=0}^{N-1} W(i,j)\right)^2} \tag{2.7}$$

Where i is the pixel index, j the image index, and N the total number of images in the exposure sequence. $W(i,j)$ is the weighting function for well-exposedness and $E(i,j)$ the radiance. If the images are aligned accurately, the regions associated with moving objects can be derived from the pixels with the variance over certain threshold.

Entropy is a statistical measure widely used in information theory. The entropy of an image remains the same if the intensity values of the image are scaled, provided that no pixels are scaled out of the pixel value limits so as to be clipped. The entropy can be expressed as

$$H(X) = -\sum_{x} P(X = x)\log\big(P(X = x)\big) \tag{2.8}$$

Where X is an intensity value of a pixel and $P(X = x)$ the probability of a pixel with intensity value of x. A local entropy for each pixel can be evaluated from the intensity histogram of a window centered about the pixel. The local entropy of the same window size should be the same for different exposures and a change in the local entropy implies a possible moving pixel. The Uncertainty Image (UI) is derived from the weighted difference in local entropy for all pixels and is used to find movement clusters. The selection of proper window size may well depend on the image scenes. If the window size is too small, local entropy may be more prone to noise and outliers, and thus is less reliable. On the other hand, using too big a window size would make it difficult to detect small objects. Another minor issue is that the entropy remains the same if the pixels are relocated in the same window.

In the HDR image composition, the weighted means of radiance values are calculated from all radiance maps for all pixels in a motion-free region. For a region with moving objects, radiance values for the whole region are taken from one single exposure to be coherent. Weighted blending along the border of the region is used to reduce possible artifacts.

Compared to the Variance Image approach, the Uncertainty Image approach can be applied directly in LDR images though with more computation. In addition, the Uncertainty Image seems to fail to detect object moving in bright or dark regions [31]. The reason may be that entropy is no longer a good indicator if some pixels become saturated.

Peci and Kautz developed the Bitmap Movement Detection (BMD) algorithm [23] for motion detection based on MTB. As the bit information for a motion-free pixel should remain the same between all MTB images, the sum of *N* MTB images should be *0* or *N* for static pixels. A motion map can be generated using *1*s if the sum is neither *0* nor *N* and *0*s otherwise. After performing some morphological processing to cluster regions and remove noise, the motion map is used to guide the HDR generation based on the exposure fusion method [18]. For a motion affected cluster, the best single exposure is selected based on the average well-exposedness of all pixels in the cluster. As the motion map may be sensitive to noise and morphological processing, possible artifacts can occur if moving objects are not well-defined.

Based on the work of Zhang and Cham [24], the orientation information of the gradient field can serve as a consistent measure for detecting object motion owing to its invariance to exposure differences. The score measuring orientation difference is expressed as exponential function:

$$S^i(x,y) = \sum_{j=1}^{N} \exp\left(\frac{-d_{ij}(x,y)^2}{2\sigma_s^2}\right) \tag{2.9}$$

Where i and j are the image indices in the image stack of size N, x and y are pixel coordinates, and σ_s is a constant. $d_{ij}(x,y)$ is the orientation difference defined as

$$d_{ij}(x,y) = \frac{\sum_{k=-m}^{m} \left| \theta^i(x+k, y+k) - \theta^j(x+k, y+k) \right|}{(2m+1)} \quad (2.10)$$

where *i* and *j* are the image indices. The window for the evaluation of the orientation measure is (*2m+1*) by (*2m+1*).

The weighting map is directly related the consistency map which is defined as

$$C^i(x,y) = \frac{S^i(x,y) \times \alpha^i(x,y)}{\sum_{j=1}^{N} S^j(x,y) \times \alpha^j(x,y)} \quad (2.11)$$

Where $\alpha^i(x,y)$ is *0* if the pixel is under- or over-exposed; and *1* otherwise. The moving pixels are not completely removed but their weights are reduced in the HDR fusion. This approach does not produce severe artifacts along the borders of moving objects because of fading weights. On the other hand, ghosting effects may still be present if the number of images is not large enough and moving objects are present in multiple images. This approach does not seem to be effective in coping with 'fluid' ghosts such as tree leaves weaving.

The images captured of the same static scene with different exposure times should appear identical except for over- and under-exposed regions if they are re-exposed using the camera response curves. In Grosch's paper [30], object motion is detected using the color difference between the re-exposed image pair. If the exposure time is Δt_j for Image *j* and Δt_k for Image *k*, a predicted image can be derived by re-exposing Image *j* using exposure time of Δt_k using the response curve as:

$$\hat{Z}_{i,k} = f(\ (f^{-1}(Z_{i,j})/\Delta t_j) \times \Delta t_k) \quad (2.12)$$

Where *i* is the pixel index, *j* the image index, and *f* the camera response function. A pixel is considered invalid (i.e. in motion) if the difference between the original image and the predicted image at exposure time of Δt_k is over certain threshold ϵ:

$$\left| \hat{Z}_{i,k} - Z_{i,k} \right| > \epsilon \quad (2.13)$$

The invalid pixels are used to create the error map. In HDR fusion, the radiance values of the pixels in a valid region are as usual calculated as the weighted means of all exposures. The radiance values for an invalid region are directly copied from a single exposure to maintain the coherence in moving regions.

The ordering information in the exposure sequence has also been used to detect moving objects by Sidibe et al. [70]. If we capture two images of the same static scene with one short and one long exposure times, the pixel values in the image with the long exposure time should be greater than or equal to the corresponding ones in the short exposure image. But if an object brighter than the background moves from one position at the first capture to another position at the second capture, the pixels associated with the bright object in the mage with the short exposure time could be darker than those in the image with high exposure time when associated with the darker background. Thus, this ordering relation can be used for the detection of moving objects.

Silk and Lang [26] have made use of both the changes in irradiance and in exposure ordering relation for the detection of object movement. For discrete motion like people or cars moving, ghost removal is done by reducing the weights based on the change maps. For fluid motion, iterative selection of the optimum exposure patches is performed for a consistent HDR composition.

Background Modelling Based

Background modeling is another common way for object detection in traffic monitoring and video surveillance [71]. Some statistical models are normally used to evaluate the probability of a pixel belonging to the background, and an optimization is usually required. A few attempts have also been made to extend background modeling techniques to HDR imaging for dynamic scenes.

A non-parametric model of background was suggested by Khan et al. [22]. The basic assumption is that the number of pixels belonging to the background is far greater than that

to moving objects. The neighborhood of a pixel should be a reasonable representation of the background. The weight contributing to the final HDR fusion is directly related to the probability of the membership of a pixel to its neighborhood. A feature vector for each pixel contains three dimensions for color and two dimensions for location. If the feature vector lies in a more densely populated region in the feature space, the probability of its membership to the background will be high. Iterative process is used to find the optimum solution.

Granados et al. [72] approached the background modeling as an energy-based labeling problem. The energy function consists of a data term, a smoothness term and a hard constraint term:

$$E(f_p) = \sum D_p(f_p) + \sum V_{p,q}(f_p, f_q) + \sum H_{p,q}(f_p, f_q) \qquad (2.14)$$

The data term penalizes the pixels with low probability along the image set, and discourages motion. The smoothness term penalizes the intensity differences between the adjacent pixels having different labels. The hard constraint term ensures local consistency so that moving objects are either included or removed completely. In the HDR composition stage, the intermediate radiance map estimated for the background is used to compare to each radiance map in the image stack. The averaging processing in HDR fusion includes a pixel only if the difference between its value and the value of the corresponding pixel in the background radiance map is less than a given threshold. The pixels associated with moving objects are disregarded.

Both the scene difference and the background based techniques are applicable to images captured with stationary cameras of scenes in the presence of moving objects. Although this thesis focuses more on generating HDR images of static scenes using a moving camera, the techniques for detecting and removing moving objects from fused HDR images may be added to the proposed spatiotemporal HDR framework. For instance, the ghost removal method for HDR images developed by Silk et al. [26] can be applied to the images warped to a common target view based on depth maps.

2.3.3 Camera and Object Movements

Some researchers have attempted to develop more general algorithms that can handle both object and camera movements simultaneously. We may classify those algorithms into two classes:

- *Motion estimation based*: find motion vectors for pixels between sequence image pair e.g. using optical flow (e.g. [28]) or block matching [25];
- *Reference alignment based:* select a reference image and align other images to the reference (though different exposure) (e.g. [42]).

Motion Estimation Based

Optical flow can be used to evaluate small movement across a sequence of images. The basic assumptions are small motion and brightness constancy. For large motion (displacement more than one pixel), multi-scale (coarse-to-fine resolution) approach can be used to estimate displacements iteratively. However, this approach can still fail to estimate large motion of small objects. The assumption of brightness constancy is violated across differently exposed images. But if the image is converted into a gradient image, the brightness (or magnitude) constancy across the gradient images derived from differently exposed images may hold. Thus, optical flow can be applied to gradient images.

Kang et al. [34] developed the framework for generating HDR videos using a hierarchal warping for the global registration and optical flow for the local registration. The LDR video stream consists of alternating short (S) and long (L) exposure frames and the neighboring three frames with two different exposures are used to generate an HDR frame for each viewpoint. There are two possible three-frame sequences: short (S^-), long (L) and short (S^+), and long (L^-), short (S), long (L^+). We use the short (S^-), long (L) and short (S^+) sequence to show the HDR imaging process. For a sequence of frames with a short (S^-), a long (L) and a short (S^+) exposures, optical flow is used to estimate the mapping between the previous frame (S^-) and the current frame (L) and create a unidirectional

warped image (S_U^-), and similarly a unidirectional warped image (S_U^+) from the next frame. Bidirectional warped images (S_B^-) and (S_B^+) are derived by combining the homography-based registration between the previous and the next frame and the optical flow based registration. The warped images and the current frame are converted into radiance maps as $\hat{S}_U^-$, $\hat{S}_U^+$, $\hat{S}_B^-$, $\hat{S}_B^+$, and $\hat{L}$. HDR images are assembled for each pixel in the following way:

- If the pixel is well-exposed in L, the radiance value of this pixel is assigned to the weighted mean of the radiance values of the corresponding pixel in $\hat{S}_U^-$, $\hat{S}_U^+$ and $\hat{L}$;
- If the pixel is over- or under-exposed in L, the radiance value of this pixel in the HDR image is set to the value of the corresponding pixel in $\hat{S}_B^-$ if the pixel is well-exposed in S^-, or to the value of the corresponding pixel in $\hat{S}_B^+$ if the pixel is not well-exposed in S^-.

The method of Kang et al. [34] seems to generate HDR videos that can handle both camera and object movement although fast moving objects still cause artifacts.

The energy-based optical flow algorithm proposed by Zimmer et al. [28] is able to achieve a sub-pixel precision for image registration and thus can be used to enhance both dynamic range and resolution from a sequence of multi-exposed images captured with a handheld camera. The energy function is expressed as

$$E(u_k, v_k) = \sum \big(D(u_k, v_k) + \alpha S(\nabla u_k, \nabla v_k) \big) \tag{2.15}$$

Where $D(u_k, v_k)$ is the data term, $S(\nabla u_k, \nabla v_k)$ is the smooth term and α is a constant. The data term models how well the displacements match the given images whereas the smoothness term enforces the smooth displacement field. Because of exposure difference across images, the data term is based on gradient constancy rather than brightness constancy. The smoothness term uses a Total Variation regularizing term. As usual, a coarse-to-fine multi-scale approach is used to accommodate large displacements.

The experiments show pleasing results in the presence of both camera movement and moving clouds. However, the approach fails to handle small objects with large movements. This is also confirmed by Hu et al. [43] and by Silk [27].

In addition to optical flow, the bi-directional motion estimation technique used for H264 [73] has also been applied to image registration for the creation of ghost free HDR video by Mangiat and Gibson [25]. Saturated holes are filled using the motion vectors from adjacent blocks. It is assumed that the displacements are small between adjacent frames but the search region based on block matching can be increased to allow for the capture of fast moving objects. A cross-bilateral filter has been used to reduce block artifacts and misregistration.

Reference Alignment Based

In this category, a good reference image is first selected and then all the rest images in the exposure stack are aligned with the reference image.

In the work of Gallo et al. [21], the images of different exposures are all divided into the same number of non-overlapping regions of equal size (*40*x*40 pixels*) and HDR images are generated by assembling all fused patches each of which is a weighted mean of the consistent patches from different LDR radiance maps in the same region. Gallo et al. [21] suggests assessing the linear relationship between the logarithmic exposures to detect possible moving regions. If two images of the same scene are captured with different exposure times, the plot of exposure ($X_{i,j} = E_i \times \Delta t_j$) values between Image j and Image k should be a straight line with the slope of *1* as

$$\ln(X_{i,j}) = \ln(X_{i,k}) + \ln\left({\Delta t_j}/{\Delta t_k}\right) \tag{2.16}$$

For a dynamic region with moving objects, the radiance (E_i) values are not the same for the corresponding pixel pairs between two images and the log-exposure plot will deviate from being linear.

Each patch (a square region) from any non-reference frame will be tested for the consistence with the patch in the reference using the above linear relationship. A fused patch for each region can be calculated as a weighted mean of the reference patch and all the other consistent patches in radiance space. The HDR radiance map can be obtained by assembling the fused patches for all regions. Since the region-wise assembly can exhibit blocking artifacts along region borders, a Poisson blending has been used to reduce possible artifacts. As the patches are selected from different images, there is also a risk of duplicating a moving object fully or partially in different regions in final HDR images.

The other reference alignment based approaches include the PatchMatch based methods of Sen et al. [42] and Hu et al. [43]. PatchMatch [66] is a fast method for the reconstruction of an image using only the pixels from another different image based on random sampling and optimal sample propagation. The approaches of Sen et al. [42] and Hu et al. [43] perform both image alignment and HDR reconstruction simultaneously and iteratively.

The HDR reconstruction approach proposed by Sen et al. [42] is to start with a good reference image, and combine with the information, that is consistent to the reference image, from all rest images in the exposure stack. This approach also utilizes the bi-directional similarity principle [67]: completeness and coherence. The completeness term requires that all radiance information in the LDR images is captured in the final HDR as much as possible. The coherence term specifies that the final HDR image contains the information only from the existing LDR images, i.e. with as little artifacts as possible. The optimization energy function enforces that the final HDR image is similar to the reference image for the well-exposed pixels and extracts the information from another image if the region is poorly exposed in the reference image. A multi-scale and iterative approach is used to find the optimal HDR image that minimizes the energy function.

(a) Input images

(b) Aligned images

Figure 2.4 Input images and aligned images generated from exposure stack using the approach by Sen et al. [42].

The source code for the algorithm is provided by Sen et al. [42]. The result of applying it to one sample dataset is shown in Figure 2.4. The images are aligned well to the middle reference image.

Hu et al. [43] has also applied the Patch-Match method [66] to generate a latent image for each exposure which is similar to the reference image geometrically and has the same dynamic range to the image with the current exposure. The latent image retains the information for the well-exposed region of a given exposure if this region is saturated in the reference image. Then an HDR image can be generated as the weighted mean of all latent images using the exposure fusion approach proposed by Mertens et al. [18]. Provided that moving objects are in the well-exposed region in the reference, the ghosting artifacts caused by moving objects can be prevented in the final HDR images.

The quality of the fused images depend very much on the quality of the reference images as can be seen by comparing Figure 2.5 and Figure 2.6. The low and high exposed images are the same but the middle reference images are different in both figures. As can been seen in Figure 2.6, when the reference image selected contains a large saturated region, the latent image with the low exposure (the left bottom image) fails to extract the information for the sky region from the low exposed image. Rather, the saturated regions are padded with gray colors. Similarly, the quality of the final fused HDR images is much reduced compared to Figure 2.5.

(a) Input Images

(b) Latent images (c) HDR image

Figure 2.5 Input images, latent images and HDR image generated using Hu et al.'s algorithm [43]

(a) Input images

(b) Latent images (c) HDR image

Figure 2.6 Sample with bad reference image using Hu et al. algorithm [43]. The latent image at the low exposure fails to extract the information from the cloud region.

Motion estimation methods such as optical flow or block matching cannot deal with large motion. As the PatchMatch based methods of Sen et al. [42] and Hu et al. [43] rely on a selected reference to align multi-view images, HDR images vary with the quality of the reference image. We would like to develop a multi-view HDR imaging framework that can produce HDR images of consistent quality across different viewpoints by extracting the radiance information from neighboring images.

2.4 Multi-view HDR Imaging

When capturing images of complex scenes from different views, a homography transformation is not adequate. Two-view or multi-view geometry should be used to find dense correspondence between images. A brief explanation will be given on some basics of stereo matching and then a review of the research will be made on the HDR imaging based on stereo correspondence.

2.4.1 Stereo Correspondence

Stereo vision is one of the research fields in computer vision [63]. Stereo matching can be used to determine the dense correspondence using the epipolar geometry. Well over 100 different algorithms for stereo matching have been proposed [75]. Scharstein and Szeliski [15] provided a taxonomy of stereo matching algorithms, a framework and test bed for quantitative evaluation of algorithms. The test benchmark and Middlebury test set can be found in [16]. Most stereo algorithms performs the following four steps (not necessary in the same order):

- Match cost computation;
- Cost (support) aggregation;
- Disparity computation/optimization;
- Disparity refinement.

Stereo algorithms may be classified in two broad classes: local and global. The local algorithms compute the disparity of a pixel within a finite window, whereas the global approaches determine disparity by minimizing a global cost function with a data term and a smoothness term.

Some constraints commonly used in stereo matching include: epipolar constraint, photo consistency, smoothness, uniqueness, ordering etc. But some of those constraints do not

hold all the time, e.g. photo consistency, ordering. The challenges faced in practical matching may include specular reflection from non-Labertian surfaces, poorly textured regions, half-occlusion etc.

The most common match costs can be evaluated using SAD (sum of absolute difference) or SSD (sum of squared difference) of pixel intensities for radiometrically similar images. To cope with possible radiometric differences caused by camera settings and illumination variation, more robust cost functions can be used, e.g. mutual information [77], normalized cross-correlation (NCC), LoG (Laplacian of Gaussian), rank filter or census filter methods [78]. According to Hirschmuller, census filter and mutual information may be the top performers in the presence of radiometric difference.

2.4.2 HDR Recovered from Multi-view Images

Only a limited number of researchers have attempted to employ stereo correspondence to the reconstruction of HDR images from multi-exposed LDR images captured from different viewpoints. Although it is useful to be able to obtain both HDR and depth information simultaneously, saturated regions can be very challenging for stereo matching in addition to exposure differences. The dilemma is to balance dynamic range and stereo matching: high dynamic range requires big difference in exposures whereas good stereo matching needs small difference in exposures.

The method propose by Troccoli et al. [33] may be one of the early attempts to make use of multi-view geometry for multi-view HDR imaging. Both (camera) radiometric response curves and relative exposures are recovered. If the camera response curve is a gamma curve, the normalized cross-correlation (NCC) is proved to be exposure-invariant. Their approach is composed of three steps. First, a reliable correspondence is found by applying multi-view stereo with the exposure invariant cost function of NCC. Then the camera response curve is derived based on the correspondence. Finally, after LDR images are converted into radiance maps using the response curve, multi-view stereo matching is performed in radiance space to determine disparity maps. Then disparity maps are used to

warp all images to the reference view and create HDR images. The evaluation of their approach conducted using synthetic data, and real data [16] seems to produce both accurate response curves and HDR images of good quality. As the test data do not have large saturated regions and the views of camera differ only slightly, it is unknown how their approach will perform on the data with large exposure and view differences.

Another attempt for recovering HDR and disparity maps was made by Lin and Chang [79]. Their algorithm consists of the following steps: derivation of the camera response curves, image normalization and correspondence matching, and ghost removal and tone-mapping. First, SIFT is used to find matching key points for the derivation of the camera response curves. Second, the response curves (in each of RGB channels) are used to normalize the stereo image pair before the stereo matching, based on the belief propagation and GC (graph cut) approach, is performed to find dense disparity maps. Finally, HDR images are generated using dense disparity maps. Before fusing radiance values from different images, the difference in radiance values is checked and used to detect and remove ghosting pixels from the HDR images. Although their algorithm is tested positively using the test data from Middlebury [16], the samples selected are limited with very small exposure and view differences. Furthermore, it is not clear how good matching points were selected from noisy matching data using SIFT that may affect the accuracy of the derived response curves.

A more recent research by Sun et al. [35] introduces an HDR imaging algorithm from two-view stereo matching. Their algorithm also includes three steps. First, perform stereo matching using LDR images based on GC (graph cut) using adaptive normalized cross-correlation (ANCC) to find initial disparity map. Second, derive the camera response curves using the initial disparity map. Finally, stereo matching based GC using census filter in radiance space to determine the final disparity map which is used to guide the assembly of HDR image. The test samples selected are also from Middlebury [16] but with exposure differences greater than the samples used in [79], although the view difference is still very limited. The percentage of invalid pixels in the final disparity map is less than 10% for

different sets of the samples. The algorithm seems to work well in spite of the presence of saturated regions.

In his Master thesis [36], Rufenacht presented two methods for recording stereoscopic HDR images using two parallel cameras with a fixed distance: temporal and spatial. In the temporal HDR approach, both cameras capture images using the same exposures alternating between the short and long exposure times. In the spatial HDR approach, one camera always use the short exposure time whereas the other uses the long exposure time. Before stereo matching, the image with high exposure is re-exposed using the short exposure time. The matching algorithm is block-based approach, similar to local algorithms and SAD is used as the cost function. Image segmentation is used to refine disparity maps.

The half-clipping concept was introduced for the region clipped in one image but well exposed in the other of a stereo pair. To fill in the depth information, iterative depth propagation (IDP) is used to expand the valid depth information from the neighboring pixels around the boundaries of the saturated regions. More advanced inpainting methods used for depth map and images [80] may be useful to fill the saturated regions than IDP.

The spatial approach allows for a higher HDR frame rate but a lower HDR quality as the image registration between two images captured from different cameras is not perfect. Some artifacts are also visible along the edges where half-occlusions occur as no explicit cross-check was carried out to remove the errors caused by half-occlusion.

Akhavan et al. [37] presented an HDR stereo matching framework with some preliminary results. Three approaches for creating disparity maps are stereo matching using HDR image pairs, using tone-mapped LDR image pairs, and using LDR image pairs. The disparity map based on the tone-mapped LDR image pair indicates more details than that derived from single optimum exposed image pair.

Based on the review above, the HDR imaging from two-view or multi-view images with multiple exposures has been limited to indoor data with small exposure gaps and with small or no significant saturated regions. In addition, two views would not allow to extract

reliable information using depth maps if a regions is saturated. Thus, we would like to extend the HDR imaging to multiple views. For a saturated region in the current view, we can find depth maps between images from other views in which the region is well-exposed. Thus depth maps from the other views can be obtained and used to extract the radiance information from the other views to the current view.

All the multi-view HDR imaging methods can generate only planar HDR images but many applications require full-view HDR images. Thus, we would like to extend similar ideas to spherical images and develop an omnidirectional multi-view HDR framework.

2.5 Omnidirectional HDR

Omnidirectional images may be generated in the following three ways [87]: stitching multiple images, using a special lens or using common camera lens with a mirror. Accordingly, omnidirectional cameras can be classified into the following three types:

- Polydioptric: multiple cameras with overlapped views of field.
- Dioptric: combination of shaped lens e.g. fisheye lens.
- Catadioptric: a standard camera combined with a shaped mirror, e.g. convex mirror.

The Ladybug 2 camera set from Point Grey falls into the Polydioptric type as it uses six independent lenses. The SpheroCam camera from Spheron-VR AG [12] captures omnidirectional images stitched from multiple images captured in sequences from a rotating camera.

Omnidirectional HDR images are fused from an exposure bracketed LDR images captured using a stationary omnidirectional camera, e.g. SpheroCam HDR [12] and Fisheye lens [89]. Okura et al. [90] tried capturing full spherical HDR images of the sky using two fisheye cameras: one pointing up and one down.

The Catadioptric type refers to a capture system built with a lens and a mirror. The mirror can be in different shape, e.g. convex, sphere. The HDR images used for computer

graphics are captured using a camera and a light probe (chrome ball) [96]. An extension of this idea has been made for HDR video capture. Software post-processing is needed for assembling the sequences of sphere images captured. Combining the concept of spatially varying sensors with beam splitter [50] and light probe [96], Unger et al. [97] developed a system with multiple sensors and light probe that is able to capture omnidirectional HDR images in one single shot. The main problem with the mirrored ball is that sphere images are highly non-uniformly sampled; the central area has high resolutions while the rim region has very low resolutions. Although omnidirectional stereo vision [91] has been explored for robotic navigation, limited work has been done on the acquisition of omnidirectional HDR stereo.

