

ADAPTIVE EXPOSURE FUSION FOR HDR IMAGING

by

Sidhdharthkumar Patel

Bachelor of Engineering, Ryerson, 2013

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the Program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2015

©Sidhdharthkumar Patel 2015

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

Adaptive Exposure Fusion for HDR Imaging

Master of Applied Science 2015

Sidhdharthkumar Patel

Electrical and Computer Engineering

Ryerson University

Abstract

HDR images are usually generated by fusing sequence of images captured with variable exposure time. Exposure Fusion is a technique that directly fuses the exposure shots into displayable image. This thesis proposes a novel Exposure Fusion algorithm that directly fuses exposure bracketed shots into a displayable image. Most techniques targeted for direct fusion do not have an effective exposure control mechanism and are only designed for exposure sequence containing adequate number of exposure shots. The proposed algorithm offers a novel approach that adaptively adjusts its parameter for the best viewing experience even for exposure sequence that do not contain adequate number exposure shots. An online user survey showed that the proposed method gave consistent superior results when compare with other methods of similar nature. The survey results were also compared against state of the art evaluation metrics (TMQI and HDR-VDP) and the survey results contradicted the evaluation metric.

Acknowledgements

First of all I would like to thank my supervisors Dr. Kyan and Dr. Androutsos for their amazing guidance in this project. I like to thank Ahmed Sagarwala and Dr. Dixit for their help with the the user survey. I would also like to thank Laura Thomas for her invaluable help in editing the thesis. Lastly, I would like thank my sister for her motivation and encouragement.

Contents

<i>Declaration</i>	ii
<i>Abstract</i>	iii
<i>Acknowledgements</i>	iv
<i>List of Tables</i>	viii
<i>List of Figures</i>	ix
<i>List of Appendices</i>	xi
1 High Dynamic Range Imaging	1
1.1 Introduction	1
1.2 Applications	2
1.3 The Pipeline	3
1.4 Organization	5
2 HDR Image Generation	8
2.1 Introduction	8
2.2 Hardware Based Methods	9
2.3 Software Based Methods	10
2.3.1 Median Threshold Bitmap	10
2.3.2 Image Fusion	11
2.3.3 Ghost removal	13
2.4 Summary	14

3	Tone Mapping	15
3.1	Introduction	15
3.2	Stevens' Power Law	16
3.3	Naka-Rushton Model	17
3.4	Retinex	18
3.5	Global Tone Mapping Operators	19
3.6	Local Tone mapping operators	21
3.7	Discussion	22
3.8	Summary	24
4	Exposure Fusion	25
4.1	Introduction	25
4.2	Existing Methods	26
4.3	Adaptive Exposure Fusion	28
4.3.1	Motivation	28
4.3.2	Proposed Method	29
4.3.3	Exposure Adjustment	30
4.3.4	Adaptive k	33
4.3.5	Contrast, Saturation and Well Exposedness	36
4.4	Summary	39
5	Experimental Setup and Results	41
5.1	Introduction	41
5.2	The Dataset	41
5.3	HDR Image Evaluation Site	42
5.4	Mean Reciprocal Rank (MRR)	42
5.5	Results	43
5.6	Comparison with other HDR image evaluation metric	44
5.7	Discussion	45
5.8	Summary	61

6 Conclusion and Future Work	62
6.1 Future Work	63
Appendices	64
References	72

List of Tables

5.1	Score for image set 15	50
5.2	Score for image set 11	50
5.3	Score image set 23	54
5.4	score fpr image set 35	54
5.5	Score for image set 3	59
5.6	Score for image set 4	59
5.7	Score for fig 30	59
5.8	Score for image set 32	59

List of Figures

1.1	HDR Image Acquisition Pipeline	4
1.2	Increasing exposure time from (a) to (h)	6
1.3	Increasing exposure time from (a) to (g)	7
3.1	Naka-Rushton Model	17
3.2	Tone mapped image generated using various TMOs	23
4.1	Mertens' Exposure Fusion algorithm	27
4.2	a) Tone mapped image; Exposure Fusion Image	30
4.3	proposed Exposure Fusion algorithm	32
4.4	Result of equation 4.1 after setting various k values	33
4.5	Graphical representation of equation 4.2	35
4.6	(a)Comparison of original exposure image before and after Exposure Adjustment a) Original exposure image;(b) After application of Exposure Adjustment stage	35
4.7	Adaptive k value for a exposure sequence	36
4.8	(a)Image generated using all three metric;(b) Resulting image with the absence of Contrast metric	37
4.9	(a)Image generated with all three quality metrics; (b) Resulting image with the absence Saturation metric	37
4.10	(a)Image generated with all three quality metrics;(b) Resulting image with the absence Well Exposedness metric	38
4.11	visualization of the quality metric: (a) Contrast; b)Saturation; c) Well Exposedness d) combined quality metric	39