2.6 Summary

A review has been given on the related research to HDR imaging with emphasis on resolving image registration techniques in the presence of camera and object movements. HDR imaging is a sub-field of computational photography which is a broad and active field for extending the capacity of digital cameras. Section 2.1 has outlined the basic idea and scope of computational photography and Section 2.2 has described the different methods for the generation of HDR images and the challenges.

As the most common way of HDR imaging is through assembling a sequence of LDR images captured with different exposures, considerable research effort has been made to address the challenges of pixel registration in the presence of camera and object movements. Section 2.3 has reviewed different techniques in three categories. The first category includes the techniques to align the scene's static background which can only deal with camera movement. The second category covers those techniques for the detection and removal of artifacts caused by moving objects in a dynamic scene. The last category refers to the techniques that can deal with both camera and object movements simultaneously, e.g. motion estimation by optical flow.

Section 2.4 has discussed the recent development in HDR imaging through stereo correspondence that can deal with large and general camera movement and can acquire both high dynamic range and 3D (depth) information. Section 2.5 has outlined omnidirectional HDR imaging techniques.

Based on the review in this chapter, we have found that most HDR imaging methods for dynamic scenes deal only with small camera movement such as camera shaking. Only limited work has been done on the multi-view HDR imaging but with small exposure and view differences. Furthermore, it seems that no attempts have been made on the generation of omnidirectional HDR imaging from multi-view and multi-exposed LDR images. Thus, this thesis is inspired to develop an omnidirectional HDR imaging framework based on the multi-view geometry for handling large and general camera movement.

Chapter 3

Data Acquisition and Omnidirectional Image Creation

Our multi-view omnidirectional HDR imaging framework requires cylindrical panoramas as input. The major steps from image data capture to cylindrical panorama generation are shown in Figure 3.1. In addition, photometric calibration for the recovery of inverse camera response curves will be covered in this chapter.

Section 3.1 will describe the experimental design of image capture that provides data for the generation of HDR images using multi-exposed images captured from a single view or multiple views.

For the recovery of scene radiance from measured pixel values, we need to ensure that pixel values in images map to exposure (product of irradiance and exposure time) consistently. Thus, Section 3.2 will describe the configuration of the camera system for consistent capture of radiance, the methods and procedures for the image data capture.

Section 3.3 will explain some common panoramic image formats used to represent omnidirectional images: cubic, cylindrical, spherical, and rhombic-dodecahedron map (RD-map). Only cylindrical and RD-map formats are used in this thesis.

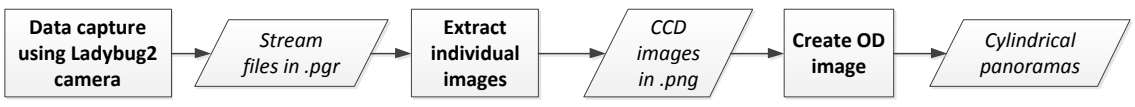


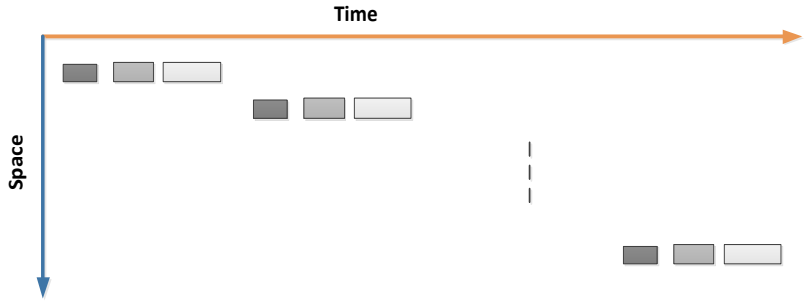
Figure 3.1 Pipeline from data capture to omnidirectional image generation.

Section 3.4 will describe the photometric calibration for the recovery of inverse camera response curves, and present calibration results. Finally, a summary will be given in Section 3.5.

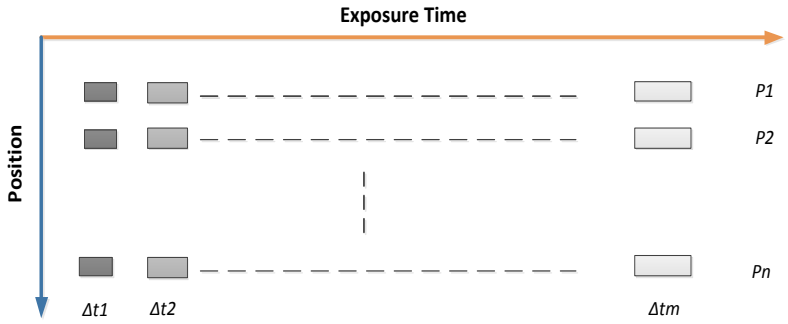
3.1 Experimental Design

HDR images can be created from LDR images captured in different ways and we can classify HDR imaging methods accordingly. The classic method for creating HDR images is to capture LDR images in a time sequence with the camera placed at a fixed position, we refer to this method as temporal HDR imaging following the definition of Rufenacht [36]. Similarly, if HDR is created from LDR images captured at the same time but at different camera positions, we call this method spatial HDR imaging. Accordingly, we would like to develop methods for spatiotemporal HDR imaging, i.e. HDR recovered from LDR images captured at different positions and at different time instances.

Our experiments were designed to capture sufficient data of different indoor and outdoor scenes which allow for the comparison of different ways of finding stereo correspondences and generating HDR. As shown in Figure 3.2, the images are captured first temporally and then spatially. In other words, the camera placed on a tripod captures different images automatically with exposure time increased from low to high (Δt_1, Δt_2, .., Δt_m) at each of positions from P_1, P_2, …, P_n. The data matrix captured allows us to create temporal HDR using LDR images for each position, and spatiotemporal HDR imaging when selecting images across different positions. We can perform stereo matching using the images with the same or different exposures at different positions to assess the impact of exposure differences on stereo matching algorithms.



(a) Photo capture flow



(b) Data matrix collected

Figure 3.2 Experimental design. Data are captured sequentially with a camera resulting in a data matrix collected for each position with the same number of exposure increments and steps.

3.2 Image Data Acquisition

The Ladybug 2 omnidirectional camera system was used to capture images data. The technical features and software SDK and tools are described in Appendix A. We will discuss the HDR option existing in the Ladybug SDK, and the data capture, the camera configuration and the capture procedures for the proposed HDR framework.

3.2.1 Existing Ladybug HDR option

HDR imaging requires the capture of images with multiple exposures to increase the dynamic range of irradiance. The Ladybug 2 system contains six lenses/sensors each of which can be controlled independently. If each sensor with auto-exposure captures a region of different radiance, the final panorama image can contain a broader range of radiance of the scene. However, this may not work if a sensor faces a region of great contrast.

The Ladybug system also offers an HDR mode which allows for an exposure bracket of four sets of user defined shutter/gain combinations. In the HDR mode, the exposure bracket is cycled through and all sensors are set to the same set of exposures. Four exposure settings are not sufficient for our data capture design as we intend to capture exposure sequences with small increments and more steps. Thus, we do not use the HDR mode built in the SDK and use a customized capture procedure.

3.2.2 Proposed Image Capture for HDR imaging

In order to allow us to evaluate different types of HDR generation methods, the data capture was designed to capture as much data as possible. Instead of capturing one image at each position, we captured the same exposure bracket at all camera positions (views). In doing so, we can use the image sequence with the same optimum exposure to obtain the disparity maps from stereo matching so that we can have a comparison standard for stereo matching of multi-exposed multi-view images. For each viewpoint (at each capture position), we can easily create an HDR image from the LDR images captured at the same position using the classic temporal HDR method.

In photography, the intensity of images can be changed by adjusting the amount of light using aperture, shutter speed or ISO speed (or gain). In the HDR community, shutter speed (or exposure time) is commonly used because aperture can affect depth of field and a high ISO speed produces more noise. A stop is used to measure the relative amount of light. An

increase in exposure by one stop means doubling the amount of light whereas a decrease by one stop implies halving the amount. As the amount of light falling on a sensor is proportional to exposure time, +1 stop means doubling the exposure time and -1 stop halving the exposure time.

The exposure bracket at different positions uses a fixed increment in stops and a fixed number of steps. The camera positions are about 20-30 cm apart along the capture path. Ladybug_vidcapture has been modified to allow us to specify the exposure increments in stops (usually 1/3 or 2/3 stop), and the number of increment steps to be applied in the configuration file.

3.2.3 Camera Configuration Control

As we intend to obtain the radiance maps from omnidirectional LDR images, we need to make sure the pixel values are affected only by the exposure (irradiance and exposure time). We need to disable the features that can affect the pixel values, e.g. auto-white balance and auto-exposure, gamma correction and black level, and then manually set the frame rate and red and blue channels to the values from a manual white balance.

The white balance is conducted manually based on the MacBeth ColorChecker test chart [94] (Figure 3.3). The test chart contains 4x6 patches of different colors with the reference color space values and reflectance data available. Therefore, it is widely used for the calibration of image capture devices such as cameras and scanners. In our work, it is used both for white balance and photometric calibration of sensors (see the next section).

The LadybugCapPro software tool was used to capture images of the test chart with different combination of red and blue channel registrar values. The bottom gray patches were used to assess the optimal white balance and select the optimal ratio of the red and blue channels. Once determined, the same values for the red and blue channels were configured before all data capture.

Figure 3.3 Photograph of the MacBeth ColorChecker test chart [94]. The bottom six gray patches from bright to dark are used for white balance and photometric calibration.

3.2.4 Image Capture Procedure

The following procedure is used for all data capture:

1. Place the Ladybug 2 camera set on a steady tripod and connect it to the PC with the SDK and Ladybug_vidcapture tool installed.
2. Start the LadybugCapPro tool to control the Ladybug camera system, and go to configuration menu to disable all auto exposure, white balance features, and set gamma to *1*, and set the red and blue channel to the optimum values based on the manual white balance. Capture a few sample images and ensure if all configuration settings are in effect and are correct. Stop LadybugCapPro tool.
3. Specify the exposure increment in stops and the number of increments and gain in the configuration file for the Ladybug_vidcapture tool.
4. Move the Ladybug camera (mounted on a steady tripod) to a required position.
5. Start the Ladybug_vidcapture tool and select image type to capture, e.g. *.pgr (Point Grey stream file) and individual sensors.

6. Start to capture images with the specified exposure sequence.
7. Once the capture is finished, go to Step 3 for another capture.

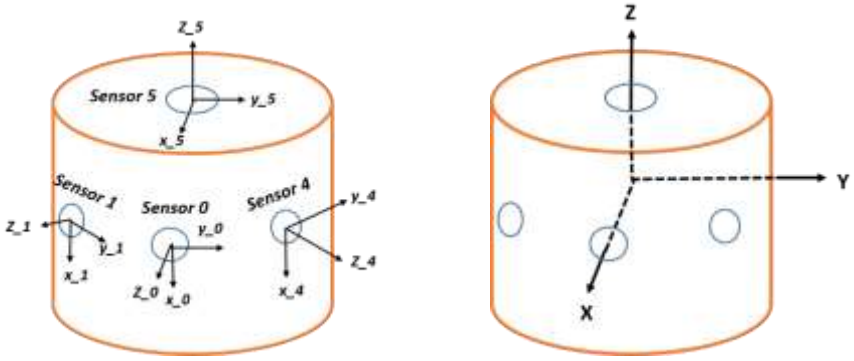
3.3 Omnidirectional Image Generation

Common images captured by consumer cameras are planar (rectilinear), i.e. the lines in a scene remains straight in planar images. When we create panoramic images with a wide field of view, i.e. over 180 degrees, planar images cannot be used. Some common panoramic image formats used to represent the omnidirectional images are cubic, cylindrical and spherical. Rhombic dodecahedron map (RD-map), recently proposed by Fu et al. [39], is able to provide more uniform sampling than other formats and is thus used in the Spherical Vision Tool developed by Brunton [40]. In the following section, we will outline the basic ideas of producing cubic, cylindrical, spherical and RD-map images and how the raw image data are converted in those formats.

We will explain how the Ladybug SDK can convert pixels in 2D images to 3D ray vectors. Then we will describe how 3D ray vectors can be projected onto different 3D surfaces and further mapped to different 2D panoramic image formats.

3.3.1 Image Pixels to 3D Rays

The Ladybug system uses different local coordinates for each sensor and the global coordinates for the system (Figure 3.4). It also provides API functions for the transformations between the local and global coordinates, and pixel locations and 3D rays in global coordinates.



(a) Sensor coordinate (b) Ladybug Coordinate

Figure 3.4 Local and global coordinates used in the Ladybug camera system. The X-Y axes are on the same planes on which the projection centres of the five side sensors are located. The Z axis is in the same direction as the z axis of the top sensor.

We can create different panoramic representations from raw image data captured using the Ladybug camera set. For a given 3D ray vector, we can determine the corresponding pixel value from the raw image data captured.

3.3.2 Panoramic Image Format

Omnidirectional images can be considered as the capture of 3D light rays reflected from all surrounding scene points. First, we can project all rays to some kind of spherical surface in 3D space and then map the surface to a 2D format of representation for display. In this section, we will outline the cubic, cylindrical, spherical and RD-map formats.

Cubic Projection

A cubic panorama is one of the common formats used for environmental maps in computer graphics. With the projection centre at the centre of the (unit) cube, all scene points visible

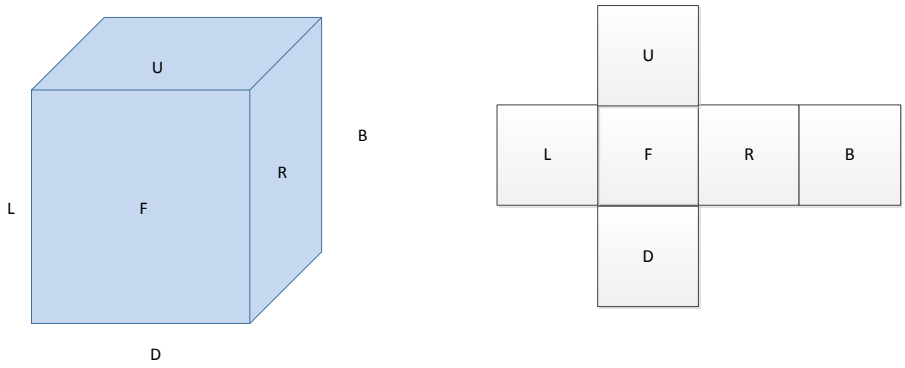


Figure 3.5 Six cubic surfaces unrolled to 2D planar representation. U, D, F, B, L and R denote up, down, front, back, left and right faces of the cube.

can be projected onto the six faces. To represent the panoramic image in a 2D image format, we can unwrap the six faces to place them in the layout as shown in Figure 3.5. As we do not use the cubic panorama, we will not go into details for the mapping of 3D ray vectors to 2D representation.

Cylindrical Projection

In a cylindrical projection, the projection centre is placed at the centre of the cylindrical surface in 3D as shown in Figure 3.6. 3D Rays intercepts the unit cylindrical surface. The cylindrical surface maps to the planar mage coordinates with the origin at the bottom left corner. The lines in a scene are not necessarily straight in the 2D cylindrical images.

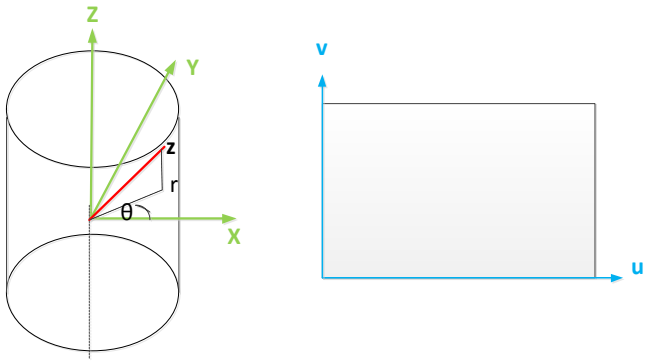


Figure 3.6 Projection of scene points to cylindrical surface and unrolled to 2D planar cylindrical panoramic image.

The transformation from the 3D coordinates (*X*, *Y*, *Z*) to the cylindrical coordinates (*θ*, *r*, *z*) can be performed using the following expressions

$$\theta = \text{atan}(Y/X) \tag{3.1}$$

$$r = \sqrt{(X^2 + Y^2)} \equiv 1 \tag{3.2}$$

$$z = Z \tag{3.3}$$

For the 2D representation, the cylindrical surface can be unrolled to 2D space and mapped to the rectangular 2D image coordinates (*u*, *v*). The u coordinate is mapped to the *θ*, and the *v* coordinate to the *z* coordinate. If the cylindrical surface is mapped to a 2D image, the mapping relationship can be expressed as:

$$u = \left(\frac{\pi - \theta}{2\pi}\right) \tag{3.4}$$

$$v = \frac{z - z_{min}}{z_{max} - z_{min}} \tag{3.5}$$

On the other hand, we need to map (*u*, *v*) to a 3D ray vector (*X*, *Y*, *Z*) in order to generate a 2D cylindrical panorama image from image data captured. Starting with 2D coordinates (*u*, *v*), we can calculate the corresponding cylindrical coordinates (*θ*, *r=1*, *z*) as

$$\theta = \pi(1 - 2u) \tag{3.6}$$

$$z = v(z_{max} - z_{min}) + z_{min} \tag{3.7}$$

Then we calculate the 3D ray vector point using the below equations:

$$X = cos\theta \tag{3.8}$$

$$Y = sin\theta \tag{3.9}$$

$$Z = z \tag{3.10}$$

For a given 3D ray vector, the corresponding pixel value from the individual Ladybug sensor images can be found by calling the Ladybug library function.

Equirectangular Projection

Similarly we can project all rays to the unit sphere surface with the projection centre at the centre of the unit sphere (Figure 3.7). The unit sphere is defined in the longitude-latitude coordinates. When mapped to 2D coordinates, the longitude coordinate is transformed to the horizontal coordinate (u), and the latitude to the vertical coordinate (v).

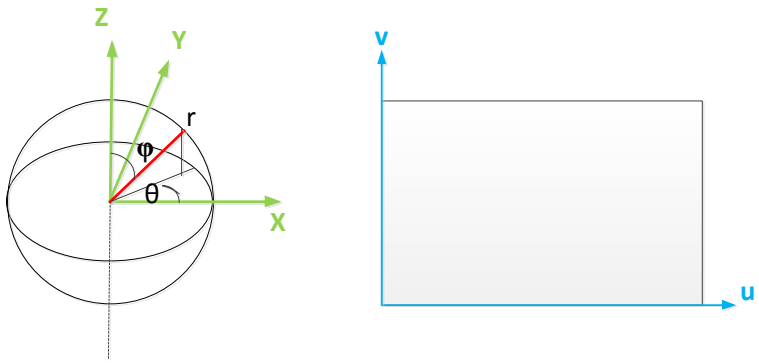


Figure 3.7 Projection of scene points to unit sphere surface and unrolled to 2D planar spherical panorama image.

The transformation from the 3D coordinates (*X*, *Y*, *Z*) to the spherical coordinates (θ, ϕ, *r*) can be performed using the following expressions

$$\theta = \text{atan}\left(\frac{Y}{X}\right) \tag{3.11}$$

$$\varphi = \text{atan}\left(\frac{\sqrt{(X^2 + Y^2)}}{Z}\right) \tag{3.12}$$

$$r = \sqrt{(X^2 + Y^2 + Z^2)} \equiv 1 \tag{3.13}$$

To map the surface in the spherical coordinates (θ, ϕ, *r*) to a 2D image, the following expressions can be used:

$$u = \left(\frac{\theta - \pi}{2\pi}\right) \tag{3.14}$$

$$v = \left(\frac{\pi - \varphi}{\pi}\right) \tag{3.15}$$

To create 2D spherical panorama image, we need to find the color information from the raw image data for a pair of 2D (*u*, *v*) coordinates. The following steps are used to this end. First we need to find the corresponding spherical coordinates (θ, ϕ, *r =1*):

$$\theta = \pi(1 - 2u) \tag{3.16}$$

$$\varphi = \pi(1 - v) \tag{3.17}$$

Then we can find the 3D ray vector as

$$X = cos\theta sin\varphi \tag{3.18}$$

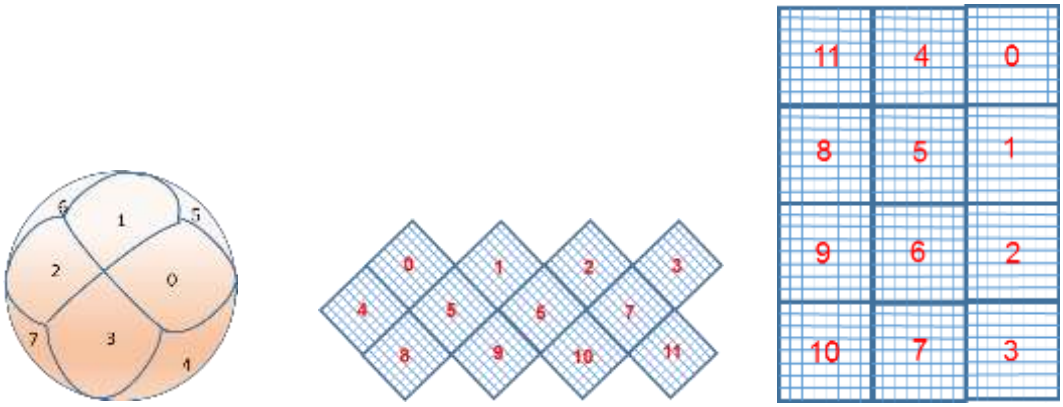
$$Y = sin\theta sin\varphi \tag{3.19}$$

$$Z = cos\varphi \tag{3.20}$$

The pixel value for (*u*, *v*) coordinates can be found by calling the Ladybug library function to extract from the raw data for the given 3D ray vector.

RD-map Projection

The most common issue with spherical or cylindrical panoramas is that the sampling region on the 3D surface represented by a 2D image are non-uniform. To achieve a more uniform sampling pattern, the rhombic dodecahedron map (RD-map) has been proposed by Fu et al. [39]. Each pixel in the RD-map (Figure 3.8) spans almost the same solid angle and thus the shape distortions of subdivided pixels are very similar and small. However, the conversion between the 2D coordinates and 3D rays is not straightforward and the technical detail can be found in the paper by Fu et al. [39]. A library is implemented in SVT (Spherical Vision Toolkit [40]) for providing common functionality for the convention.



(a) Surface division (Rhombi) (b) Unrolling of Rhombi (c) Displayed as 2D image

Figure 3.8 Rhombi and 2D representation for RD-map [39].

3.3.3 Creation of Omnidirectional Images

In-house software tools have been developed and modified for the generation of omnidirectional images in required formats.

First, we need to extract the individual images for each sensor using the Extracter (by Prof. Lang). This tool takes .pgr images captured from the Ladybug camera set as input

and extracts the individual images for each sensor and saves them in .png format. The configuration file can be used to specify the image size and color processing pattern etc.

Second, the Create_Pano_Ladybug in-house tool is used to generate the omnidirectional images using the individual sensor images and capture parameters, and camera calibration properties and stitching masks. A configuration file is also provided to allow for the selection of panoramic image type and size.

3.4 Photometric Calibration

To recover the radiance values in an LDR image, we need to know the response function that map pixel values to exposure (product of exposure time and irradiance). In the following section, we will describe the method and the tool used for the recovery of the response curves, and photometric calibrations based on raw sensor images and cylindrical images. We also check if the same radiance values are mapped to pixel values consistently across the six different sensors in the Ladybug 2 camera set.

3.4.1 Calibration Method and Tool

There are a number of different methods for the recovery of camera response curves. The approach proposed by Debevec et al. [3] is selected for its simplicity and robustness. The camera response curve is expressed as a function of irradiance and exposure time:

$$Z_{ij} = f(E_i * \Delta t_j) \tag{3.21}$$

In HDR imaging, the inverse response curve, i.e. exposure to pixel relationship as

$$\ln X_{ij} = g(Z_{ij}) \tag{3.22}$$

A software tool based on the above method was implemented in MATLAB by Eitz and Stripf [95] and will be used to determine the camera response curves.

3.4.2 Calibration Based on Sensor Images

Common camera sensors are linear to exposure but the processing pipelines may introduce nonlinearity e.g due to gamma correction. According to the calibration performed by Silk [26] using the Robertson's method, the Ladybug 2 sensors are approximately linear.

In the current work, the linearity is evaluated in two approaches. First, the MacBeth ColorChecker test chart (Figure 3.3) was used to measure directly how the pixel values for each of six gray patches change with exposure time. The result is shown in Figure 3.9. As can been observed, it is obvious that a linear relationship between pixel value and exposure time holds for each of 6 patches (one curve for each patch). The gray patches from low (dark) to high (bright) reflectance exhibit different slopes. If the illumination is approximately the same, the radiance values of patches are proportionally to the reflectance values. Thus, higher reflectance leads to a steeper slope.

The inverse camera response curves are recovered from the images with different exposure times for Sensor 0 and plotted in a linear plot and a semi-logarithmic plot in Figure 3.10. The curve exhibits linearity up to pixel values of 200 or more, which is consistent with what is found by Silk [26].

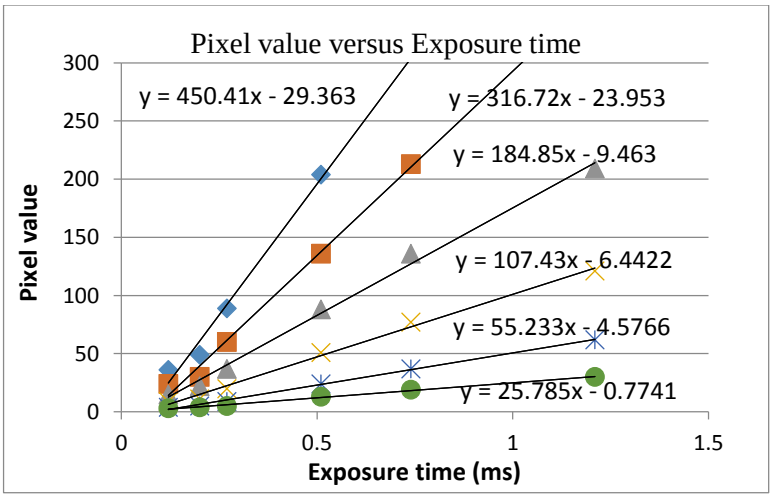
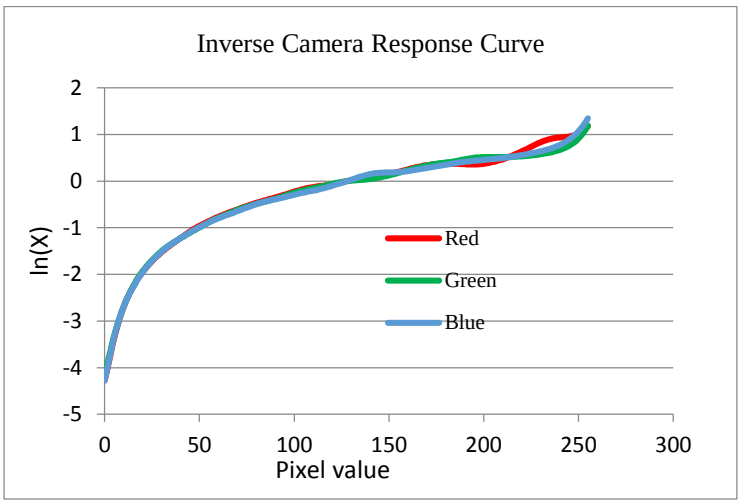
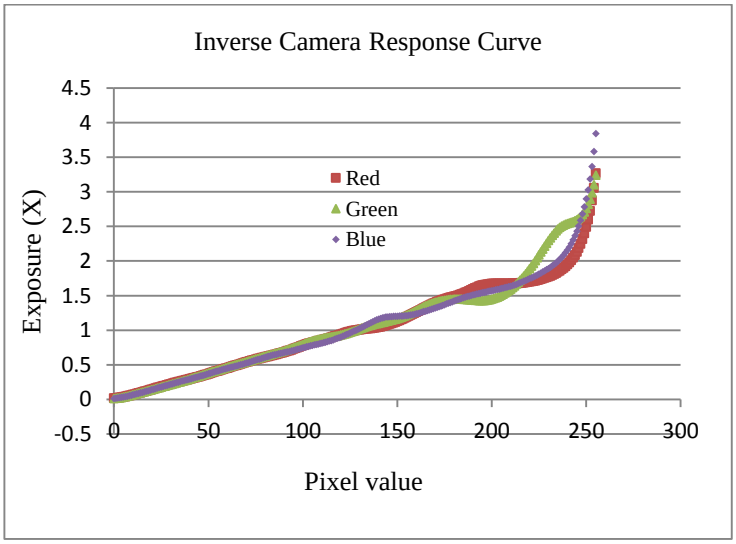


Figure 3.9 Evaluation of linearity of Sensor Images.



(a) Semi-logarithmic plot



(b) Linear plot

Figure 3.10 Inverse camera response curves (pixel value versus exposure) for a senor in (a) semi-logarithmic plot and (b) linear plot.

In addition, we would like to check if different sensors have similar responses to the same radiance values. As the MacBeth ColorChecker test chart (Figure 3.3) provides standard color patches and is widely used for calibration of cameras, we placed the chart under the same illumination and turned each sensor in the Ladybug 2 camera set directly towards the chart to take one image in turn. The cropped chart images are shown in Figure 3.11. Those chart images look almost identical across different sensor images. The detailed values of each gray patch are compared across sensors in Figure 3.12. The differences between sensors are less than 5%. Thus, it is reasonable to assume that the six sensors exhibit the same response.

(a) Sensor 0; (b) Sensor 1; (c) Sensor 2; (d) Sensor 3; (e) Sensor 4; (f) Sensor 5.

Figure 3.11 Cropped images of the test chart from the images captured by each sensor in the Ladybug 2 camera set.

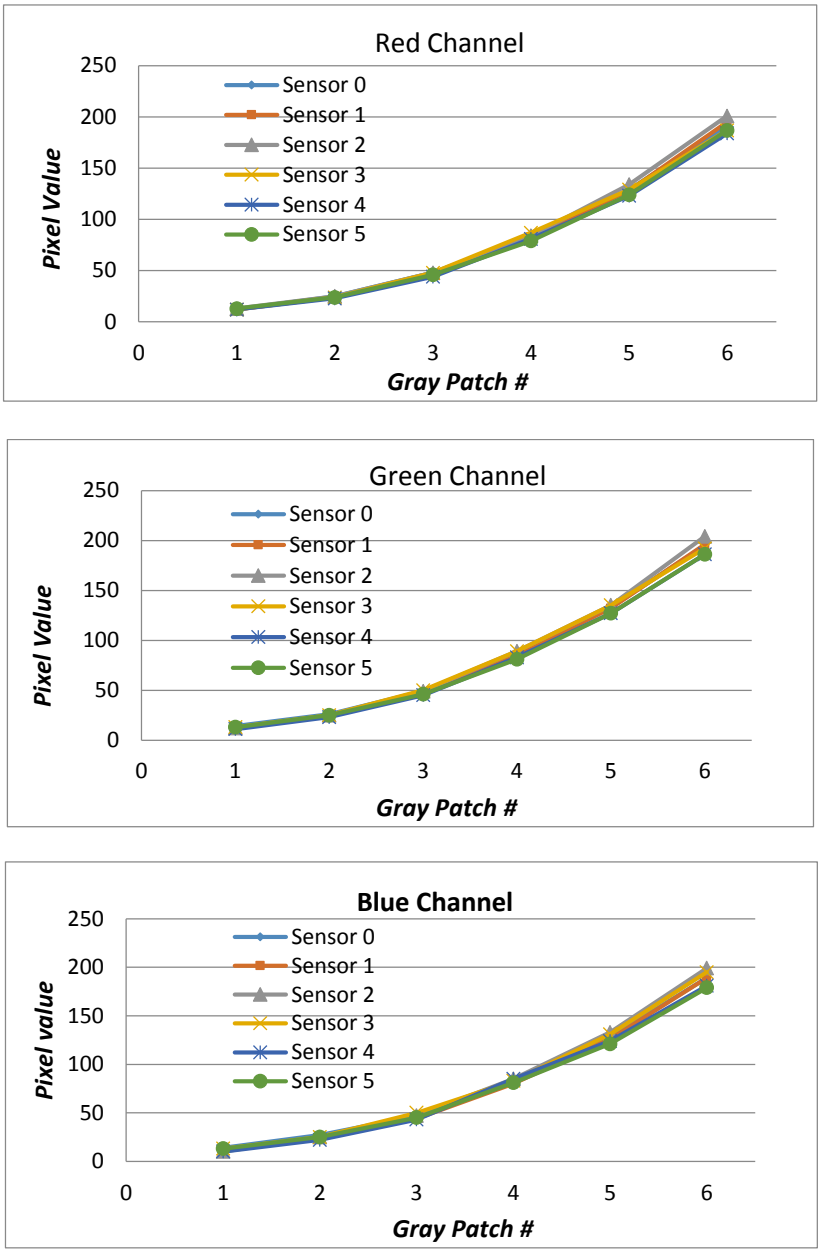


Figure 3.12 Assessment of consistency across sensors. Comparisons of pixel values in red, green and blue channels for six gray patches in test chart images each captured by one sensor.

3.4.3 Calibration Based on Panoramic Images

As we use cylindrical panoramic images as input to stereo matching in the determination of disparity maps before assembling HDR images, we need to recover the camera response curves based on the cylindrical panoramic images.

The camera response curves are different if the gain is set to different values. For an outdoor scene, the gain was set to *0* and for the indoor scene the gain was set to 9 dB. The inverse response curves recovered from cylindrical panoramic images of the outdoor and indoor scenes are displayed in Figure 3.13 and Figure 3.14, respectively. Although the low end of the response curve for the red channel in Figure 3.14 appears to be away from being linear, all the rest response curves appear to be linear. Thus, approximately linear relationships are assumed between logarithmic exposure and pixel value for the inverse response curves and the best-fit equations in Figure 3.13 and Figure 3.14 will be used to recover radiance values from LDR images with known exposure times.

Please note that the relative radiance is set to *1* as the reference point when the shutter speed is *1 second* at the gray level of 128 in each case. Relative radiance values cannot be compared between the images with different camera settings (e.g. different gains).

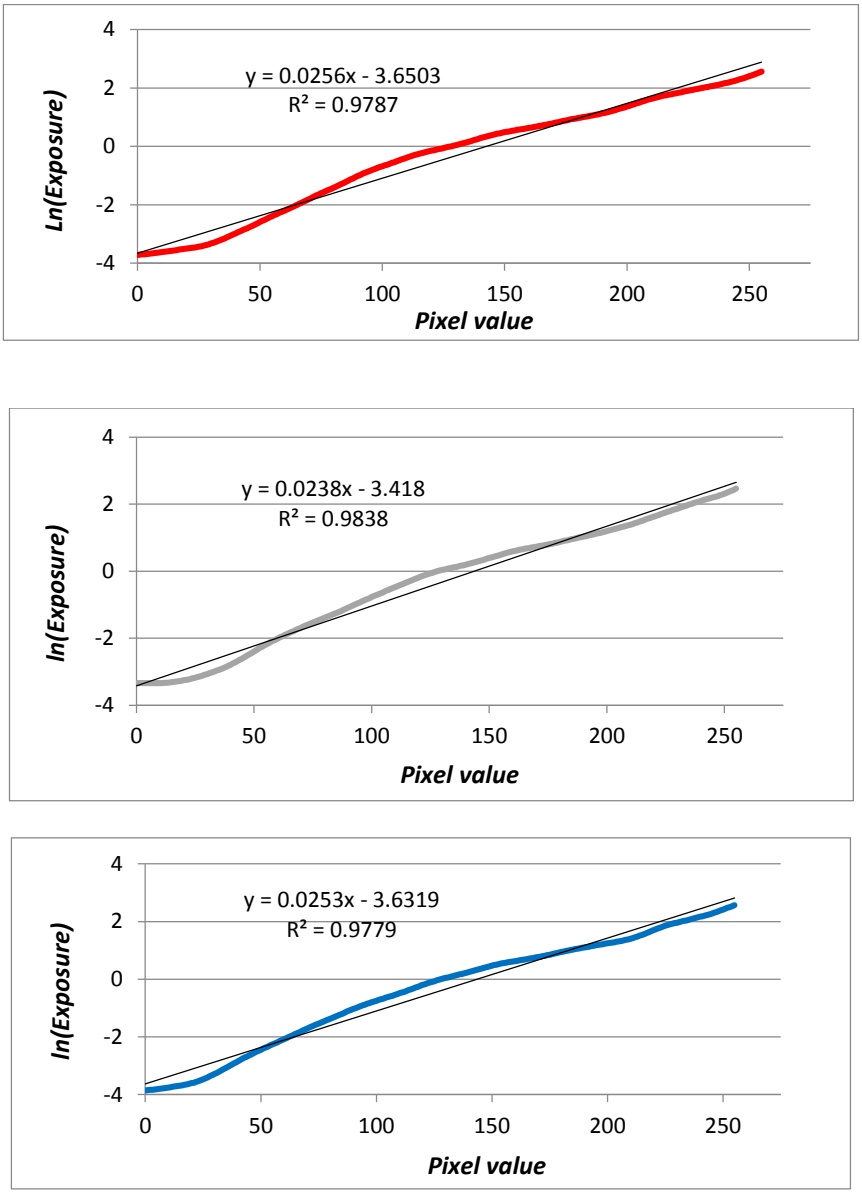


Figure 3.13 Inverse camera response curves for different color channels based on cylindrical panoramic images for the front yard outdoor scene (gain: 0 dB). From up to down: red, green and blue channels.

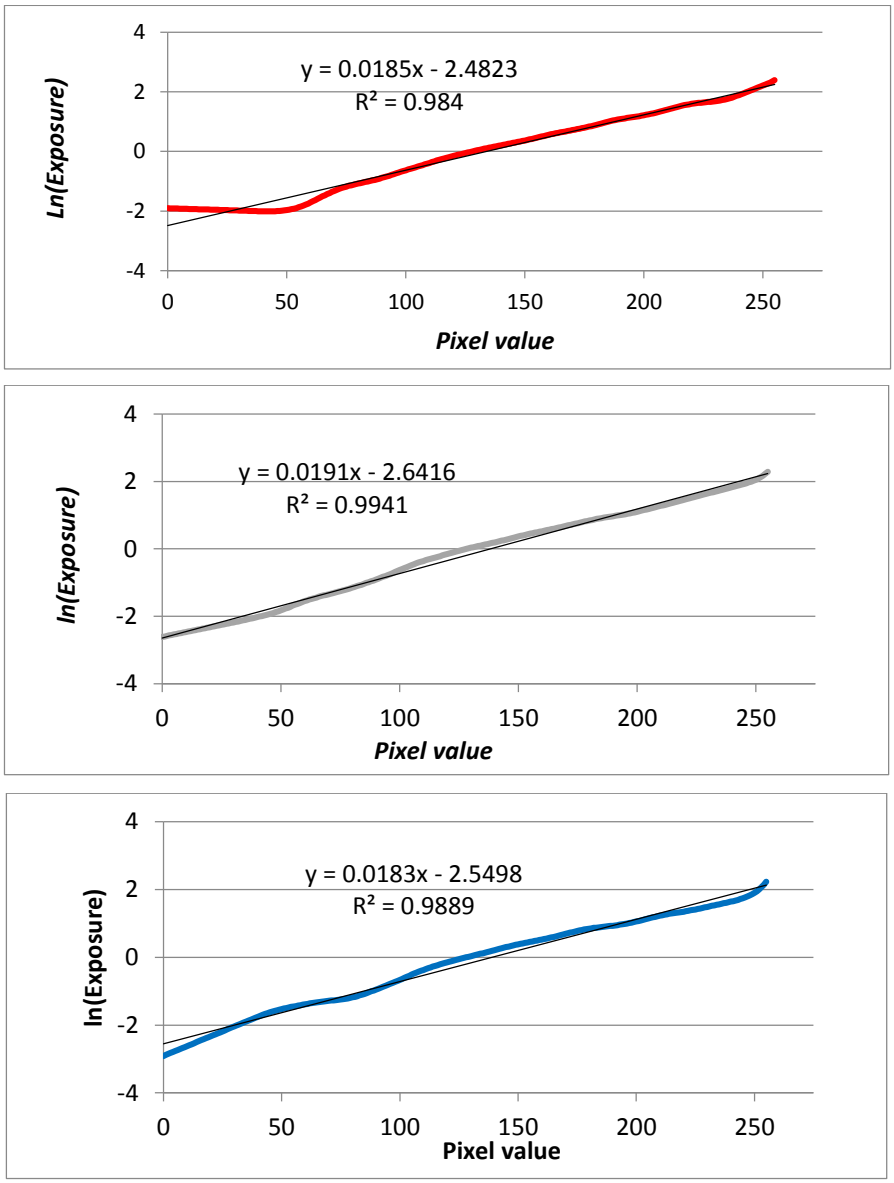


Figure 3.14 Inverse camera response curves for different color channels based on cylindrical panoramic images for the Ruth E. Dickinson public library indoor scene (gain: 9 dB). From up to down: red, green and blue channels.

3.5 Summary

This chapter has covered the experimental design in Section 3.1, and the data capture methods and procedures in Section 3.2. In Section 3.3, we have described the different omnidirectional projections in 3D and image formats in 2D such as cubic, cylindrical, spherical and RD-map, even though only cylindrical and RD-map panoramas are used in this thesis. The image data captured have to be converted into cylindrical panoramas in order to serve as input to spherical stereo matching.

To generate high dynamic range images, we need to convert LDR images in pixel values to radiance space. To this end, we have carried out the photometric calibration to determine the camera response curves. The calibration method, the calibration tool and the results of response curves have been presented in the Section 3.4. The response curves for the sensor images show a linear relationship between exposure and pixel value whereas the response curves based on the cylindrical omnidirectional images exhibit approximately a semi-logarithmic relationship. The response curves will be used to recover radiance values from LDR images.

Chapter 4

Stereo Correspondence of Omnidirectional Images

To compose an HDR image from multi-view and multi-exposed images, we need to find dense correspondence of the scene points between images from different viewpoints. We can use stereo matching to find the dense correspondence between a stereo pair which can be represented in a disparity map. Disparity is defined as the difference in the horizontal location for two rectified planar images. As disparity is inversely proportional to depth, disparity map may also be referred to as depth map. For spherical stereo vision, different types of disparity will be defined in Section 4.2. The pipeline from cylindrical input images to disparity maps is shown in Figure 4.1.

There has been significant progress in stereo vision and over 100 algorithms have been are proposed for stereo matching [15]. However, most algorithms are applicable only to planar images. Lines in a scene remain to be lines in planar images but may become curves in omnidirectional images. In addition, pixel sampling in omnidirectional images may be

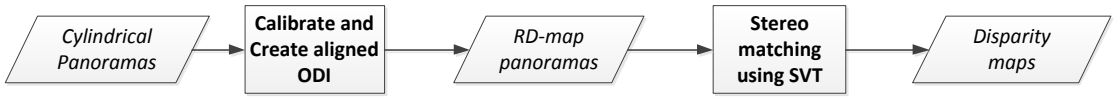


Figure 4.1 Pipelines of stereo correspondences.

distorted and non-uniform. Therefore, research on stereo matching of omnidirectional images is limited.

A multi-scale spherical stereo matching framework recently proposed by Brunton et al. [41] seems effective to construct spherical disparity maps and has been used for view synthesis in virtual navigation. A software tool, the Spherical Vision Toolkit (SVT), has also been developed. However, the framework is developed and tested for stereo matching of similarly well-exposed spherical images. Thus, we apply this framework to stereo matching of spherical images captured with different exposure times. We need to determine camera poses at each capture position and align the spherical images in the rhombic dodecahedron map (RD-map) before feeding to the SVT pipeline for stereo matching.

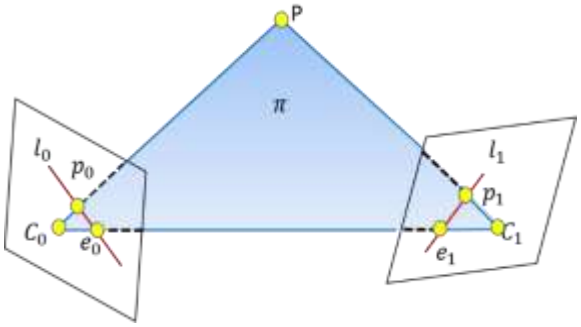
Thus, Section 4.1 will explain how to perform geometric calibration to determine the poses (rotation and translation) of each view, and how to create RD-map images and how to rotationally align them using the pose information. In Section 4.2, the multi-scale stereo matching framework will be outlined and the pipeline for SVT will be described in detail.

The multi-scale stereo matching framework has been tested successfully for both planar and spherical image data [41]. Although the algorithm is designed to handle small variation in illumination, it has not been tested on images with exposure differences and saturated regions. In Section 4.3, an experimental evaluation will be presented to show the different stages and the performance of SVT using images with exposure difference and saturated regions. Finally a summary will be given in Section 4.4.

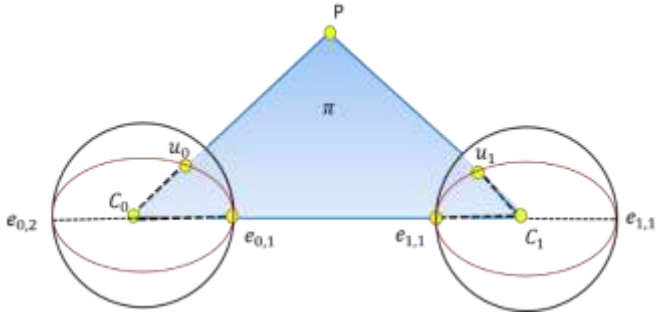
4.1 Camera Calibration and Alignment

4.1.1 Geometric Calibration

In stereo vision, geometric calibration can be used to determine intrinsic parameters such as focal length, optical centres, and extrinsic parameters such as translations and rotations. Because the Ladybug 2 camera system is calibrated and the intrinsic parameters are provided, we need to determine only the extrinsic parameters in the calibration. Camera rotation and translation between two views can be recovered from the essential matrix which is defined by the epipolar geometry (Figure 4.2). The projection centres of two cameras $\boldsymbol{C_0}$ and $\boldsymbol{C_1}$ and the line connecting the two centres is referred to as baseline. The



(a) Planar image pair



(b) Spherical image pair

Figure 4.2 Epipolar geometry for planar and spherical stereo matching pair [97].

epipoles are the projected points on the current image plane from the other projection centre, and they are on the baseline. For a scene point *P* and its projection to one image plane as $\boldsymbol{p_0}$, the matched point $\boldsymbol{p_1}$ on the other image plane should be on the epipolar line $\boldsymbol{l_1}$. If we know the camera pose, we can determine the matching point in the other image for any given point in one image.

On the other hand, if we know a sufficient number of matching pairs, we can derive the translation and rotation. If a scene point projects to the left image plane as $\boldsymbol{p_0}$ and the right plane as $\boldsymbol{p_1}$, there exists epipolar geometry defined as

$$\boldsymbol{p_0^T E p_1} \tag{4.1}$$

Where $\boldsymbol{E}$ is essential matrix and $\boldsymbol{p_0}$ is the ray vector in the left camera coordinate frame with origin of $\boldsymbol{C_0}$ and $\boldsymbol{p_1}$ as the ray vector in the right camera coordinate frame with origin of $\boldsymbol{C_1}$. A similar epipolar geometric relationship exists for a spherical image pair if the two matching points $\boldsymbol{u_0}$ and $\boldsymbol{u_1}$ are defined as 3D vectors in the camera coordinates.

$$\boldsymbol{u_0^T E u_1} \tag{4.2}$$

However, the epipolar lines become epipolar circles as the epipolar plane, defined by the two projection centres ($\boldsymbol{C_0}$ and $\boldsymbol{C_1}$) and the scene point $\boldsymbol{P}$, intersects the two spherical surfaces forming two epipolar circles. The epipolar circles are great circles because the intersection plane passes through the centre points of the spheres.

The essential matrix has 7 degrees of freedom and thus needs only 4 matching pairs to solve the Equation 4.2 using the so-called 8-point algorithm [62]. Thus sparse correspondence is sufficient. There are a number of features descriptors available such as SIFT (Scale Invariant Feature Transform [64]), SURF (Speed-Up Robust Feature [86]). SIFT leads to a more reliable feature detection whereas SURF is much faster. Because there exist some outliers in the matching points found, RANSAC (RANdom SAmple

Consensus [65]) is commonly used to determine model parameters. The basic idea is to select randomly the minimum set of data to find the model parameters and use the model to determine the number of the data points fitting to the model within a tolerance. The process is iterative and stops if the percentage of the data points fitting is over a threshold.

We rely on an in-house tool, CaliPano, developed for spherical stereo vision in [40] for calibration. As for multi-view and multi-exposed image matching, wide view panoramic images are expected to lead to more robust and accurate camera poses than are planar images. This is because there is a sufficient overlap even for great changes in camera poses and the wide field of view makes the calibration problem better conditioned. Thus, calibration is performed using our stitched panoramic images rather than using individual planar images.

The CaliPano tool uses OpenSurf [110] to find matching points between cylindrical image pairs. The matching points in pixel coordinates can be converted to 3D (three dimensional) ray vectors in the Ladybug coordinates for each view. The 3D ray vectors are used to estimate the essential matrix using RANSAC [65]. Once the essential matrix is determined, the rotation and translation can be solved using the single value decomposition of the essential matrix.

Subsequently, a spherical bundle adjustment is applied to refine the pairwise calibration parameters and the positions of the 3D points by directly minimizing the re-projection error, as measured by the angle between the detected feature direction and that of the re-projected scene point. Specifically, CaliPano minimizes the energy

$$E_r(C,u) = \sum_i \sum_j v_{ij} \arccos(\hat{\boldsymbol{y}}_{\boldsymbol{ij}} . \hat{\mathbf{u}}_{ij}) \tag{4.3}$$

Where C is the set of camera extrinsic parameters, $\hat{y}_{ij}$ is the location of feature j as detected in image i, $\hat{u}_{ij}$ is the projection of 3D scene point u_j into image i, and v_{ij} is a visibility term indicating whether feature j was detected in image i.

4.1.2 Rectification and Rotational Alignment

For planar images, stereo images can be resampled onto a common plane such that the search for matching point can be done along scan lines for efficiency.

It is also possible to rectify spherical images so that the epipolar circles becomes vertical lines in an equirectangular projection. This is achieved by rotating the two spheres so that poles align with the baseline and the epipolar circles become latitudes.

For spherical images represented with a RD-map, there is no simple way of turning epipolar circles into vertical or horizontal lines on the RD-map. Thus only rotational alignment is performed so that the optical axis of the camera for each view can be aligned. Once the rotation matrix is determined from the geometric calibration, we can apply it on the ray vectors to resample from raw data to create a rotationally aligned RD-map image for each view. Aligned panoramic images in RD-map are used as input to the multi-scale spherical stereo vision to find the dense correspondence between multi-exposed images captured at different viewpoints.

4.2 Multi-scale Stereo Matching Framework

While there are numerous choices for stereo matching, we use the multi-scale spherical stereo matching framework proposed recently by Brunton et al. [41]. This framework uses a hybrid approach: matching cost evaluated locally and aggregating matching cost globally

using multi-scales (i.e. multi-resolutions). The efficiency and effectiveness are achieved by multi-scales and the use of the distance transform. Details of the algorithm can be found in [40].

4.2.1 Spherical Stereo

A 3D scene point imaged in a spherical stereo pair has an angular disparity γ between two rays from the two projection centres to the scene point. Two new disparity measures introduced in [40] are radial disparity and normalized radial disparity. As shown in Figure 4.3, the radial disparity is defined as:

$$d = b/r_i \tag{4.4}$$

Where b is the baseline and r_i is the depth at view i. The normalized radial disparity is

$$\hat{d} = d/b = 1/r_i \tag{4.5}$$

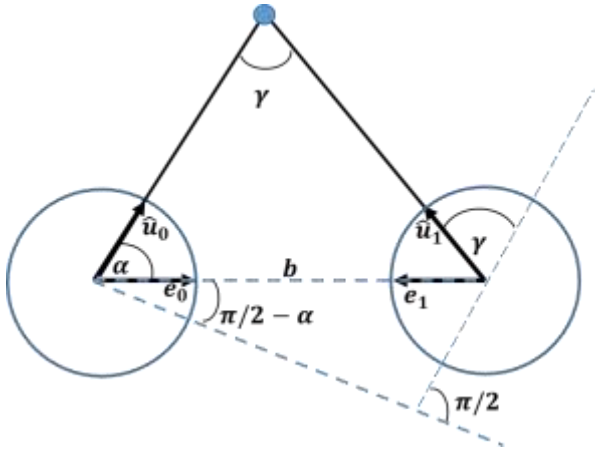


Figure 4.3 Epipolar geometry and spherical disparity [40].

The normalized radial disparity is not dependent on the baseline, and thus is convenient to use when fusing depth maps from multi-views, and is used in SVT. The relationship between the angular disparity and radial disparity can be expressed

$$\tan(\gamma) = \frac{dsin\,(\alpha)}{1 - d\cos(\alpha)} \tag{4.6}$$

Where α is the angle between $\hat{\boldsymbol{u}}_{\boldsymbol{i}}$ and the baseline.

For a given sampling direction in the reference $\hat{\boldsymbol{u}}_{\boldsymbol{0}}$, the corresponding sampling direction $\hat{\boldsymbol{u}}_{\boldsymbol{1}}$ in the matching image can be determined from the angular disparity:

$$\hat{\boldsymbol{u}}_{\boldsymbol{1}} = \hat{\boldsymbol{u}}_{\boldsymbol{0}}\cos(\gamma) + \hat{\mathbf{v}}\sin(\gamma) \tag{4.7}$$

Where

$$\hat{\mathbf{v}} = \hat{\mathbf{n}} \times \hat{\boldsymbol{u}}_1 \tag{4.8}$$

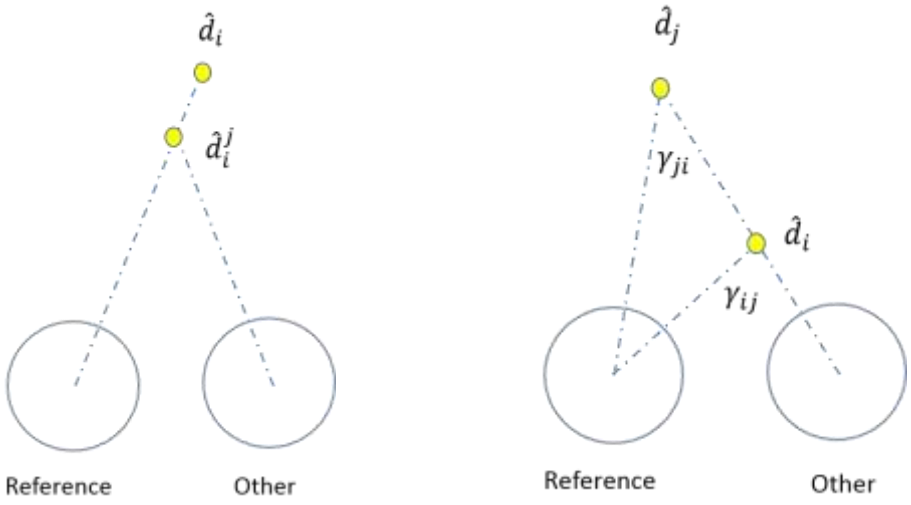
$$\hat{\mathbf{n}} = \hat{\boldsymbol{e}}_0 \times \hat{\boldsymbol{u}}_0 \tag{4.9}$$

4.2.2 Disparity Map Fusion and Filtering

SVT produces a disparity map (RD-map image storing normalized disparity defined in Section 4.2.1) after an initial spherical stereo matching and a confidence map after a cross-checking. The cross-checking compares the disparity map obtained using the first image of the stereo pair as the reference and the second as the matching image, to the disparity map obtained using the second of the stereo pair as the reference and the first as the matching image. The confidence map stores the scale of agreements between the two disparity maps for each pixel in an RD-map image. Initial disparity maps contain incorrect matches or missing data due to poorly textured regions or half-occlusions, and thus need to be refined to remove artifacts and fill holes. The fusion stage takes as input a set of disparity maps

and confidence maps, and outputs a fused disparity map and a fused confidence map, after removing outliers (incorrect matches).

The stability-based fusion algorithm was proposed by Merrell et al. [99] and adapted to the spherical disparity by Brunton [40]. One view is selected from *N* input views as a reference and warp all the disparity maps from the other views to the reference view. For each ray (or pixel), we have *N* disparity (or depth) estimates. We can evaluate each of the disparity estimates and find the optimal one in terms of stability. If we select Estimate #*k* from *N* estimates as the current estimate, we count the number of the estimates that are large than the current estimate, which is the number of the times the current estimate violates the occlusion (Figure 4.4.a). Then for each of the other views, we can check the matching ray to see if its disparity is smaller than the estimate disparity. If true, this is a violation of free-space (Figure 4.4.b).



(a) (b)

Figure 4.4 Illustration of occlusion (a) and free-space violation (b).

The stability is measured as the difference between the number of occlusion violations and the number of free-space violations. The fused disparity is the disparity estimate with the smallest stability greater or equal to zero. This is done independently for each ray (or pixel) for the reference view. The same steps are repeated for each of the views available.

The near viewpoint (high-disparity) outliers are removed after the fusion stage. Low-disparity outliers and small-scale noise remain in the disparity maps and can cause visually unpleasing distortions for close surfaces. A redundant wavelet transform is incorporated in the pipeline by Brunton [40] toward this end. This stage can be used to produce smooth results in view synthesis.

4.3 Experimental Evaluation

In the following section, we will show camera geometric calibration results. We will demonstrate that the calibration is able to find a sufficient number of feature point matches for the accurate determination of rotational and translational matrices. Then we will compare two radiometrically invariant matching cost functions for stereo matching. Finally, we will show the disparity maps obtained after different stages of SVT.

4.3.1 Geometric Calibration and Alignment

In order to test the geometric calibration, we use the cylindrical images captured at three different positions. First we use the images with the same exposure for calibration. The three cylindrical images with the same exposure captured at three different positions are shown in Figure 4.5. As can be observed, there are big changes in camera pose for three images captured. The three images are fed to the CaliPano tool for calibration and a rotational matrix and a translational vector are found for each position. Once the rotational

matrices are found, the Create_Pano_Ladybug tool is used to create rotationally aligned RD-map images (Figure 4.6).

Then we test with the three images (Figure 4.7) captured with different exposures (2 stops in exposure difference). After the calibration and rotational alignment, the aligned RD-map images are shown in Figure 4.8. Comparing Figure 4.6 and Figure 4.8, we can observe that exposure differences do not affect the image alignment. It seems that the calibration and alignment are robust to exposure difference. More evaluation will be presented in Chapter 6.

Figure 4.5 Cylindrical images captured at three positions with the same exposure used as input for geometric calibration.

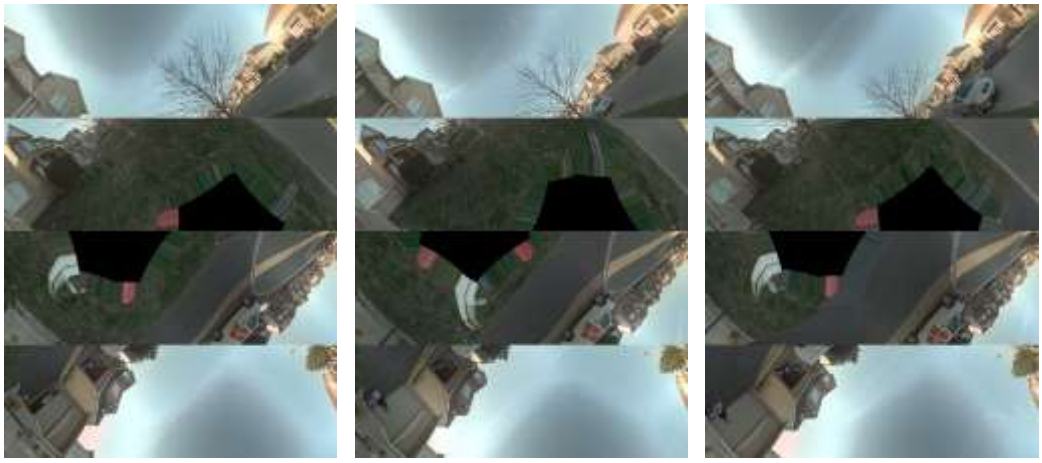


Figure 4.6 Aligned RD-map images with the same exposure at three positions.

Figure 4.7 Cylindrical images captured at three positions with different exposures used as input for geometric calibration.

Figure 4.8 Aligned RD-map images with different exposures at three positions.

Table 4.1 includes the numbers of the feature points detected for the images capture at three positions with the same exposure and the different exposures. As can be seen, the low exposed image has fewer feature points detected, and high exposure has more points detected. Although the saturated regions may produce fewer feature points, the grass field with increasing exposure can lead to more feature points. The numbers of matched feature points between image pairs are greatly affected by the exposure difference as shown in Table 4.2.

Table 4.1 Numbers of feature points detected in each of three images captured at different positions with the same exposure and different exposures.

Position #	Same exposure	Different exposure
0	1417	937
1	1352	1352
2	1301	1572

Table 4.2 Numbers of matched feature points detected between image pairs captured at three positions with the same exposure and different exposures.

Position #	Same exposure	Different exposure
0, 1	334	125
1, 2	342	151

However, the geometric calibration is not affected as can be seen by comparing Figure 4.6 and Figure 4.8 because we only need a small number of good matching points for camera calibration.

4.3.2 Multi-View Stereo Matching

The normalized cross-correlation related algorithm may be one of the most common methods for handling variation in brightness for stereo matching of images with illumination or exposure variation [35]. As discussed in Hirschmuller [76], mutual information is one of the best cost functions for handling illumination variation. The mutual information used as a matching cost function is defined as a function of entropies of two image blocks (I_1, I_2) and their joint entropy (H_{I_1,I_2}) as

$$MI_{I_1,I_2} = H_{I_1} + H_{I_2} - H_{I_1,I_2} \tag{4.10}$$

For handling exposure difference, the use of gradient field has been attempted in optical flow in the state-of-the-art work by Zimmer et al. [28]. Brunton et al. [41] has proposed as cost function the weighted sum of absolute differences in the radiometric adjustment image and the gradient image derived from the original images. The radiometrically adjusted image is derived from subtracting each pixel from a local mean. Therefore, a simple

comparison was conducted to compare the performance of the radiometric adjustment and gradient approach, and the mutual information approach.

Figure 4.9 shows three RD-map images captured from different views with different exposures. Figure 4.10 displays the effect of the radiometric adjustment and the three images look much similar after the adjustment.

The disparity maps generated using the radiometrically adjusted and gradient based approach and the mutual information approach are shown in Figure 4.11 and Figure 4.12. The disparity maps look very close although there are fewer black regions in the disparity map produced by the radiometric adjustment and gradient approach than that by the mutual information. In the work of this thesis, the radiometric adjustment and gradient approach has been used.

Figure 4.9 RD-map images with three different exposures. From left to right: low, medium and high exposures.

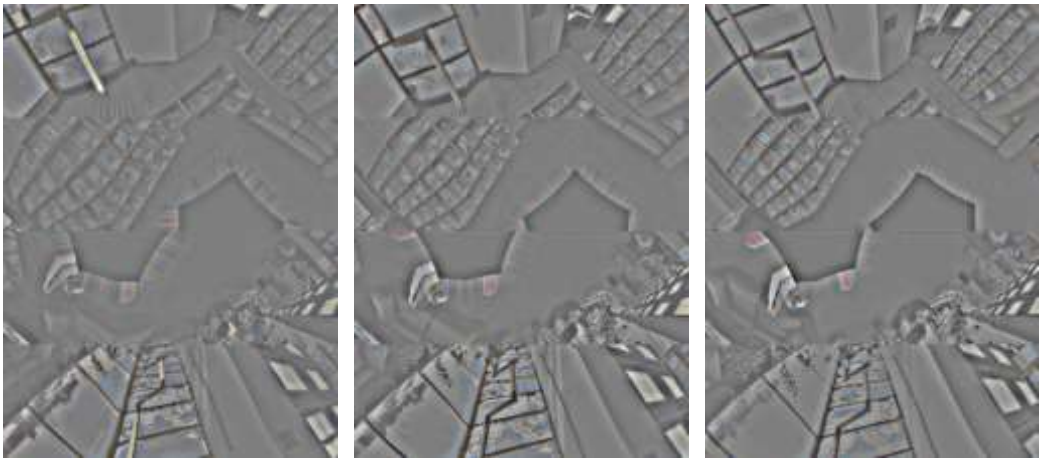


Figure 4.10 Radiometrically adjusted images. From left to right: low, medium and high exposures.

Figure 4.11 Disparity map derived using radiometrically adjusted and gradient match method. The color bar on the right indicates the minimum to the maximum disparity values upwards.

Figure 4.12 Disparity map derived using mutual information method. The color bar on the right indicates the minimum to the maximum disparity values upwards.

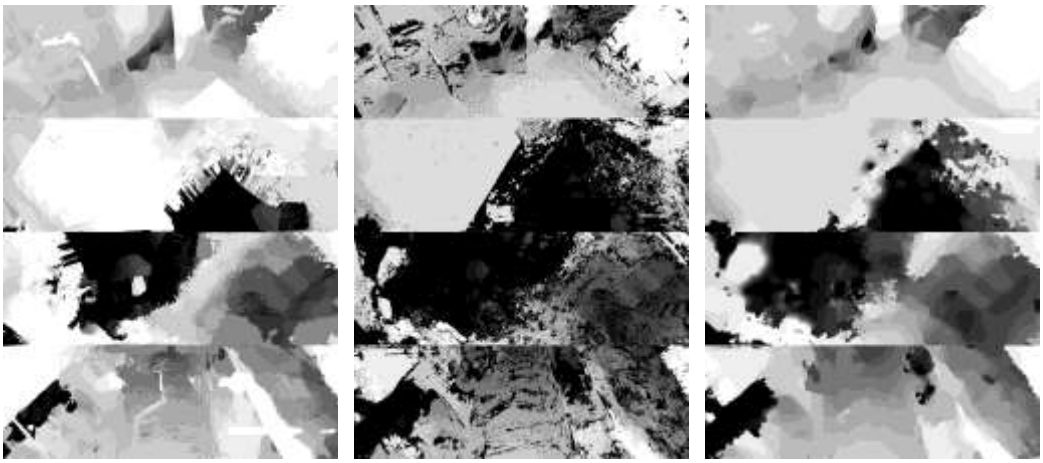


Figure 4.13 Disparity maps after different stages of SVT. From left to right: pairwise matching, fusion and filtering. The brightness from black to white indicates increasing disparity values.

To illustrate the different stages of SVT, Figure 4.13 shows the disparity maps generated after each of the three stages in SVT: initial stereo matching, fusion and filtering. The fusion stage removes invalid matches but leaves more holes. The filtering stage seems to smooth out steep disparity changes.

4.4 Summary

The multi-scale omnidirectional stereo matching framework of Brunton [40] has been introduced and related tools are described. An experimental evaluation has been presented in this chapter. The framework includes geometric calibration and alignment and omnidirectional stereo matching. To extend to multi-exposure and multi-view HDR imaging, we need to evaluate the impact of exposure differences and saturated regions on both calibration and stereo matching, and the problems with disparity maps. Possible techniques for disparity map enhancement for HDR imaging will be addressed in next chapter.

Chapter 5

Omnidirectional HDR Imaging

This chapter describes the temporal and spatiotemporal HDR imaging pipelines in detail. Some results will be presented to show the flow of the HDR pipelines and more results and comparisons will be given in next chapter.

Although our work is focused on multi-view HDR imaging, the experimental design was made to allow us to create both temporal and spatiotemporal HDR images. On one hand, the temporal HDR images can serve as the ground truth data for the assessment of multi-view HDR results. On the other hand, some algorithms used are the same for both temporal and spatiotemporal HDR imaging, in particular, radiance conversion, HDR fusion and tone mapping, and those algorithms can be tested easily using temporal HDR data.

Section 5.1 will describe some quality metrics commonly used for the assessment of image quality with attention to the comparisons of tone mapped HDR images. Section 5.2 will cover the temporal HDR imaging pipeline. Section 5.3 will explain the spatiotemporal HDR pipeline. A summary will be given in Section 5.4.

5.1 Image Quality Metrics

To evaluate the performance of different HDR imaging methods, we need some image metrics to assess the quality of tone mapped HDR images. The peak signal-to-noise ratio (PSNR) may be one of the simplest ones. But it may not correlate well with the perceived quality by human visual system (HVS). The SSIM (Structural SIMilarity) proposed by Wang et al. [112] is an HVS based metric and is widely used for image quality assessment [113]. Mantiuk et al. [114] proposed a visual metric, HDR-VDP-2, that compares images

pair to measure the visible difference and quality degradation with respect to the reference image. According to Rufenacht [36], HDR-VDP-2 seems to be better than SSIM but it is very slow. It is implemented only in MATLAB. We like to have some good metric to be implemented in our HDR pipeline in C/C++. Thus, the SSIM index is included in the HDR pipeline and used to measure the quality differences of tone-mapped HDR images.

5.1.1 Peak Signal-to-Noise Ratio

For the assessment of image quality, some metrics can be used to facilitate a more quantitative comparison. One common metric is the peak signal-to-noise ratio (PSNR) which can be expressed as

$$PSNR = 10log_{10}\left(\frac{S_{max}^2}{MSE}\right) \tag{5.1}$$

Where S_{max} is the peak value of the signal, e.g. 255 for an 8-bit image. The mean squared error (MSE) may be defined for gray images of M by N as:

$$MSE = \frac{1}{M \times N}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(I_{ij} - \hat{I}_{ij})^2 \tag{5.2}$$

where I_{ij} denotes the original or the ground truth image and $\hat{I}_{ij}$ the distorted image.

PSNR measures pixel level errors and may not correlate well with the perceived quality by the human visual system (HVS). For instance, PSNR shows big errors when an image is shifted by a few pixels or rotated slightly.

5.1.2 Structural Similarity (SSIM)

Among the HVS based metrics, the SSIM (Structural SIMilarity) index proposed by Wang et al. [112] is one of the most commonly used. The basic concept behind SSIM assumes

that similarity measurements can be expressed as a function of three comparisons: luminance, contrast and structure as

$$S(x,y) = f(l(x,y), c(x,y), s(x,y)) \tag{5.3}$$

Where x and y are two nonnegative image signals. $l(x,y)$, $c(x,y)$ and $s(x,y)$ are luminance, contrast and structure functions, respectively. The SSIM proposes to use the following weighted product

$$SSIM(x,y) = l(x,y)^{\alpha}\, c(x,y)^{\beta}\, s(x,y)^{\gamma} \tag{5.4}$$

The above expression can be simplified into the common form for the SSIM index as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5.5}$$

Where C_1 and C_2 are constants. μ_x is the average of *x* and μ_y of *y*. σ_x is the variance of *x* and σ_y the variance of *y*. σ_{xy} is the covariance of *x* and *y*. Those components can be evaluated from the following equations with a window of size *N*:

$$\mu_x = \frac{1}{N}\sum_{i=0}^{N-1} x_i \tag{5.6}$$

$$\mu_y = \frac{1}{N}\sum_{i=0}^{N-1} y_i \tag{5.7}$$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=0}^{N-1} (x_i - \mu_x)^2 \tag{5.8}$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=0}^{N-1} (y_i - \mu_y)^2 \tag{5.9}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=0}^{N-1} (x_i - \mu_x)(y_i - \mu_y) \tag{5.10}$$

5.2 Temporal HDR imaging

5.2.1 HDR Pipeline

To generate an HDR image from multi-exposed images captured from the same viewpoint, we employ the temporal HDR imaging pipeline shown in Figure 5.1. First, rotationally aligned RD-map images are converted into LDR radiance maps using the camera response curves and exposure times. Then the LDR radiance maps are fused into one single HDR radiance map. Finally, the HDR radiance map is tone mapped to an LDR image for display. The pipeline was implemented in C++.

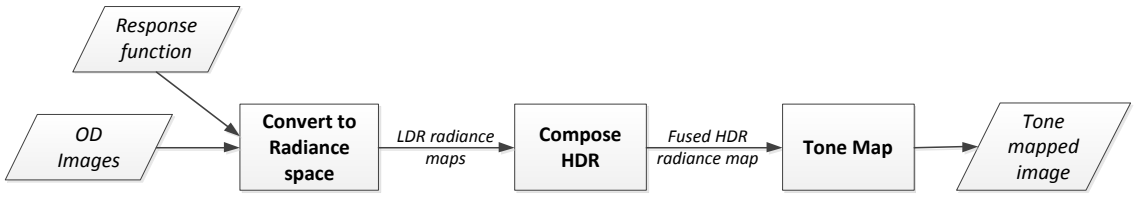


Figure 5.1 Temporal HDR imaging pipeline

5.2.2 HDR Composition

The multiple radiance maps are fused into one single HDR radiance map using the weighted HDR fusion method proposed by Ward et al. [29]. C++ code similar to the MATLAB pseudo-code in [95] was realized for the HDR fusion but the handling of saturated pixels was modified. For temporal HDR imaging, we can assume the images are captured using a stationary camera of a scene with no or few small moving objects. In addition, we also assume the scene illumination is constant.

Considerable research efforts have been made to develop algorithms for the detection and removal of artifacts caused by moving objects in the fused HDR images [13]. As the ghosting effects are caused by the presence of multiple occurrences of moving objects, ghosts can be removed by keeping single occurrence of moving objects or using weighted means of all images. The weighting approach is simple and works well if the number of shots are large and the moving objects are small. As our images contain only small moving objects in some of the exposures, we use the weighting approach in the composition of the temporal HDR images.

In addition, the radiance values near over- and under-exposed pixel values are filled with more noise and are less reliable than the rest of the image. Thus, we use a similar approach as suggested by Silk [26] and exclude the pixels with the values under 10 and over 245 in the HDR composition.

5.2.3 Tone Mapping

In order to display HDR radiance maps, we need to convert them into displayable 8-bit images. Tone mapping operators can be used to compress HDR radiance maps into 8-bit LDR images. Reinhard's global tone mapping operator [59] (covered in Section 2.2.2) is simple and produces images with consistent mapping that makes the comparison of HDR imaging methods easy. The similar procedures to the MATLAB implementation [95] of the Reinhard's global tone mapping operator were implemented in C++ in our HDR pipeline.

5.2.4 Evaluation

A set of omnidirectional LDR images of an outdoor scene are shown in Figure 5.2. The software tool of our own implementation was used to generate the temporal HDR images from the LDR images. If the pixels with one or two channels are saturated, they cause the artifacts shown in the top image in Figure 5.3 . This effect is more profound in the bright

regions. As can be seen in the bottom image, the artifacts are removed when the pixels with saturated channels (pixel value >250) are excluded from the HDR image.

As our C++ implementation of the HDR fusion and the tone mapping is based on similar methods used by Eitz and Stripf [95] in their MATLAB implementation, we can verify our implementation by compare our result with the result generated from their MATLAB code. As can be seen, our tone mapped HDR image (Figure 5.3.b) looks very similar to the result generated with their MATLAB code (Figure 5.4).

Figure 5.2 Exposure bracketed LDR images captured from single view.

(a)

(b)

Figure 5.3 Impact of over- or under-exposed pixels on the HDR fusion. The artifacts in the top image (a) can be removed by excluding the pixels with one or more saturated channels in HDR fusion (b).

Figure 5.4 Tone mapped HDR image created with the MATLAB implementation of Eitz and Stripf [95]

5.3 Spatiotemporal HDR Imaging

5.3.1 HDR Pipeline

For the recovery of HDR from LDR images captured from multiple views, we use the spatiotemporal HDR pipeline shown in Figure 5.5. The pipeline consists of six stages: stereo matching, disparity map enhancement, image warping, radiance conversion, HDR composition and tone mapping. As the SVT based stereo matching has been covered in Chapter 4 and the radiance space conversion and tone mapping are the same as for temporal HDR imaging, we will cover only three of the six stages: disparity map enhancement, image warping and HDR composition.

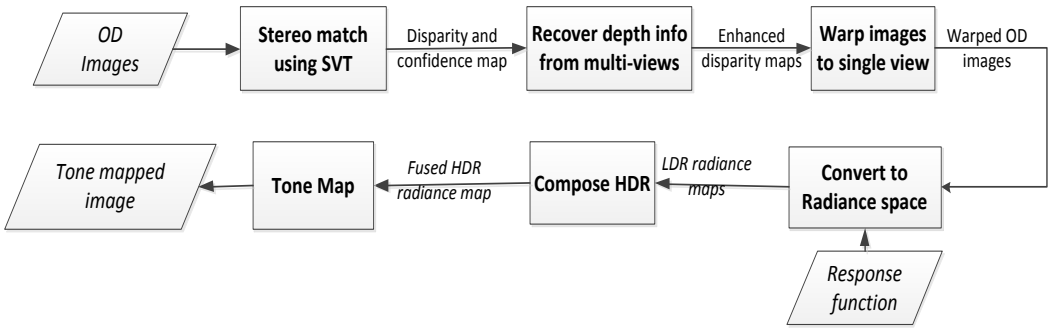


Figure 5.5 Spatiotemporal HDR imaging pipeline

5.3.2 Disparity Map Enhancement

Before the HDR fusion stage, disparity maps are required to warp images from the neighbouring views to one target view for which an HDR image will be created. The stereo matching stage in SVT produces initial disparity maps and confidence maps. Because of exposure difference and saturated regions, disparity maps may contain large regions with invalid disparity values even after the fusion and filtering stages in SVT. For instance, an

outdoor scene can have a large sky region lacking textures for stereo matching which results in large regions with no valid disparity information. The following techniques have been explored to improve the disparity maps:

- Remove invalid disparity values based on the confidence maps generated from the disparity cross-checking algorithm in SVT.
- Detect sky regions based on the prior information of the sky color and fill in with the minimum disparity value for the sky regions identified.
- Inverse distance weighted interpolation to fill in holes based on superpixels found in color images using the SLIC tool developed by Achanta et al. [115] .

Based on the confidence maps obtained using SVT, we can filter out invalid disparity values from the disparity maps. In addition, we can fill in holes in each superpixel based on the inverse distance weighted interpolation. The SLIC (Simple Linear Iterative Clustering) method developed by Achanta et al. [115] seems to produce pixel clusters respecting salient boundaries. Because most disparity discontinuities align with object boundaries, we can perform disparity interpolation only inside superpixels without averaging out the disparity changes across object boundaries. A superpixel technique has also been used in depth synthesis by Chaurasia et al. [116].

For an outdoor scene, it is difficult to obtain disparity information for sky regions and some uses the information about the location of sky regions in typical images as a guide for disparity map improvement [116]. It is more complex to use the location information to identify sky regions in RD-map images. As sky regions are normally smooth and contain some unique color [117], we use the color information combined with the superpixel technique for disparity map enhancement for the sky regions. The color model for the sky is based on the following assumptions:

- A pixel in a sky region should be brighter than the average pixel.

- The red value of the pixel should be less than the blue and the green components.
- The difference between the green and the blue components should be small.

We first convert RGB color space into normalized RGB space (rg chromaticity space) as

$$r = R/(R + G + B) \tag{5.11}$$

$$g = G/(R + G + B) \tag{5.12}$$

$$b = B/(R + G + B) \tag{5.13}$$

A pixel is a sky pixel if the pixel is bright than the average pixel in the image and the following relations hold:

$$(r < g)\ \&\&\ (r < b)\ \&\&\ (0.25 < r < 0.32) \tag{5.14}$$

and

$$abs(g - b) < 0.04 \tag{5.15}$$

The color model for sky is based on color analysis of our captured data which do not contain direct sunlight. Thus, the model is limited only to similar scenes. For any superpixel with a percentage of sky color pixels over certain threshold, we can classify the superpixel as part of a sky region and fill in with the minimum disparity value.

To demonstrate the proposed disparity enhancement method, two images with different exposures captured at two positions are used as shown in Figure 5.6. The segmented images with superpixels are shown in Figure 5.7. As can be seen, the superpixels found indeed respect the boundaries of houses. The initial and enhanced disparity maps are compared in Figure 5.8 and a considerable improvement is made in the enhanced disparity map. The warped images based on the initial and enhanced disparity maps are compared in Figure 5.9. The warping method is forward warping which will be covered in next section. As can be seen, the quality of the warped image based on the enhanced disparity map shows few holes although there is still room for a further improvement.

Figure 5.6 Differently exposed images from two different views.

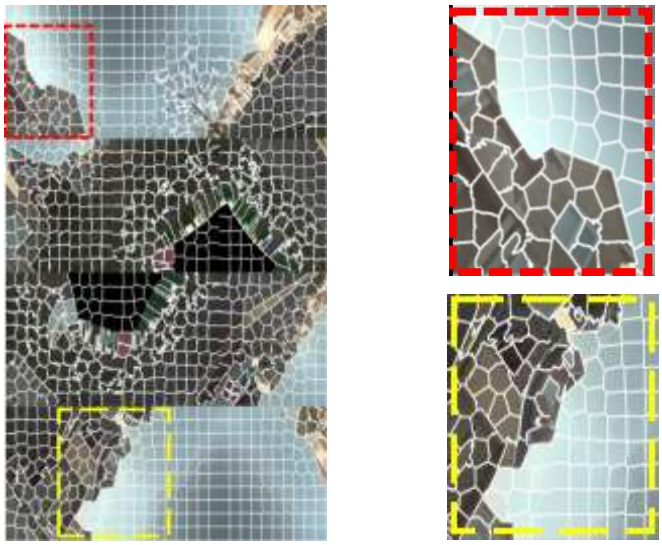
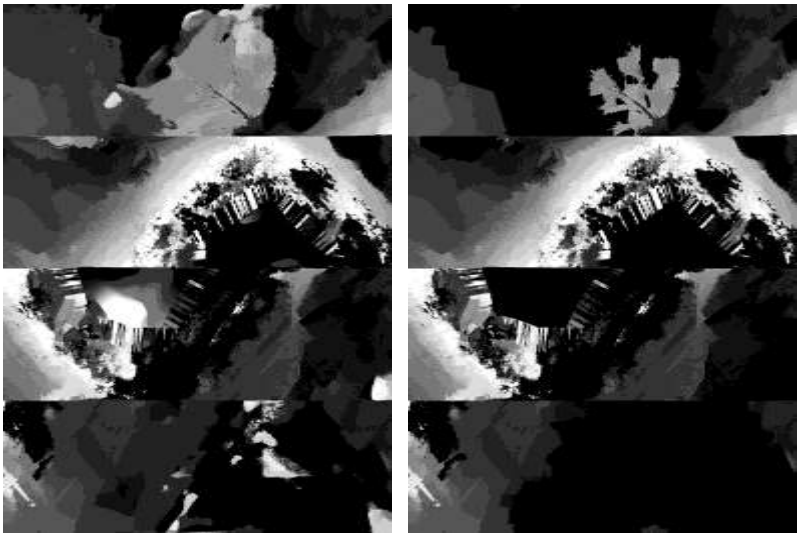


Figure 5.7 Segmented color (whole and partial) images used to guide disparity map enhancement.



(a) (b)

Figure 5.8 (a) Initial disparity map obtained using SVT; (b) Enhanced disparity map with the proposed algorithm.

(a) (b)

Figure 5.9 Warped images based on (a) initial disparity map and (b) enhanced disparity map.

5.3.3 Disparity-based Image Warping

To combine the radiance maps from LDR images captured from different views, we need to warp the images to a common target view. Image warping is a common technique in computer graphics [118] and disparity-based image warping has been used in many applications [119]. If the disparity maps are available, the images capturing from other views can be warped to the target view selected. Depending on the view with disparity map, warping methods are divided into forward and backward warping. The algorithms for the forward and backward image warping were added to SVT and implemented in CUDA C/C++ based on the epipolar geometry for spherical stereo matching in Figure 5.10. CUDA (Compute Unified Device Architecture) created by NVIDIA provides a parallel computing platform and programming model [120]. CUDA C/C++ API facilitates the parallel programming and allows users to use GPU (graphic processing unit) for general purpose programming. As GPU is growing more powerful with hundreds of cores, the spherical stereo matching and HDR imaging can be sped up and the core functions in SVT are implemented in CUDA C.

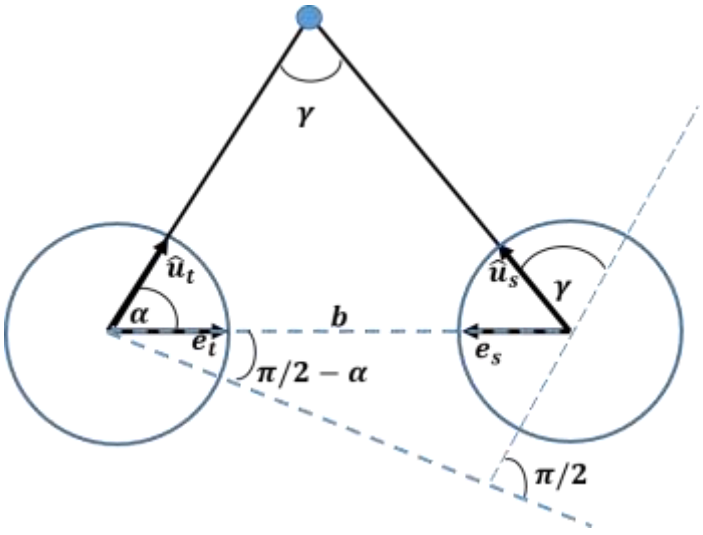


Figure 5.10 Epipolar geometry and spherical disparity used in SVT [40]. The ray vectors can be mapped between the two views if disparity and baseline information is known.

Forward Warping

Forward warping refers to the warping of a source image to the target view based on the disparity map for the source view. The pseudo-code for the algorithm is shown in Figure 5.11. If the disparity map from the source view is available, we can find the disparity value and the color for each pixel in the source image. We can convert the pixel location in the RD-map to a 3D ray vector using the RD-map function. Based on the epipolar geometry shown in Figure 5.10, we can find the corresponding ray vector for target view (see Section 4.2.1 for details). Then we can convert the 3D ray vector to the pixel location in the target RD-map image, and transfer the color information from the source to the target. Forward warped images may show holes because of invalid disparity information or many-to-one mapping. Further post-processing is required. For small and scatted holes, disparity values in neighboring valid pixels can be used to fill holes.

Input: RD-map Image I_r from a target view, RD-map Image I_n and RD-map Disparity map D_n from a source view

Output: Warped image $\hat{I}_{n\to r}$ to the target view

Procedure for Disparity-Based Forward Image Warping:

for each pixel (r_n, x_n, y_n) in D_n ***do***

Disparity d =D_n (r_n, x_n, y_n)

Convert (r_n, x_n, y_n) to 3D ray $\vec{u}_n$

Derive the ray $\vec{u}_r$ for the reference view using the equations in Section4.2.1

Find RD coordinates (r_r, x_r, y_r) for $\vec{u}_r$

Set $\hat{I}_{n\to r}(r_r, x_r, y_r) = I_n(r_n, x_n, y_n)$

end for

Figure 5.11 Disparity-based forward warping algorithm.

Backward Warping

If the disparity map for the target view is used to find the correspondence between the source image and the target image, the warping algorithm is denoted as backward warping as shown in Figure 5.12. If the disparity map from the target view is known, we find the disparity value for each pixel in the target view. We can convert the pixel location in the RD-map to a 3D ray vector using the RD-map function. Based on the epipolar geometry shown in Figure 5.10 (see Section 4.2.1 for details), we can find the corresponding ray vector for the source view. Then we can convert the 3D ray vector to the pixel location in the source RD-map image, and transfer the color information from the source to the target.

Input: RD-map Image I_r and RD-map disparity map D_r from a target view, and RD-map Image I_n from a source view

Output: Warped image $\hat{I}_{n\to r}$ to the target view

Procedure for Disparity-Based Backward Image Warping:

for each pixel (r_r, x_r, y_r) in D_r ***do***

Disparity d =$D_r(r_r, x_r, y_r)$

Convert (r_r, x_r, y_r) to 3D ray $\vec{u}_r$,

Derive 3D ray $\vec{u}_n$ for the source view using the equations in Section4.2.1

Find RD coordinates (r_n, x_n, y_n) for $\vec{u}_n$

Set $\hat{I}_{n\to r}(r_r, x_r, y_r) = I_n(r_n, x_n, y_n)$

end for

Figure 5.12 Disparity-based backward warping algorithm.

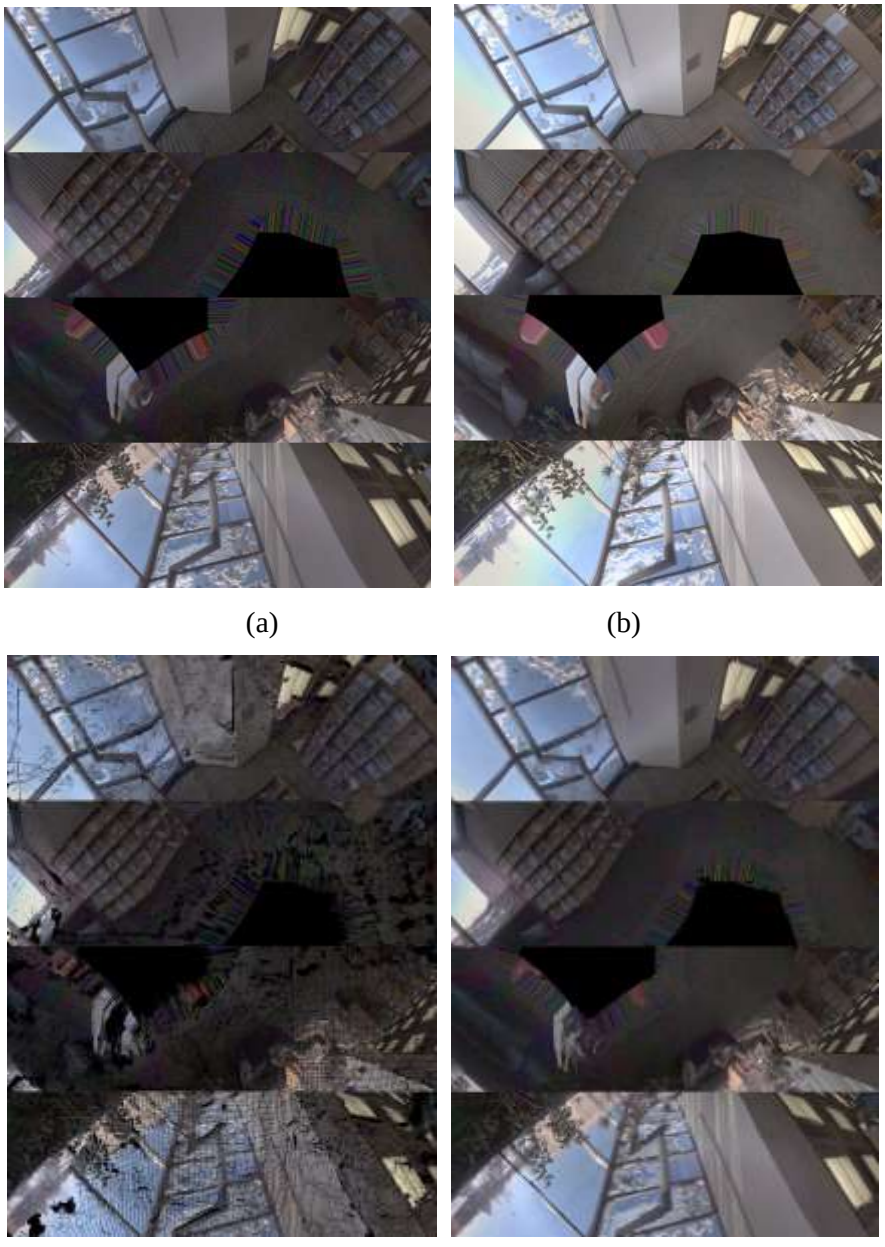
Comparison

Two images captured from different views and exposures are shown in Figure 5.13.a and Figure 5.13.b. The disparity maps for two views found using SVT are used to perform the forward and backward warping using the algorithms described above. The forward warping algorithm produces images with some holes whereas the backward warping algorithm generates a smooth image without noticeable holes. Thus, we will use mainly the backward warping algorithm for the spatiotemporal HDR imaging framework although the forward warping algorithm will be compared for some cases. The forward warping method is useful when there is no disparity map available for a target view. One such application is new view synthesis.

5.3.4 HDR Composition

Once we have the reference image and the images warped from the neighboring views, we can convert them into radiance space and compose them to obtain HDR in radiance space. We use the weighted mean for each pixel in the HDR image similarly as in temporal HDR imaging and exclude the saturated pixels.

When fusing an HDR image from a sequence of LDR images captured using a stationary camera, the variation in the entropy, intensity or radiance value can be used to detect the moving pixels. In the proposed spatiotemporal HDR imaging framework, we use stereo matching to find disparity maps used to register images. Stereo matching assumes static scenes. Therefore, the proposed framework deals only with camera movement and no moving objects are explicitly handled. We may include optical flow to handle moving objects in future. But if the moving objects can be manually marked in black, the objects can be removed in the HDR fusion. In our approach, we fuse the current image for the target view and two or more images warped from the other views. Saturated pixels will be removed from the final HDR images.



(a) (b)

(c) (d)

Figure 5.13 Comparison of disparity-based image warping. Two different exposed images (a-b) from two views and the image from the first view warped to the second view: forward warping (c) and backward warping (d).

5.4 Summary

This chapter has described the pipelines for the temporal and the spatiotemporal omnidirectional HDR imaging. We have explained the major stages in the temporal HDR imaging pipeline: radiance conversion using the inverse camera response curve, HDR composition in radiance space, and tone mapping of HDR radiance maps to displayable images. We have also described in detail our spatiotemporal HDR pipeline. For spatiotemporal HDR imaging, an image registration is required to carry out on images from different views. We have proposed to use disparity maps to warp images to a selected target view from other views and then use the current and warped images to generate HDR images. Two warping algorithms: forward and backward have been proposed and implemented. We have explored some possible techniques for improving disparity maps.

Chapter 6

Results

In this chapter we will show more results with different datasets captured indoors and outdoors, compare the different HDR imaging methods, and evaluate different factors influencing the quality of the HDR results. In addition to quantitative comparison in dynamic range of radiance, we will also assess the quality of HDR images by visual inspection and an image quality metric: the SSIM (Structural Similarity) index (see Section 5.1).

Section 6.1 describes the two of the datasets captured: one for indoors and one for outdoors, which are used for the HDR imaging evaluation. The results of geometric calibration for each capture position will be presented for the two datasets. In addition, the numbers and percentages of under- and over-exposed pixels will be calculated for different exposures.

Temporal HDR is the basic method for recovering HDR. Some HDR processing steps such as radiance conversion and tone mapping are used both in the temporal and the spatiotemporal HDR approaches. Thus we will start with the presentation of the HDR imaging results in Section 6.2. The HDR images generated and the dynamic ranges recovered will be shown for an indoor scene and an outdoor scene.

Section 6.3 will present the results of the spatiotemporal HDR imaging approach. As the spatiotemporal HDR approach relies on disparity maps for pixel correspondence, the quality of the HDR images depends strongly on disparity maps. On one hand, capturing the full dynamic range of a high contrast scene requires a sufficiently large difference in exposures. On the other hand, an increasing exposure difference will lead to less accurate

disparity maps and will result in more artifacts in the final HDR images. Different options will be discussed and compared in terms of dynamic range and image quality.

We are not aware of a suitable omnidirectional HDR imaging benchmark for comparison. However, the recent PatchMatch based HDR methods proposed by Hu et al. [43] and Sen et al. [42] seem to produce plausible results even in the presence of camera movement and object movement. We will compare these methods to our approach in Section 6.4. Finally, we will have discussions in Section 6.4 and a summary in Section 6.5.

6.1 Datasets

Good quality LDR images of high contrast scenes are very important for the evaluation of HDR imaging techniques. As we are not aware of any benchmark dataset available for the omnidirectional HDR imaging, we have captured the omnidirectional LDR images of a number of indoor and outdoor scenes of high brightness contrast.

For the indoor scene selection, we have tried to choose indoor rooms of large size for multiple captures, and with windows for high dynamic range, and with sufficient textured regions for stereo matching. The indoor scenes include a public library, and the halls in a community centre.

As for the outdoor scene selection, we need to consider the limitation of the Ladybug 2 camera set. It is found that CCD smear and veiling occur when the sensors of the Ladybug 2 camera set face directly strong sunlight. Thus we chose to capture outdoor scenes in mornings or afternoons when the sun did not directly shine into the sensors which implies the true dynamic range of an outdoor scene in a typical sunny day can be higher than what we have captured.

In the following sections, the two datasets used for the HDR imaging evaluation, one for an indoor scene and one for an outdoor scene will be described and analyzed.

6.1.1 Public Library

The library scene captured in this dataset is located inside the Ruth E. Dickinson public library in the Walter Baker community centre, Nepean. There are large side windows, bookshelves, sofas and tables in the scene. Nine captures were performed each at a different position (about 30 cm apart) with the Ladybug 2 camera set placed steadily on a tripod. At each position, 26 images were captured with the exposure time starting from the shortest shutter time of 0.117 *ms* in an increment of 1/3 stops. The gain was set to 9 dB for an indoor scene.

As the camera set was placed on a tripod, the camera movement was approximately on the same plane. This can be seen from Table 6.1 as the Z (upwards) coordinates of the translational vectors are nearly zero for the nine capture positions. Thus, the rotational angles are estimated approximately as rotations around the Z-axis in the reference frame. The rotational matrix can be simplified as

$$R = \begin{bmatrix} \cos\theta & -sin\theta & 0 \\ sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.1}$$

The rotational angle can be found from the rotational matrix obtained from the camera calibration.

Table 6.1 shows the rotational angles and translational vectors (up to a scale) for the nine capture positions inside the Ruth E. Dickinson public library based on the geometric calibration using the CaliPano tool. Please note that the 3D reconstruction, e.g. rotational matrix and translational vectors, based on the shape from motion approach, are up to an arbitrary scale [62]. Both the translations and rotations are evaluated with respect to the world coordinate frame set to be the same as the coordinate frame at the first capture position. Thus, the rotational angles and translational vectors are relative to the camera coordinate frame at Position #0.

Table 6.1 Rotational angles and translational vectors of nine capture positions with respect to the first position for the Ruth E. Dickinson library scene (The calibration is up to a scale but the observed translation is about 30 cm between views).

Position #	**Rotation angle (degree)**	**Translational vector (up to a scale)**		
		X	*Y*	*Z*
0	0	0	0	0
1	12.8	0.45	-0.902	-0.0023
2	20.6	0.6389	-1.6811	-0.0086
3	35.1	0.9623	-2.5722	-0.0062
4	42.7	1.1258	-3.2611	-0.0068
5	48.3	1.5149	-3.8758	-0.0054
6	49.4	2.3090	-4.8016	0.02112
7	70.6	3.3546	-5.9600	0.0449
8	87.7	4.0271	-6.9817	0.0649

In HDR imaging, we use different exposure times to capture the different ranges of scene radiance values. Reducing exposure time captures high radiance values whereas increasing exposure time recovers low radiance values. The exposure time is long enough if there is no under-exposed region, and it is short enough if there is no over-exposed region.

Table 6.2 lists the numbers and percentages of under- and over-exposed pixels counted in the images with the different exposure times selected in the HDR evaluation. As can be seen, the percentage of under-exposed pixels reduces to zero when the exposure time increases to 3.083 *ms* for Exposure #14. The percentage of over-exposed pixels reduces with a decrease in exposure time. At the lowest exposure time of 0.117 *ms*, the percentage of over-exposed pixels is down to 0.014%.

Table 6.2 Numbers and percentages of under- and over-exposed pixels in the cylindrical images of the Ruth E. Dickinson public library scene with different exposure times at Position #4. The image size is 1152x1536 pixels.

Exposure #	**Exposure time (ms)**	**Under-exposed pixels (<5 gray levels)**		**Over-exposed pixels (>250 gray levels)**	
		#	%	#	%
0	0.117	193819	10.953	251	0.014
2	0.194	128452	7.259	7391	0.418
5	0.350	42142	2.382	37206	2.103
7	0.584	17249	0.975	54856	3.100
9	0.975	1535	0.087	103735	5.862
12	1.912	42	0.002	219050	12.379
14	3.083	1	0.000	352526	19.923

6.1.2 Front Yard

This dataset was captured in the front yard of a house in an afternoon. Eight captures were performed each at a different position (about 30 cm apart) with the Ladybug 2 camera set placed steadily on a tripod. Although the number of capture positions for this dataset is less than that for the indoor dataset by one, there is no impact on the HDR image comparison as we use only the three neighboring image. At each position, 26 images were captured with the exposure time increasing from the shortest shutter time of 0.117 ms in an increment of 1/3 stops. The gain was set to 0 dB for an outdoor scene.

Table 6.3 shows the rotational angles and the translational vectors (up to a scale) for the eight capture positions in the front yard scene based on the geometric calibration performed using the CaliPano tool. Please note that the 3D reconstruction, e.g. rotational matrix and translational vectors, based on the shape from motion approach is up to an arbitrary scale [62]. The images of Exposure #7 taken from each of the eight capture positions are used for

the geometric calibration. As the camera set was placed on a tripod, the camera movement can be assumed to be on the same plane approximately.

This can be seen in Table 6.3 as the Z (upwards) coordinate of translational vectors does not change much at different positions. Thus, the rotational angles are estimated approximately as in-plane rotational angles from the calibrated rotational matrices similarly as for the indoor scene using Equation 6.1.

The over-exposed pixel count in the images can be used to assess if the exposure time is sufficiently short to capture the highest radiance of the scene. On the other hand, the under-exposed pixel count can be used to check if the exposure time is long enough to capture the lowest radiance. Table 6.4 lists the numbers and percentages of under- and over-exposed pixels in the images of the front yard scene with the exposure times selected for the HDR evaluation. As can be seen, the percentage of under-exposed pixels reduces to zero when

Table 6.3 Rotational angles and translational vectors of eight capture positions with respect to the first position for the front yard outdoor scene. The images of Exposure #7 at each capture position are used for the calibration.

Position #	**Rotational angle (degree)**	**Translational vector**		
		X	*Y*	*Z*
0	0	0	0	0
1	49.1	0.0109	-0.9988	-0.0473
2	3.6	-0.9945	-2.1536	-0.1264
3	-38.6	0.1114	-3.4422	-0.1891
4	19.9	0.6827	-4.7383	-0.2756
5	75.2	1.4701	-5.8735	-0.2771
6	129.2	2.2337	-6.7791	-0.2842
7	138.0	3.0952	-8.3211	-0.2545

Table 6.4 Numbers and percentages of under- and over-exposed pixels in the images captured at Position #4 for the front yard scene with different exposure times. The image size is 1152x1536 pixels.

Exposure #	**Exposure time (ms)**	**Under-exposed pixels (<5 gray levels)**		**Over-exposed pixels (>250 gray levels)**	
		#	%	#	%
0	0.117	345983	19.553	0	0.000
2	0.194	166566	9.413	12	0.001
5	0.350	63819	3.607	21914	1.238
7	0.584	18158	1.026	75180	4.249
9	0.975	2644	0.149	233272	13.183
12	1.912	42	0.002	589548	33.318
14	3.083	0	0.000	725491	41.000

the exposure time increases to 3.083 *ms* for Exposure #14. The percentage of over-exposed pixels reduces with a decrease in exposure time and is down to zero at the shortest exposure time of 0.117 *ms*.

6.2 Temporal HDR imaging

6.2.1 Library Scene

Compared to spatial or spatiotemporal HDR imaging approaches, temporal HDR imaging should be able to capture dynamic range of radiance more accurately and produce HDR images of higher quality. We will first evaluate the impact of exposure time range on the dynamic range of radiance recovered and the HDR image quality. Figure 6.1 shows the cylindrical images generated from the image data captured with the Ladybug 2 camera set for the Ruth E. Dickinson public library scene.

Since there is no under-exposed pixels after Exposure #14, only the images with Exposures #0 to #14 are included to recover the dynamic range of the scene radiance. The range of exposure time for Exposures #0 to #14 is 0.117-3.083 *ms* and the dynamic range recovered is 89-77897. If only Exposures #2 to #12 are selected, the exposure time range is reduced to 0.194-1.912 *ms* and the dynamic range is greatly reduced to 89-48187. If only Exposures #4 to #8 are used, the exposure time is 0.272-0.740 *ms* and the dynamic range is further reduced to 96-34369.

Figure 6.1 Eight cylindrical images from Exposures #0 to #14 at Position #5 of the Public library dataset. Exposure time range is 0.117-3.083 *ms* and dynamic range is 89-77897.

(a) Exposures #0 to #14

(b) Exposures #2 to #12

(c) Exposures #4 to #8

Figure 6.2 Tone mapped HDR images generated from the LDR images of different exposure range for the public library scene.

Figure 6.2 displays the tone mapped HDR images recovered from the LDR images in different exposure time ranges. The quality of HDR images may be affected if the exposure time range of the LDR images is reduced. Thus, the SSIM index is used to measure the quality change. If the tone mapped HDR image for Exposures #0 to #14 is selected as the reference, the SSIM index is 0.98 for the HDR image recovered from the images within Exposures #2 to #10, and 0.934 for HDR image with Exposures #4 to #8. Despite the

drastic change in dynamic range, there is no significant reduction in the quality of tone mapped HDR images. The reason may be that the number of high radiance scene points is too small to have a noticeable impact on image quality.

6.2.2 Front Yard Scene

We may expect a higher dynamic range of radiance of an outdoor scene than that of an indoor scene. Figure 6.3 shows the cylindrical images generated from the image data captured with the Ladybug 2 camera set for the front yard scene.

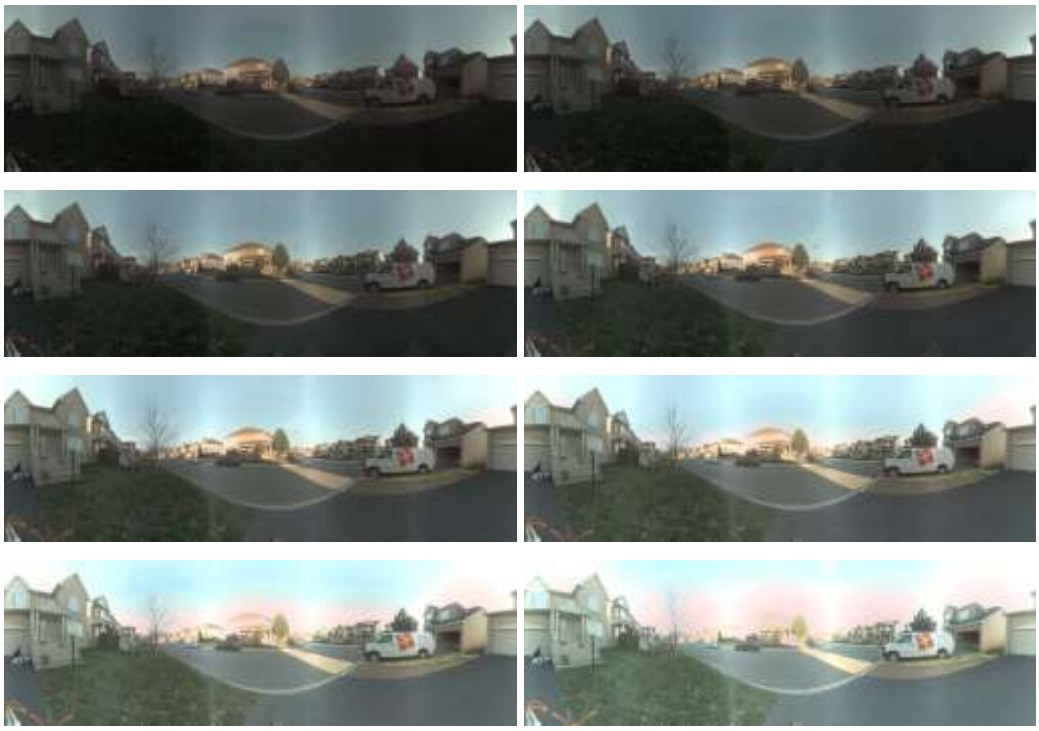


Figure 6.3 Eight cylindrical panoramas from Exposure #0 to #14 at Position #5 in the front yard outdoor dataset. Exposure time range is 0.117- 3.083 *ms.*

Since the over-exposed pixel count is zero in the image with Exposure #0 and the under-exposed pixel count is zero with Exposure #14, the exposure time range between Exposures #0 and #14 already covers the entire dynamic range of the scene. The range of exposure time for Exposures #0 to #14 is 0117-3.083 *ms* and the dynamic range recovered is 34-90668. If only Exposures #4 to #8 are selected, the exposure time range is reduced to 0.272-0.584 *ms* and the dynamic range is 41-65343.

Please note that the radiance is set to 1 when the exposure time is 1 *s* and the gray level is 128 in the camera response curve. The gain influences the reference point and the radiance values for the indoor and outdoor scenes are based on two different curves and they cannot be compared. As the gain is set to 0 dB, the radiance values for the outdoor scene are likely much higher than for the indoor scene presented in the last Section.

There are some artifacts observed in the cylindrical panoramic images (Figure 6.3). First, there are bright stripes along panoramic mosaic stitching regions. This artifact seems less noticeable in the RD-map images. Second, reddish colors appear in the saturated regions. This may be caused by the incorrect tone mapping in the generation of cylindrical panoramas dealing with the pixels with one or two channels saturated.

Figure 6.4 displays the tone mapped HDR images generated from the LDR images of different exposure ranges. Although the dynamic range is reduced by about one third, there is no noticeable difference by a visual inspection. If we use the tone mapped HDR image recovered from Exposures #0 to #14 as reference, the SSIM index of the tone mapped HDR images with Exposures #4 to #8 is 0.97. It is likely that the number of scene points within the high range of radiance is too small to have a significant impact on the image quality.

(a) Exposures # 0 to #14

(b) Exposures # 4 to #8

Figure 6.4 Tone mapped HDR images generated from LDR images of different exposure time ranges. Dynamic range measured is (a) 34-90668 and (b) 41-65343.

6.3 Spatiotemporal HDR Imaging

We will first show some examples of the forward and backward warping algorithms and their impact on the quality of the recovered HDR images. Then we will look into the effects of exposure increment step size, baseline and exposure sequence ordering. In addition, we would like to see how the proposed HDR framework performs with the outdoor data. Because the RD-map has more uniform sampling than any other panoramic formats such as cylindrical and equirectangular, SVT uses the RD-map for spherical stereo matching. Therefore, all panoramic images presented in the following sections are in RD-map. Please refer to Section 3.3 and Figure 3.8 for details.

6.3.1 Impact of Warping Method

To demonstrate the performance of the disparity-based forward and backward warping algorithms, three images of Exposures #2, #7 and #12 are taken from the Ruth E. Dickinson public library dataset for Positions #0, #1 and #2, respectively. To generate an HDR image for Position #1, we warp images from Positions #0 and #2 to Position #1. The images warped with the forward and backward warping algorithms are displayed in Figure 6.5 and Figure 6.6, respectively. As can be seen, the forward warping method produces an image with many non-mapped pixels (in small black holes or large black regions). Because the disparity maps are not perfect due to half-occlusion and mismatches, the correspondence mapping from a source view to the target view is not one-to-one. Thus, some pixels in the image warped are left black without color information. On the other hand, the backward warping algorithm generates a much smoother warped image. The reason is that each pixel in the warped image gets mapped to some pixel in the source image.

Once we have the warped images and one original image for Position #1, we convert them into LDR radiance maps and fuse the radiance maps into an HDR radiance map for Position #1. In doing so, we can generate two HDR radiance maps based on the images warped using the forward warping and the backward warping algorithms. The temporal HDR radiance is produced from images of Exposures #2, #7 and #12 captured at Position #1. For visual comparison, we convert all HDR radiances into tone mapped HDR images. Figure 6.7 compares the tone mapped temporal HDR image and spatiotemporal HDR images created using the forward and the backward warping algorithms. The HDR image created with the backward warping algorithm is better than that with the forward warping. Some artifacts are removed in the HDR composition because the under- and over-exposed pixels are disregarded when combining radiance values from multiple images.

(a) (b)

Figure 6.5 Images warped to the middle view created with the forward warping algorithm from (a) Position #0 and (b) Position #2. Non-mapped pixels in black.

(a) (b)

Figure 6.6 Images warped to the middle view using the backward warping algorithm from (a) Position #0 and (b) Position #2.

Figure 6.7 Comparison of tone mapped temporal HDR image (left) and spatiotemporal HDR images obtained using the disparity-based forward (middle) and backward (right) warping algorithms. Whole (top row) and partial (bottom row) images for Position #1.

6.3.2 Impact of Exposure Increment Step Size

For stereo matching, increasing exposure time differences may reduce the quality of stereo correspondence. From a practical viewpoint, we would like to capture the smallest number of images possible with sufficient exposure differences to cover the dynamic range of scene radiance. We will investigate the impact of the exposure time increment step size with the dataset of the Ruth E. Dickinson public library scene as described in Section 6.1.1.

Small Step Size

We select a sequence of images each from one different capture position with different combinations of exposure times to generate an exposure sequence. For an exposure time sequence of short, medium and long with small step size (0.8 stops), we can select Exposures #5, #7 and #9 from Positions #0, #1 and #2, respectively. We can repeat this exposure sequence once to add Exposures #5, #7 and #9 from Positions #3, #4 and #5. Finally, we have six images with Exposures #5, #7, #9, #5, #7 and #9 for Positions #0 to #5. The RD-map images for this sequence are shown in Figure 6.8. We will use the disparity-based backward warping algorithm to warp the images from neighboring positions using the disparity maps found from the stereo matching using SVT.

To generate an HDR image for Position #1, we warp images at Positions #0 and #2 to Position #1. Then we create HDR images from the two warped images and the original image at Position #1. Similarly we can generate an HDR image for Position #2 using the images at Positions #1, #2 and #3, and for Position #3 using the images at Positions #2, #3 and #4.

Table 6.5 includes the radiance ranges recovered from single RD-map LDR images (Figure 6.8) and HDR radiance maps. As expected, the short exposure image captures higher band of the dynamic range while the long exposure image contains the low band of the dynamic range. The HDR images indeed recover the dynamic ranges contained in the three single exposed LDR images. As can be seen, the range in HDR radiance maps may not be identical to the dynamic range covered in the LDR images used to create the HDR radiance maps. This may be attributed to the HDR fusion.

(a) Position #0 (b) Position #1 (c) Position #2

(d) Position #3 (e) Position #4 (f) Position #5

Figure 6.8 RD-map images at six different positions with alternating exposures. From left to right: Exposures #5, #7 and #9. Exposure time increment step size is 0.8 stops.

Table 6.5 Radiance range contained in single images and recovered in spatiotemporal HDR radiance maps with sequence of small exposure time increment (increment of 0.8 stops). Temporal HDR radiance range is 73-26710 for Positions #1, #2 and #3.

Exposure #	Position #	Exposure time (ms)	Radiance range		HDR	
			Min - Max	Position #1	Position #2	Position #3
5	0	0.350	204-26710	73-26710		
7	1	0.584	122-16007		75-26222	
9	2	0.975	73-9588			75-26220
5	3	0.350	204-26710			
7	4	0.584	122-16007			
9	5	0.975	73-9588			

Figure 6.9 displays the temporal HDR images for each position using the LDR images from the same position which can serve as the basis for the assessment of the spatiotemporal HDR images as shown in Figure 6.10. The HDR images look fairly similar to the corresponding temporal HDR images. Close inspection reveals some artifacts present due to the inaccurate registration. The SSIM index measured is 0.86, 0.86 and 0.87 for the tone mapped spatiotemporal HDR images for Positions #1, #2 and #3 respectively. It seems that the proposed HDR framework produces HDR images consistently across all capture positions using the neighboring images.

(a) Position #1 (b) Position #2 (c) Position #3

Figure 6.9 Tone mapped temporal HDR images fused from multi-exposed images for each of three camera positions.

(a) Position #1 (b) Position #2 (c) Position #3

Figure 6.10 Tone mapped spatiotemporal HDR images fused from multi-view and multi-exposed LDR images for each of three camera positions.

Large Step Size

Similarly to the last section, we select six images each from a different position with the selected exposure sequence. We test a large step size of 1.6 stops by selecting Exposures #2, #7, #12, #2, #7 and #12 from Positions #0 to #5, respectively. The RD-map images for the sequence are shown in Figure 6.11.

An HDR image for Position #1 is generated from the images captured at Positions #0, #1 and #2. Similarly an HDR image for Position #2 is created from the images at Positions #1 #2 and #3. Table 6.6 shows the radiance ranges recovered from the single exposed images (Figure 6.11) and from the HDR images using the camera response curves recovered in Chapter 3 for the indoor scene. The HDR images indeed recover a higher dynamic range than each of the three single exposed LDR images. Compared to Table 6.5, the dynamic ranges of HDR images for Positions #1 and #2 in Table 6.6 are considerably higher than those for the same positions. As expected, the dynamic range increases as the exposure time range increases.

Table 6.6 Relative radiance range in single images and recovered in HDR images with series of a larger exposure increment (1.6 stops). Temporal HDR radiance range is 37-48187 for Positions #1 and #2.

Exposure #	Position #	Exposure time (*ms)*	Radiance range	Spatiotemporal HDR	
				Position #1	Position #2
2	0	0.194	367-48187	37-47885	
7	1	0.584	122-16007		38-42026
12	2	1.912	37-4889		
2	3	0.194	367-48187		
7	4	0.584	122-16007		
12	5	1.912	37-4889		

(a) Position #0, short (b) Position #1, medium (c) Position #2, long

(d) Position #3, short (e) Position #4, medium (f) Position #5, long

Figure 6.11 RD-map images at six different positions with alternating exposures of short, medium and long exposure times. Exposure increment of 1.6 stops. From left to right: Exposures #2, #7 and #12.

Figure 6.12 displays the temporal HDR images for Positions #1 and #2 created using three neighboring LDR images which can serve as the basis for the assessment of the spatiotemporal HDR images shown in Figure 6.13. Again the HDR images look fairly similar to the corresponding temporal HDR images. Slightly more artifacts are noticeable due to the reduced quality in disparity maps compared to the previous sequence with 0.8 stops. As expected, the SSIM index is 0.83 and 0.80 for the tone mapped HDR images for Positions #1 and #2, respectively. The SSIM indices of the HDR images for the increment step of 1.6 stops are lower than those of the HDR images for the increment of 0.8 stops for the same positions. It seems that the proposed omnidirectional spatiotemporal HDR image framework is able to handle a reasonably large exposure time increment step.

(a) Position #1 (b) Position #2

Figure 6.12 Tone mapped temporal HDR images in RD-map with exposure time increments of 1.6 stops. HDR images generated from three images captured at the same position.

(a) Position #1 (b) Position #2

Figure 6.13 Tone mapped spatiotemporal HDR images in RD-map with exposure time increments of 1.6 stops. HDR images created using three neighboring images.

6.3.3 Impact of Baseline

If the baseline between image pairs increases, the quality of stereo correspondence may be reduced due to possibly increasing half-occlusions. If the exposure differences between the adjacent images are too big, we may have to find disparity maps obtaining from stereo matching of image pairs with the same or close exposures but with wide baselines. Thus we like to evaluate the impact of baseline on the quality of HDR images. The image sequence is the same as in Section 6.3.2.1. The image at a position is matched with the image with the same exposure but further apart. For instance, the image at Position #1 is matched with the image at Position #4 to find disparity map for Position #1.

Figure 6.14 compares the temporal HDR and spatiotemporal HDR images for Position #1. As expected, more artifacts are shown in the window regions compared to the temporal HDR image on the left in Figure 6.14 and the fatting effect around the corner of the light tube on the top left part of the images. Using the temporal HDR image as reference, the

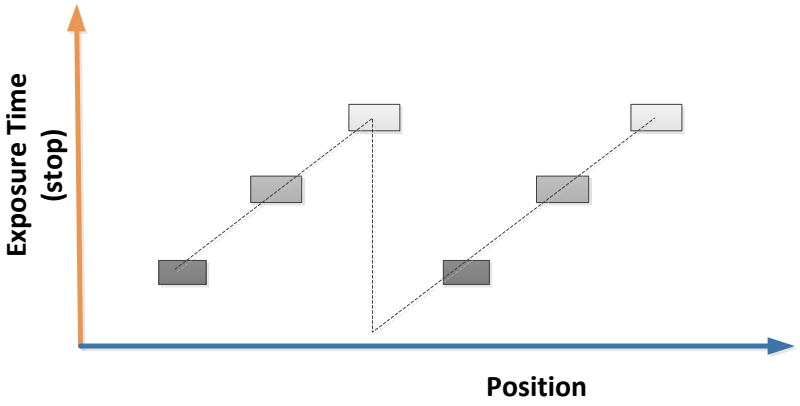
(a) Temporal HDR (b) Spatiotemporal HDR

Figure 6.14 Comparison between the temporal and spatiotemporal HDR images in RD-map with disparity maps evaluated from images with the same exposure but a large baseline.

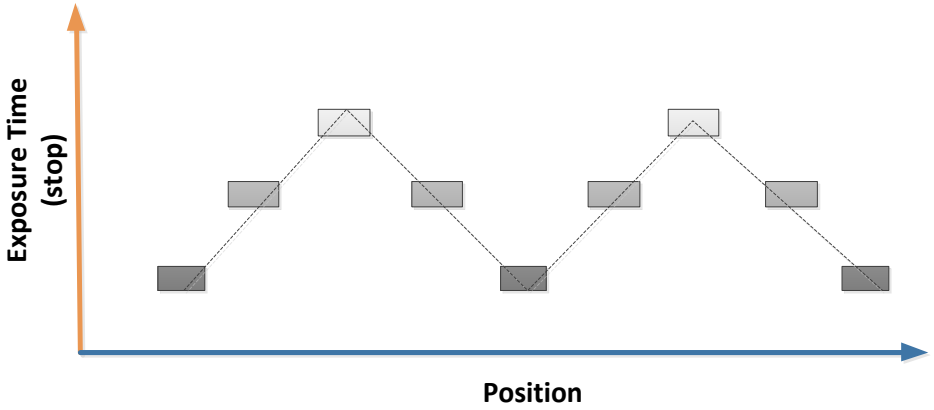
SSIM index of the spatiotemporal HDR image is 0.83. This is a slight reduction compared to the case in Section 6.3.2.1. An increasing baseline leads to a reduction in the quality of HDR images.

6.3.4 Impact of Exposure Sequence Ordering

When capturing an image sequence of two exposure times, we can alternate between short and long exposures. If an exposure sequence contains more than two exposure times, we can have different ordering of the exposure sequence. The simple way is to have the exposure times in an increasing order and repeat the sequence as shown in Figure 6.15.a. We may refer this type as saw tooth order. Another way is to change exposure times from the shortest to the longest and decrease from the longest down to the shortest in a triangle order as in Figure 6.15.b.



(a)



(b)

Figure 6.15 Two types of exposure sequence: (a) Saw tooth; (b) Triangle shape.

The triangle order has the same exposure time difference between adjacent images whereas the saw tooth shows a much bigger difference between the longest and the shortest exposure times. As the exposure difference increases, the camera geometric calibration may be affected. Thus, we will compare the impact of exposure sequence ordering on camera poses (translations and rotations) estimated form geometric calibration. We use the camera poses calibrated from the data of the same Exposure # for each viewpoint as the reference values and measure the errors as the differences between the reference values for

the similarly exposed sequence and the calculated values for the differently exposed sequence.

For a reference sequence of images captured, the rotational angle and the distance at each Position #*i* with respect to the first camera position are θ_i^r and d_i^r, respectively. If the rotational angle and the distance at Position #*i* for the current sequence are θ_i and d_i, the error in angle and the relative error in distance (%) are evaluated from the following expressions:

$$\delta\theta_i = (\theta_i - \theta_i^r) \tag{6.2}$$

and

$$P_i = \frac{d_i - d_i^r}{d_i^r} 100\% \tag{6.3}$$

Because the translations estimated from camera calibration are up to a scale, the CaliPano tool sets the translation between the first and second images in the sequence to 1. To be able to compare with the reference sequence, we include the same two images before the start of the different exposure sequences.

Table **6.7** and Table 6.8 show the results of the errors in rotational angles and relative translation for the front yard outdoor scene with the small and large exposure time increments. There are no significant errors in rotations at different capture positions for both the small and large exposure time increments. The errors in translations are also reasonably small (<5%) for two positions in the triangle ordering case. Although there could be accumulated errors in a sequence of images captured at different positions, we cannot see such a trend in the tables. Although the number of matching feature points in the saw tooth ordering is reduced, the calibration results are not affected as expected. The reason could that the estimation of the camera pose (translation and rotation) only requires a small number of matched feature points and the quality rather than the quantity of the matched feature points determines the accuracy of the geometric calibration.

Table 6.7 Impact of exposure sequence ordering on camera geometric calibration with small exposure time increments (0.8 stops) for the outdoor scene. The errors in rotation and translation are with respect to the sequence with the same exposure at the same position.

Ordering	**Position #**	**Exposure #**	**Error in rotation (degree)**	**Error in translation (%)**
Saw tooth	0	7	0.0	0
	1	7	0.0	0.0
	2	5	0.0	2.8
	3	7	0.0	1.8
	4	9	-0.2	2.9
	5	5	-0.2	2.4
	6	7	-0.1	2.3
	7	9	-0.3	4.3
Triangle	0	7	0.0	0
	1	7	0.0	0.0
	2	5	-0.0	2.8
	3	7	-0.0	1.8
	4	9	-0.2	2.9
	5	7	-0.2	4.3
	6	5	-0.1	6.0
	7	7	-0.3	9.8

Table 6.8 Impact of exposure sequence ordering on camera geometric calibration for the large exposure time increment sequence (1.6 stops) for the outdoor scene. The errors in rotation and translation are with respect to the sequence with the same exposure at the same position.

Ordering	**Position #**	**Exposure #**	**Error in rotation (degree)**	**Error in translation (%)**
Saw tooth	0	7	0.0	0
	1	7	0.0	0.0
	2	2	-0.1	1.7
	3	7	0.0	4.6
	4	12	0.0	3.3
	5	2	-0.3	2.2
	6	7	-0.1	1.7
	7	12	0.0	1.6
Triangle	0	7	0.0	0
	1	7	0.0	0.0
	2	2	-0.1	1.7
	3	7	0.0	4.6
	4	12	0.0	3.3
	5	7	-0.1	3.0
	6	2	-0.1	3.9
	7	7	-0.1	4.0

As can be seen in Table 6.9 for the indoor scene with the exposure sequence of saw tooth ordering, the errors in rotation and translation seem slightly greater for the large exposure time increments than for the small increments.

Table 6.9 Impact of exposure difference on camera geometric calibration for the indoor scene. The errors in rotation and translation are with respect to the sequence with the same exposure at the same position.

Exposure increment	**Position #**	**Exposure #**	**Error in rotation (degree)**	**Error in translation (%)**
Small (0.8 stops)	0	7	0.0	0
	1	7	0.0	0.0
	2	5	0.0	0.7
	3	7	0.1	1.1
	4	9	0.1	0.5
	5	5	0.2	1.1
	6	7	0.3	1.9
	7	9	0.3	2.5
	8	5	0.2	2.2
Large (1.6 stops)	0	7	0.0	0
	1	7	0.0	0.0
	2	2	0.0	0.8
	3	7	-0.1	2.3
	4	12	0.0	1.0
	5	2	0.1	2.0
	6	7	0.2	4.6
	7	12	0.2	6.0
	8	2	0.0	6.6

6.3.5 Outdoor Scene

The exposure sequence ordering of triangle type is taken from the front yard scene with Exposures #2, #7, #12, #7, #2, #7, #12 and #7 for the eight capture positions (Positions #0 to #7). The images at Positions #4, #5 and #6 are shown in Figure 6.16 and some tone mapping errors are noticeable in the image at Position #6 with the higher exposure. For each position, we can warp the other images to the current view and fuse the warped images and the current image to an HDR image in radiance space. The temporal HDR images are also created as references for the quality assessment of the spatiotemporal HDR images. Figure 6.17 displays the tone mapped temporal and spatiotemporal HDR images for Positions #4, #5 and #6. The spatiotemporal HDR images for different positions are not noticeably different when the quality of the current image changes. Compared to the temporal HDR images, the SSIM index of the spatiotemporal HDR images is 0.87, 0.91 and 0.89 for Positions #4, #5 and #6, respectively. It seems that the proposed framework can produces consistent quality HDR images for an outdoor scene as well in spite of some tone mapping errors present in the panoramic images.

(a) (b) (c)

Figure 6.16 RD-map images of the front yard scene captured at three positions with different exposures. (a) Position #4, Exposure #2; (b) Position #5, Exposure #7; (c) Position #6, Exposure #12.

(a) Temporal

(b) Spatiotemporal

Figure 6.17 Tone mapped temporal and spatiotemporal HDR images in RD-map generated using the proposed HDR framework. From left to right: Positions #4, #5 and #6.

6.4 PatchMatch Based HDR Methods

6.4.1 Hu et al.'s Method

Hu et al. [43] proposed the PatchMatch [66] based method for deghosting HDR images. The basic idea behind the method is to use a reference image as the template to create a latent image for each exposure. The latent image looks similar to the reference image in structure but is similar in brightness and color to the image of the given exposure. Once the latent images are generated, the exposure fusion method can be used to combine them into an HDR content image. In this approach, the ghosting effects caused by a moving camera or by moving objects are removed through freezing the moment when the reference image was captured.

One main issue with this method is that the quality of HDR images is strongly dependent on the quality of the reference image. If the reference image contains saturated regions, this method fails to extract the useful information from the short exposure images but produces grayish flat area instead for the saturated regions. Figure 6.18 shows the three RD-map images captured at different positions with varying exposure times used to evaluate the impact of quality of reference image on the HDR images. Figure 6.19 shows the latent images generated based on different reference images.

Figure 6.18 RD-map images from different views with increasing exposures used to test Hu et al.'s method [43].

(a) Short exposure image as reference

(b) Medium exposure image as reference

(c) Long exposure image as reference

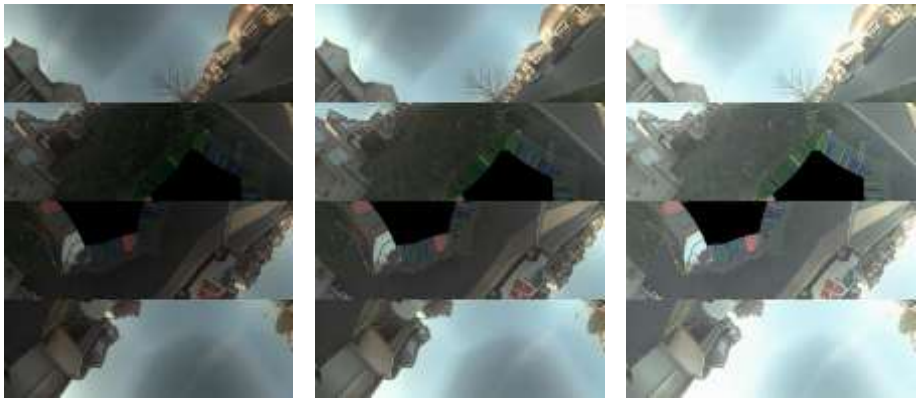
Figure 6.19 Impact of reference image on the latent images in RD-map for different positions generated using Hu et al.'s method [43].

As can be seen in Figure 6.19, the windows regions are filled with the grayish color in the bottom three latent images when using the long exposure image as the reference. Even the latent image with the short exposure time does not keep the color information from the features outside the windows. Figure 6.20 displays the final HDR images by applying the exposure fusion method [18] to the latent images. The HDR image using the long exposure image as the reference produces artifacts in the window areas and fails to capture the features outside the widows.

To compare the performance of Hu et al.'s method [43] with outdoor scene images, the same image sequence for the front yard outdoor scene in Section 6.3.5 is used. The latent images with different references are shown in Figure 6.21.

(a) (b) (c)

Figure 6.20 HDR content images in RD-map by applying the exposure fusion method [18] to the latent images created using Hu et al.'s method [43]. The impact on HDR image quality based on the reference image. (a) Short exposure; (b) Medium exposure and (c) Long exposure images.



(a) Exposure #2 as reference

(b) Exposure #7 as reference

(c) Exposure #12 as reference

Figure 6.21 Impact of reference image on the latent images in RD-map generated using Hu et al.'s method. From left to right: Positions #4, #5 and #6 in the outdoor scene dataset.

Figure 6.22 shows the comparison of tone mapped temporal and spatiotemporal HDR images generated using Hu et al.'s method [43] for the front yard scene. We can see the reddish artifacts in the spatiotemporal HDR image for Position #6 using the image with Exposure #12 as reference. Noticeable differences can be observed between the spatiotemporal HDR images by visual inspection and are also confirmed by the SSIM measurement. The SSIM index of the spatiotemporal HDR images is 0.88, 0.99 and 0.92 for Positions #4, #5 and #6, respectively.

(a) Temporal HDR

(b) Spatiotemporal HDR

Figure 6.22 Comparison of temporal and spatiotemporal HDR images in RD-map generated using Hu et al.'s method. From left to right: Positions #4, #5 and #6 in the outdoor scene dataset.

6.4.2 Sen et al.'s Method

Sen et al. [42] proposed a ghost-free HDR imaging method based on the PatchMatch [66] and the bidirectional similarity [67]. The HDR imaging method selects a reference image as the template to create an aligned (latent) image for each of the exposure sequence. An aligned image contains the brightness information from the current image of a given exposure but keeps the structure of the well-exposed regions in the reference image.

Figure 6.23 shows the impact of reference image. Similarly to Hu et al.'s method [43], the aligned images are strongly affected by the quality of the reference image. The use of the long exposure image as the reference results in a failure in capturing color information for the region, which is saturated in the reference image but is well-exposed in the short exposure image.

As an outdoor scene may have a great brightness contrast, the same image sequence for the front yard outdoor scene in Section 6.3.5 is used. The aligned images with different references are shown in Figure 6.24. Compared to the indoor scene result in Figure 6.23, the use of the long exposure image as reference does not propagates the artifacts for saturated regions to the aligned image for the low exposure view. Thus the impact of the reference image seems less visible.

(a) Short exposure image as reference

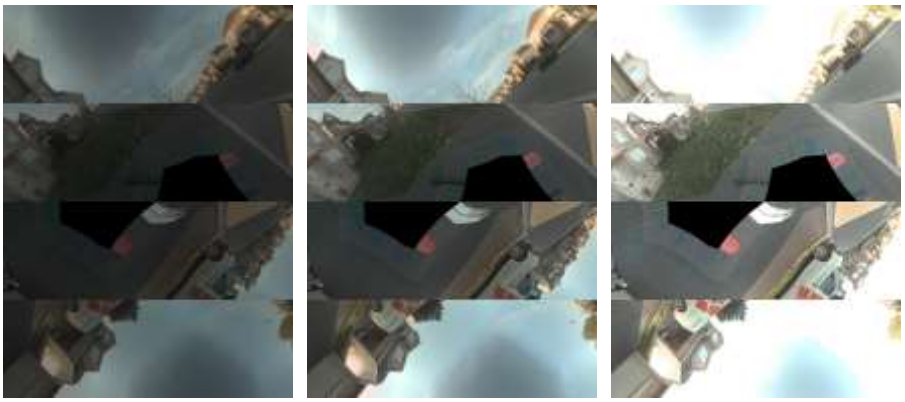
(b) Medium exposure image as reference

(c) Long exposure image as reference

Figure 6.23 Impact of reference image on aligned images in RD-map for different positions generated using Sen et al.'s Method [42].

(a) Short exposure image as reference

(b) Medium exposure image as reference



(c) Long exposure image as reference

Figure 6.24 Impact of reference image on aligned images in RD-map generated using Sen et al.'s method [42]. From left to right: Positions #4, #5 and #6 in the outdoor scene dataset.

Figure 6.25 shows the comparison of tone mapped temporal and the spatiotemporal HDR images generated using the exposure fusion method from the aligned images shown in Figure 6.24. There are big differences between the spatiotemporal HDR images with different reference images. Thus the quality of reference image has a great impact on the quality of HDR images. As can be seen, most noticeable artifacts are observed in the HDR image with the long exposure reference image. The SSIM index of the spatiotemporal HDR images is 0.79, 0.99 and 0.87 for Positions #4, #5 and #6, respectively.

(a) Temporal

(b) Spatiotemporal

Figure 6.25 Comparison of temporal and spatiotemporal HDR images in RD-map created using Sen et al.'s method. From left to right: Positions #4, #5 and #6 in the outdoor scene dataset.

6.5 Discussions

The goal of this thesis is to develop an omnidirectional framework for HDR imaging from multi-exposed LDR images captured with a moving camera. For each scene, the omnidirectional camera placed on a tripod was used to capture images of the same exposure sequence at each capture position along the selected capture path. The translations between two consecutive capture positions are 20~30 cm and rotational angles are in the range of 10 to 60 degrees. To form an image sequence of multiple exposures and multiple views, each image of a specific exposure time taken from the exposure sequence is selected from a different capture position. As a static scene was assumed and the maximum exposure time was less than 1/10 seconds, a sequence of differently exposed images selected from different capture positions in the dataset of a scene should be equivalent to the images of same exposure sequence captured with a moving camera.

The geometric calibration for the determination of camera poses is critical for stereo matching. In contrary to what is expected, the calibration is not sensitive to exposure sequence order and exposure time difference in the test. The possible reason may be that the determination of camera translation and rotation needs only a small number of correctly matched points (e.g. greater than or equal to 4). Although a large difference in exposure time may reduce the number of matching points, panoramic images with a wide field of view may still provide more matching points of good quality than what the geometric calibration needs.

Although the disparity-based forwarding warping algorithm produces warped images of considerably lower quality than does the backward warping algorithm, the HDR images produced by the former algorithm appear only slightly worse than those generated by the latter one. Most artifacts present in the warped images are removed in the HDR fusion and are not noticeable in the HDR images.

The proposed disparity-based spatiotemporal framework is able to recover the dynamic range of the scene radiance contained in the multi-view and multi-exposed images of a scene. It produces the radiance range of a scene from the multi-exposed images captured at

adjacent viewpoints similar to that produced from the multi-exposed images captured at the same viewpoint by the classic temporal HDR method. However, the tone mapped HDR images generated by the proposed framework have a lower quality than those by the classic method. As disparity maps can have errors due to half-occlusions and poorly textured or saturated regions, multi-view HDR imaging is expected to produce worse HDR images than is the classic temporal HDR imaging.

The HDR images generated for the outdoor scene look worse than those for the indoor scene as expected. The outdoor scene contains large poorly textured regions such as sky, road, grass which may also become over- or under-exposed regions in some images. Poorly textured regions and saturated regions in images can yield disparity maps of poor quality in stereo matching. Thus, additional prior information such as object boundaries or color models may be useful for disparity map refinement and HDR image improvement.

In contrast to the HDR content images generated with exposure fusion, HDR radiance maps are widely used in image-based rendering to produce a more realistic virtual environment [3]. Furthermore, HDR radiance maps can be used to recover the reflectance properties of a scene using the global inverse illumination method [4].

6.6 Summary

This chapter has presented an evaluation of the classic temporal HDR imaging method and the proposed spatiotemporal HDR imaging framework using the omnidirectional images captured for an indoor scene and an outdoor scene. We have compared the forward and backward warping algorithms and examined different factors such as exposure increment step size, baseline and exposure sequence ordering, outdoor and indoor scenes. The spatiotemporal HDR framework is able to recover the full dynamic range in radiance from multi-view and multi-exposed LDR images. The proposed framework can also recover the scene radiance ranges similar to the classic temporal HDR method but produces HDR images with more artifacts than the results the classic temporal HDR method. Compared to

the two state-of-the-art HDR imaging methods, our proposed HDR imaging framework seems able to produce more consistent HDR images across different camera positions. In addition, our framework captures disparity information which may be used to create HDR images for new positions other than the capture positions.

Chapter 7

Conclusions and Future Work

The goal of this thesis is the reconstruction of omnidirectional high dynamic range (HDR) images from multi-exposed low dynamic range (LDR) images captured with a moving camera. To this end, we have developed a spatiotemporal HDR imaging framework for generating omnidirectional HDR images based on disparity maps found from spherical stereo matching. In addition, we have created a novel experimental design to allow us to use the same datasets to create different image sequences for the evaluation of different HDR imaging methods. We have examined different factors influencing the proposed framework and demonstrated the success of reconstruction of HDR images both for indoor and outdoor scenes. Compared to two state-of-art HDR imaging methods, our HDR imaging framework can generate more consistent HDR images at all capture positions tested.

After summarizing the thesis in Section 7.1, we will outline the major contributions and draw the conclusions, and discuss the limitations of the proposed framework in Section 7.2. Finally, we will recommend some directions for future work in Section 7.3.

7.1 Summary

High dynamic range (HDR) imaging is a sub-field of computational photography and aims to overcome the limitations of digital cameras in capturing full dynamic ranges of real world scenes. We start the review of the related work in Chapter 2 with an introduction to computational photography and HDR imaging, and an outline of the basic methods and challenges of HDR imaging. The most common HDR imaging approach is to recover HDR

from multi-exposed LDR images. The challenges in HDR imaging are possible artifacts caused by object or camera movement during image capture. After an extensive review of the existing HDR image research, we have found that most of the HDR imaging methods deal with object movement and no or small camera movement. Although a few researchers have tried to develop HDR imaging techniques based on stereo matching, their work has been limited to planar images with small exposure differences and small camera movement. Thus, we are inspired to develop an omnidirectional multi-view HDR imaging framework so that we can expand both dynamic range and field of view with depth (3D) information.

Our proposed spatiotemporal HDR imaging framework is composed of three main stages: geometric calibration and alignment, multi-view stereo correspondence and HDR composition. Chapter 3 described the experimental design and method for image capture and photometric calibration, and Chapter 4 covered geometric calibration and alignment and spherical stereo matching. After Chapter 5 described the proposed disparity-based spatiotemporal HDR framework, Chapter 6 presented the evaluation of the proposed framework.

In Chapter 3 we first discussed the Ladybug 2 omnidirectional camera system and the experimental design for data capture. Then, we described the methods and the software tools for the generation of omnidirectional images and for the recovery of the inverse camera response curves that map pixel values to radiance and exposure time. Finally, we presented the results of the photometric calibration at different levels and for two camera settings used for the indoor and outdoor data captures.

Chapter 4 covered calibration and alignment, and stereo matching of omnidirectional images. The multi-scale stereo matching framework and the Spherical Vision Toolkit (SVT) developed by Brunton et al. [26] have been used to find stereo correspondence. After geometric calibration, panoramic images were aligned rotationally before feeding into SVT. Thus, we first described the methods and the software tool (CaliPano) for the geometric calibration. Then we explained the principles behind the multi-scale stereo

matching framework and the major stages in the SVT pipeline. Finally, some results of applying SVT to images with different exposures were presented.

Chapter 5 described the respective pipeline for the temporal and the spatiotemporal omnidirectional HDR imaging methods. We first described the major stages in the temporal HDR imaging pipeline: radiance conversion based on inverse camera response curves, HDR composition in radiance space, and tone mapping of HDR radiance maps to displayable images. We then described our spatiotemporal HDR pipeline. For spatiotemporal HDR imaging, images from multiple views must be registered before HDR fusion. We proposed to use disparity maps to warp images from neighboring views to any selected target view, and then use the current and warped images to generate HDR images. We proposed two warping algorithms: forward and backward. We also discussed some possible methods for improving disparity maps. Large sky regions present in outdoor scenes would result in disparity maps of poor quality. We suggested using a superpixel technique and a color model for sky as prior information to fill in holes in the sky regions with correct disparity values.

In Chapter 6 we presented an evaluation of the classic temporal HDR imaging method and the proposed disparity-based spatiotemporal HDR imaging framework using the omnidirectional images captured for an indoor scene and an outdoor scene. The forward and backward warping algorithms were compared and the backward warping algorithm seems to produce better results. We also examined different influencing factors such as exposure increment step size, baseline and exposure sequence ordering. The proposed HDR framework is able to recover the full dynamic range of radiance contained in multi-view and multi-exposed LDR images. Compared to the temporal HDR method, the proposed framework produces very similar dynamic ranges in scene radiance. The quality of tone mapped HDR images generated by the proposed framework is generally lower than that by the temporal HDR method. The reason may be that over- or under-exposed areas and poorly textured regions lead to no or invalid values in disparity maps. Furthermore, our

proposed HDR imaging framework produces more consistent HDR images across different viewpoints than do two state-of-the-art HDR imaging methods.

7.2 Conclusions and Limitations

In this thesis, an omnidirectional spatiotemporal HDR imaging framework has been developed. It has been evaluated using the panoramic datasets captured in an indoor scene and an outdoor scene, and compared with two state-of-the-art HDR methods. The proposed framework allows us to capture HDR radiance maps, disparity maps and full field of view, which can have many applications such as HDR view synthesis and virtual navigation.

Our major contributions may be summarized as follows:

- A disparity-based spatiotemporal HDR imaging framework that can generate omnidirectional HDR images from multi-exposed LDR images captured with a moving camera.
- Disparity-based forward and backward image warping algorithms for spherical stereo vision.
- Evaluation of the temporal and the spatiotemporal HDR imaging methods and examination of the different factors such as warping method, baseline and exposure sequence ordering, indoor and outdoor scenes.
- Use of a superpixel method and a color model as prior information for disparity map refinement for the outdoor scenes with large sky regions.

We may draw the following conclusions:

- The proposed disparity-based omnidirectional spatiotemporal HDR framework is able to recover the full range of radiance contained in multi-view and multi-exposed LDR images captured with a moving camera both for indoor and outdoor scenes.

- Compared to the classic temporal HDR imaging method, the proposed framework is more efficient to produce HDR images, and is still able to recover similar scene radiance ranges. As expected, however, the proposed framework generates HDR images with more artifacts than does the classic method as expected.
- The framework yields more consistent HDR images across different viewpoints for both indoor and outdoor scenarios than two state-of-the-art HDR imaging methods.

There are some limitations in the work presented in the thesis. First, we focus on omnidirectional HDR imaging with a moving camera, and develop the techniques based on stereo correspondence for registering images captured from different viewpoints. Thus, we do not include algorithms to deal with moving objects and we may see artifacts if moving objects are present. However, the removal of moving objects is possible in the framework if an error mask for moving regions can be provided. Second, the backward warping algorithm produces smooth warped images whereas the forward warping algorithm generates warped images with more invalid regions. Although the warping algorithms has a much less impact on the quality of fused HDR images than on that of warped images, the forward warping algorithm still produces some noticeable artifacts on the HDR images. Thus, image inpainting methods or combined disparity map and image inpainting methods can be used to improve the quality of HDR images if the forward warping algorithm is used.

There are also some limitations in the software tool used in the proposed framework: the SVT (Spherical Vision Toolkit) and CaliPano. Although SVT can cope with some variation in brightness, it is not designed for stereo matching of images with large exposure differences. Therefore, the under- or over-exposed regions in images are not considered explicitly in the stereo matching processing. Furthermore, SVT produces less accurate disparity maps for outdoor scenes due to large poorly textured regions such as sky, roads. As for the calibration tool, CaliPano uses OpenSURF as feature extractors on cylindrical images. OpenSURF should be able to cope with exposure differences because it extracts

features in the gradient field. However, OpenSURF is designed for planar images and some new feature extractor specifically designed for spherical images [104] may perform better.

Furthermore, the Ladybug 2 camera system has some limitations. First, it cannot capture very high radiance in most outdoor scenes as it has a fixed aperture and one can adjust only shutter speed and gain. In other words, even we select the minimum shutter speed and the gain of 0 dB, we still get over-exposed regions. Second, CCD smear and veiling glare can occur under direct bright sunlight. Thus, a scene in direct sunlight cannot be captured.

7.3 Future work

There are a number of interesting directions for future work that can improve the proposed multi-view omnidirectional HDR imaging framework:

- **Optical flow or block matching for detecting moving objects:** Our current framework deals only with camera movement and does not consider object movement. One extension to the current work is to include optical flow [28] or block matching [25] for the detection of moving objects or for refining the registration based on stereo correspondence. As in the work of Kang et al. [34], optical flow is used for local registration to complement the homography based global registration.
- **Improvement of stereo matching of differently exposed images in SVT**: The current radiometrically adjusted and gradient based approach may not be sufficient when exposure differences are too big. The census transform and the rank filter [78] may be helpful. In addition, the information on the under- or over-exposed regions and exposure ordering may be taken into account.
- **Improvement of camera calibration in CaliPano**: Further improvement of camera calibration may be critical for the generation of HDR images with long exposure sequences. In the CaliPano software tool OpenSURF is used. OpenSURF

is fast but is less effective than SIFT for detecting feature points. Thus, one modification is to implement SIFT [64] in the CaliPano tool. Another interesting work is to explore the recently proposed spherical SIFT [104].

- **Disparity map and color image enhancement:** image inpainting techniques [80] can be explored to fill in the holes in disparity map and color images simultaneously. Epipolar geometry may be used to guide and speed up the inpainting processing.

Appendix A: Ladybug 2 Camera System and Software

The Ladybug 2 camera system (Figure A.1) used for the raw data capture is an omnidirectional camera system developed by Point Grey [38]. It consists of one upward camera, and five sideward cameras and covers about 80% of the full view.

Figure A.1 Photograph of the Ladybug 2 camera system used for the image capture.

Each camera has a wide angle lens with a fixed focal length, a fixed aperture and a Sony ICX204AK 1/3" sensor. The main technical features of the Ladybug 2 camera system [38] are listed as:

- **Image sensor** Six Sony ICX204AQ 1/3" 1024x768 progressive scan CCDs

• **A/D Converter**	Six analog devices 12-bit analog-to-digital converters
• **Data Output (maximum)**	15 FPS uncompressed 8bpp Bayer-tiled data
	30 FPS JPEG compressed 8bpp Bayer-tiled data
• **Interfaces**	Head Unit to Compressor: 1.2Gbps optical link
	Compressor to PC: 800 Mbps IEEE-1394b link
• **Frame Rates**	30, 15, 7.5, 3.75 FPS
• **Optics**	High quality micro lenses of focal length of 2.5 mm
	Cameras in the horizontal ring are in portrait orientation
• **Shutter**	Automatic/Manual Shutter modes
• **Gain**	Automatic/Manual Gain modes,
	0~26 dB

Unlike common commercial cameras, all configuration and capture operations for the Ladybug 2 camera system have to be controlled using the software tools or library functions in the SDK (Software Development Kit). The Point Grey releases the Ladybug SDK which is a software package designed specifically for the use with Ladybug cameras. The package includes device driver, SDK, a number of example projects of common applications, and the LadybugCapPro software tool. LadybugCapPro allows us to configure the camera system and perform image and video capture and saving functions.

The SDK provides a complete programming library for image capture, processing, saving, and display, and supports standard C/C++/C# interfaces. All library functions are included in the Ladybug API (Application Programming Interface) header files with the following functions:

- Configure various camera parameters such as shutter, gain and white balance.
- Control the recording of images.
- Color process images using a variety of different color processing algorithms.

- Stream images off the camera to disk, access and download images in stream files to the host computer.
- Display fully stitched panoramic and spherical images in real time or from a saved file.

Ladybug_vidcapture is an in-house tool that uses the Ladybug SDK to configure, capture and save image data. It provides some basic GUI (Graphic User Interface) for selecting capture options, displaying individual images, and outputting camera settings and results. A user can use the GUI to select what images to capture and what format to save, and perform image capture in the sequence specified in the configuration file. Ladybug_vidcapture was modified to accommodate the capture procedures and bracketed exposures required for the experimental data capture in Section 3.1.

Bibliography

[1] R. Raskar and J. Tumblin, *Computational Photography: Mastering New Techniques for Lenses, Lighting and Sensors*. A.K. Peters Press, 2010.

[2] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, and G. Ward, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann, 2010.

[3] P.E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH*, 1997, pp. 369–378.

[4] Y. Yu, P. Debevec, J. Malik, and T. Hawkins, "Inverse global illumination: recovering reflectance models of real scenes from photographs," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 215-224. ACM Press/Addison-Wesley Publishing Co., 1999.

[5] B. Hoefflinger, Ed., *High Dynamic Range (HDR) Vision: Microelectronics, Image Processing*, Computer Graphics. Springer, 2007.

[6] S. Hrabar, P. Corke, and M. Bosse, "High dynamic range stereo vision for outdoor mobile robotics," in *Proceedings of IEEE International Conference on Robotics and Automation*, Kobe, Japan, 2009, pp. 430-435.

[7] R. Lukac, Ed., Computational Photography: Methods and Application. CRC Press 2011.

[8] J.A. Ferwerda, S.N. Pattanaik, P. Shirley, and D.P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, 1996, pp. 249–258.

[9] Photofocus, *1.4 Megapixel Photonfocus monochrome CMOS image sensor*. [Online]. Available: http://www.photonfocus.com/upload/flyers/flyer_A5_A1312_en_1.1.6.pdf [Accessed: 3 March, 2014].

[10] The J. Paul Getty Museum, *Gustave Le Gray, photographer*, The J. Paul Getty Museum. [Online]. Available: http://www.getty.edu/art/exhibitions/le_gray/ [Accessed: 3 March, 2014].

[11] Image Source, *Canon T4i Review*, Image Source. [Online]. Available: http://www.imaging-resource.com/PRODS/canon-t4i/canon-t4iA.HTM [Accessed: 3 March, 2014].

[12] SpheronVR AG, *Image Based Lighting with Spheron HDR*, SpheronVR AG. [Online]. Available: http://www.grafixgear.com/HTML/PDF/SpheroCam_HDR.pdf [Accessed: 3 March, 2014].

[13] A. Srikantha and D. Sidibe, "Ghost detection and removal for high dynamic range images: recent advances," in *Signal Processing Image Communication*, EURASIP, 2012.

[14] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, "Google street view: Capturing the world at street level," Computer, vol. 43, no. 6, pp. 32–38, 2010.

[15] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. Jour. Computer Vision*, 47(1/2/3):7–42, 2002.

[16] Middlebury, *Middlebury Stereo Evaluation - Version 2*, Middlebury. [Online]. Available: http://vision.middlebury.edu/stereo/eval/ [Accessed: 3 March, 2014].

[17] A.A. Goshtasby, "Fusion of Multi-Exposure Images," *Image and Vision Computing*, 23, 2005, pp. 611–618.

[18] T. Mertens, J. Kautz, and F.V. Reeth, "Exposure fusion: a simple and practical alternative to high dynamic range photography," *Computer Graphics Forum*, 28, 2009, pp. 161–171.

[19] M. Piccardi, "Background subtraction techniques: a review," in *Systems, Man and Cybernetics*, 2004 IEEE International Conference on, vol. 4, Oct. 2004, pp. 3099–3104.

[20] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV Library*, 1st ed., O'Reilly Media, 2008.

[21] O. Gallo, N. Gelfand, W.-C. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *Proceedings of the IEEE International Conference on Computational Photography* (ICCP), 2009, pp. 1–7.

[22] E.A. Khan, A.O. Akyuz, and E. Reinhard, "Ghost removal in high dynamic range images," in *Proceedings of the IEEE International Conference on Image Processing*, 2006, pp. 2005–2008.

[23] F. Pece and J. Kautz, "Bitmap movement detection: HDR for dynamic scenes," in *Proceedings of Visual Media Production* (CVMP), 2010, pp. 1–8.

[24] W. Zhang and W.-K. Cham, "Gradient-directed composition of multi-exposure images," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2010, pp. 530–536.

[25] S. Mangiat and J. Gibson, "High dynamic range video with ghost removal," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7798, August 2010.

[26] S. Silk and J. Lang, "Fast high dynamic range image deghosting for arbitrary scene motion," in *Proceedings of Graphics Interface*, 2012, pp. 85-92.

[27] S. Silk, "High dynamic range panoramic imaging with scene motion," *M.A.Sc. thesis*, EECS, University of Ottawa, 2011.

[28] H. Zimmer, A. Bruhn, and J. Weickert, "Freehand HDR imaging of moving scenes with simultaneous resolution enhancement," *Computer Graphics Forum*, 2011.

[29] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from handheld exposures*," Journal of Graphics Tools*, 8(2), 2004, pp.17–30.

[30] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," in *Vision, Modeling and Visualization*, RWTH Aachen, 2006, pp. 277–284.

[31] K. Jacobs, C. Loscos, and G. Ward, "Automatic high-dynamic range image generation for dynamic scenes," *IEEE Computer Graphics and Applications*, 28:84–93, 2008

[32] A. Tomaszewska and R. Mantiuk, "Image registration for multi-exposure high dynamic range image acquisition," *WSCG*, January 2007.

[33] A. Troccoli, S.B. KANG, and S.M. Seitz, "Multi-view multi-exposure stereo," in *3DPVT* 2006, pp. 861–868.

[34] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, 22:319–325, July 2003.

[35] N. Sun, H. Mansour, and R.K. Ward, "HDR image construction from multi-exposed stereo LDR images," in *ICIP*, 2010, pp. 2973–2976.

[36] D. Rufenacht, "Stereoscopic high dynamic range video," *Master's thesis*, EPFL, Lausanne, Switzerland, 2011.

[37] T. Akhavan, H. Yoo, and M. Gelautz, "A framework for HDR stereo matching using multi-exposed images," in *HDRi2013 - First International Conference and SME Workshop on HDR imaging*, 2013.

[38] Point Grey Research Inc., "Ladybug2 Getting Started Manual," Online: http://www.ptgrey.com/support/downloads/documents/Ladybug2%20Getting%20Started%20Manual.pdf. [Accessed: 3 March, 2014].

[39] C.-W. Fu, L. Wan, T.-T. Wong, and C.-S. Leung, "The rhombic dodecahedron map: An efficient scheme for encoding panoramic video," *IEEE Transactions on Multimedia*, vol. 11, no. 4, June 2009.

[40] A. Brunton, "Multi-scale methods for omnidirectional stereo with application to real-time virtual walkthroughs," *PhD's thesis*, EECS, University of Ottawa, 2012.

[41] A. Brunton, J. Lang, and E. Dubois, "Efficient multi-scale stereo of high-resolution planar and spherical images," *3DIMPVT 2012*: 120-127, 2012.

[42] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," in *SIGGRAPH Asia*, 2012.

[43] J. Hu, O. Gallo, K. Pulli, and X. Sun, "HDR deghosting: how to deal with Saturation?" in *Proceedings of CVPR*, Portland, OR, June 23-28, 2013.

[44] R. Raskar, "Computational photography: epsilon to coded photography," in *ETVC 2008*, LNCS 5416, 2009, pp. 238-253.

[45] N. Ren, "Fourier Slice Photography," in *ACM SIGGRAPH*, 2005.

[46] M. Brown and D. Lowe, "Recognizing panoramas," In *Ninth International Conference on Computer Vision*, Nice, France, October 2003, pp. 1218-1225.

[47] S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts, and J. Bogaerts, "A logarithmic response CMOS image sensor with on-chip calibration", *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 8, August 2000.

[48] A Martinez-Sanchez, C Fernandez, P J. Navarro, and A. Iborra, "A novel method to increase LinLog CMOS sensors' Performance in High Dynamic Range Scenarios," *Sensors* 2011, 11, pp. 8412-8429.

[49] E. Ikeda, "Image data processing apparatus for processing combined image signals in order to extend dynamic range," *U.S. Patent 5801773*, September 1998.

[50] M. Aggarwal and N. Ahuja, "Split aperture imaging for high dynamic range," in *IEEE International Conference on Computer Vision* (ICCV), 2, 10–17 July 2001.

[51] M.D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," *ACM Trans. on Graph*. 30, 4, July 2011.

[52] D.D. Wen, "High dynamic range charge-coupled device," *US Patent 4873561*, 1989.

[53] R.A. Street, "High dynamic range segmented pixel sensor array," *U.S. Patent 5789737*, August 1998.

[54] S.K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. Of IEEE Conf. on Computer Vision and Pattern Recognition 2000*, 1:472–479, June 2000.

[55] S.K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: optical control of pixel exposures over space and time," *IEEE International Conference on Computer Vision (ICCV)*, Vol.2, pp.1168-1175, Oct, 2003.

[56] T. Mitsunaga and S.K. Nayar, "Radiometric self calibration," *in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, Los Alamitos, CA, USA: IEEE Computer Society, 1999.

[57] S. Mann and R.W. Picard, "On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures," in *Proceedings of IS&T*, 1995, pp. 442–448.

[58] M.A. Robertson, S. Borman, and R.L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proceedings of 1999 International Conference on Image Processing*, vol. 3. IEEE, 1999, pp. 159–163.

[59] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proc. of SIGGRAPH '02*, ACM Press, 2002.

[60] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of ACM SIGGRAPH '02*, ACM Press, 2002, pp. 249-256.

[61] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proceedings of ACM SIGGRAPH '02*, ACM Press, 2002, pp. 257-266.

[62] R.I. Hartley and A. Zisserman, *Multiple View Geometry*. Cambridge University Press, Cambridge, UK, 2004.

[63] R. Szeliski, *Computer Vision: Algorithms and Applications*. Springer London, 2010.

[64] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[65] M.A. Fischler and R.C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, 24(6), 1981, pp. 381-395.

[66] C. Barnes, E. Shechtman, A. Finkelstein, And D.B. Goldman, "Patchmatch: a randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.* 28, 2009, pp. 24:1–24:11.

[67] D. Simakov, Y. Caspi, E. Shechtman, And M. Irani, "Summarizing visual data using bidirectional similarity," in *CVPR 2008*, pp. 1–8.

[68] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scenes with hand-held cameras," in *ECCV*, 2012.

[69] K. Karaduzovic-Hadziabdic, J.H. Telalovic, and R. Mantiuk, "Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging," *HDRi2014 - Second International Conference and SME Workshop on HDR imaging*, 2014.

[70] D. Sidibe, W. Puech, and O. Strauss, "Ghost detection and removal in high dynamic Range images," in *Proceedings of the 17th European Signal Processing Conference - EUSIPCO*, 2009, pp. 2240–2244.

[71] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of Gaussians for foreground detection - a survey," *Recent Patents on Computer Science*, Volume 1, No 3, pp. 219-237, November 2008.

[72] M. Granados, H.-P. Seidel, and H.P. Lensch, "Background estimation from non-time sequence images," in *Proc. GI*, 2008, pp. 33–40.

[73] I.E. Richardson, *H.264 and MPEG-4 video compression*, Wiley, 2003.

[74] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI*, 1981.

[75] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2006, pp. 519–526.

[76] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[77] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328–341, 2008.

[78] H. Hirschmüller and D. Scharstein. "Evaluation of stereo matching costs on images with radiometric differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1582-1599, September 2009.

[79] H.-Y. Lin and W.-Z. Chang, "High dynamic range imaging for stereoscopic scene representation," in *ICIP*, 2009, 4249–4252.

[80] L. Wang, H. Jin, R. Yang, and M. Gong, "Stereoscopic inpainting: joint color and depth completion from stereo images," In *CVPR '08: Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[81] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles, "Image inpainting," In *SIGGRAPH '00*, ACM, 2000, pp. 417–424.

[82] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *CVPR '03: Conference on Computer Vision and Pattern Recognition*, 2003, pp. 721–728.

[83] S. Ikehata, J.-H. Cho, and K. Aizawa, "Depth map inpainting and super-resolution based on internal statistics of geometry and appearance," in *Proceedings of IEEE International Conference on Image Procession (ICIP)*, 2013.

[84] J. van de Weijer and T. Gevers, "Robust optical flow from photometric invariants," *ICIP*, 2004

[85] Y. HaCohen, E. Shechtman, D.B. Goldman, and D. Lischinski, "NRDC: non-rigid dense correspondence with applications for image enhancement," *SIGGRAPH*, 2011.

[86] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," *ECCV*, 2006.

[87] Y. Yasushi, "Omnidirectional sensing and its applications," *IEICE Transactions on Information and Systems*, vol. E82-D, no. 3, 1999, pp.568-579.

[88] D. Scaramuzza, Omnidirectional Camera, GRASP Lab, University of Pennsylvania. [Online]. Available: http://rpg.ifi.uzh.ch/docs/omnidirectional_camera.pdf. [Accessed: 18 March, 2014].

[89] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec, "Direct HDR capture of the Sun and sky", *Afrigraph* 2004, Cape Town, South Africa, November 2004.

[90] F. Okura, M. Kanbara, and N. Yokoya, "Full spherical high dynamic range imaging from the sky," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2012, pp. 325 – 332.

[91] H. Ishigura and S. Tsuji, "Omni-directional stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, 1992.

[92] S. Kang and R. Szeliski, "3D scene data recovery using omnidirectional multi-baseline stereo," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1996.

[93] M. Aly and J.-Y. Bouget, "Street view goes indoors: Automatic pose estimation from uncalibrated unordered spherical panoramas," in *IEEE Workshop on the Applications of Computer Vision*, 2012.

[94] XRite Incorporated., "ColorData-1p_EN," XRite Incorporated. Online: http://xritephoto.com/documents/literature/en/ColorData-1p_EN.pdf. [Accessed: 3 March, 2014].

[95] M. Eitz and C. Stripf, "High dynamic range imaging and tone mapping," 2007. Online: http://cybertron.cg.tu-berlin.de/eitz/hdr/. [Accessed: 3 March, 2014].

[96] P.E. Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography," In *SIGGRAPH 98*, July 1998.

[97] J. Unger and S. Gustavson, "High dynamic range video for photometric measurement of illumination," in *Proceedings of Sensors, Cameras and Systems for Scientific/Industrial*

Applications X, IS&T/SPIE 19th Inernational Symposium on Electronic Imaging, vol. 6501, 2007

[98] J. Unger, J. Kronander, P. Larsson, S. Gustavson, and A. Ynnerman, "Temporally and spatially varying image based lighting using hdr-video," in *Proceedings of EUSIPCO '13, Marrakech*, Morocco, 9-13 September, 2013.

[99] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J.-M. Frahm, R. Yang, D. Nister, and M. Pollefeys, "Real-time visibility-based fusion of depth maps," in *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[100] Z. Arican and P. Frossard, "Dense disparity estimation from omnidirectional images," in *2007 IEEE AVSS*, 2007.

[101] J. Heller, "Stereo reconstruction from wide-angle images," *Master Thesis*, 2008, Department of Software and Computer Science Education, Charles University in Prague, Czech.

[102] A. Pagani and D. Stricker, "Structure from motion using full spherical panoramic cameras," *ICCV Workshops* 2011: 375-382.

[103] J. Fujiki, A. Torii, and S. Akaho, "Epipolar geometry via rectification of spherical images," in *MIRAGE*, 2007.

[104] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire, and J.-P. Thiran, "Scale invariant feature transform on the sphere: Theory and applications," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 217–241, 2012.

[105] J. Zhu, G. Humphreys, D. Koller, S. Steuart, and R. Wang, "Fast omnidirectional 3D scene acquisition with an array of stereo cameras," *IEEE Computer Society 3DIM*, 2007, pp. 217-224.

[106] N. Menzel and M. Guthe, "Freehand HDR photography with motion compensation," in *Proceedings of Vision, Modeling, and Visualization (VMV)*, AKA Heidelberg, 2007, pp. 127–134.

[107] P. Sand and S. Teller, "Video matching," in *Proc. ACM SIGGRAPH*, 2004, pp. 592–599.

[108] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," In *CVPR*, pages 261–268, 2004.

[109] M. A. Lourakis and A. Argyros, "SBA: a software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, no. 1, pp. 1–30, 2009.

[110] C. Evans, Notes on OpenSURF Library, 2009. Online: http://www.chrisevansdev.com/computer-vision-opensurf.html. [Accessed: 3 March, 2014].

[111] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR Upper Saddle River, NJ, USA, 1998.

[112] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[113] A.C. Brooks, X. Zhao, and T.N. Pappas, "Structural similarity quality metrics in a coding context: exploring the space of realistic distortions," *IEEE Transactions on Image Processing*, VOL. 17, NO. 8, AUG. 2008.

[114] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. ACM Trans. Graph., 30(4), 2011.

[115] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels compared to state-of-the-art superpixel methods, " IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, 2012, pp. 2274-2282.

[116] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation," ACM Transactions on Graphics, Vol. 32, No. 3, 2013.

[117] F. Schmitt and L. Priese, "Sky Detection in CSC-segmented Color Images," *VISAPP 2*, INSTICC Press, 2009, pp. 101-106.

[118] M. Levoy and P. Hanrahan, “Light field rendering,” In Proc. SIGGRAPH'96, Los Angeles, CA, 1995, pp. 29-38.

[119] G. Schaufler and M. Priglinger, “Efficient displacement mapping by image warping,” in *Proceedings of the Eurographics Workshop in Granada*, Spain, 1999, pp. 175-186.

[120] NVIDIA, CUDA Programming Guide, 2010.