4.12	Resulting pyramids decomposition of one exposure shot	40
5.1	evaluation site	43
5.2	Mean Reciprocal Rank (MRR) score of the survey	44
5.3	TMQI Score	46
5.4	HDR-VDP Score	47
5.5	Image set 15 generated using various EF methods	48
5.6	Image set 11 generated using various EF methods	49
5.7	Image set 23 generated using various EF methods	52
5.8	Image set 35 generated using various EF methods	53
5.9	Image set 3 generated using various EF methods	55
5.10	Image set 4 generated using various EF methods	56
5.11	Image set 30 generated using various EF methods	57
5.12	Image set 32 generated using various EF methods	58

List of Appendices

1	Gaussian Pyramid	64
2	Laplacian Pyramid	66
2.1	inverse Laplacian	67

Chapter 1

High Dynamic Range Imaging

1.1 Introduction

The primary purpose of an image or a photograph is to capture, record and present the incoming light in order to accurately represent the scene or object in question. A conventional camera can capture light in an 8-bit format (per channel) which, for the most part, can represent a reasonable amount of incoming light accurately in most situations. However, the natural light has a vast range of intensities which a conventional camera cannot possibly capture. A sunlit scene vs. a moonlit scene differs by a factor 10^9 in amount of light hitting the object [1] and thus, representing this range using only 8 bits will cause a lot of relevant information to be lost.

The Human Visual System (HVS) deals with this enormous range of incoming light by dynamically changing its photoreceptor sensitivity and contraction or expansion of the pupil [1]. However, the dynamic response of the eye is still not fully understood and the attempts to replicate the results lie even below our current understanding of the eye. The goal of High Dynamic Range (HDR) imaging is to mitigate the limitation of the standard camera by capturing the full range of incoming light [1].

Recent cameras are equipped with functionality to automatically detecting the scene brightness, allowing the camera to adjust its exposure settings automatically when taking the shots. However, the auto-exposure functionality of the camera still has limitations. A scene containing dark regions as well as bright regions can lose a significant amount of details. For example, a photograph of a room with sunlight streaming through a window will either make the scene outside the window or the room itself

unrecognizable. HDR imaging is a technique that tries to lessen these effects.

HDR imaging attempts to capture the light as faithfully as possible in order to preserve maximum information (which in most cases refers to the delineation of adequate structural detail defining the scene). This can be achieved in two ways: 1) use a camera which can capture the light with more than 8 bits or 2) capture images at a multiple exposures and then combine the most relevant part of each of the image into one final image [1]. The first approach is an expensive solution to HDR imaging, usually requiring a camera with a multiple layers of sensors, each with different sensitivity [1].

Furthermore, most display devices can only display 8-bit images so capturing image with HDR becomes problematic when converting into a range capable for rendering. However, the second approach can be employed by existing cameras. The issue of displaying the HDR image still remains for both methods and therefore, the image needs to be converted back into an 8-bit image before displaying. Reducing the dynamic range for displaying the image is referred to as tone mapping and the traditional 8-bit image is referred to as a low dynamic range (LDR) images.

1.2 Applications

Applications of an HDR imaging are countless. HDR imaging is especially useful in applications involving: 1) Fusion of multiple images or 2) Accurate rendering of light and colour. Examples of applications that perform one of the two functions mentioned above are very wide and numerous including:

- **Scientific and Medical Visualization**

HDR imaging has found many uses in medical imaging where high accuracy is essential. For example, a mammography images are usually taken in multiple shots each captured using different exposure level. Mammography images are taken this way because it is essential to capture every subtle detail. An expert specializing in inspecting the different bands of light is then required for diagnosis. With the advances in HDR imaging techniques (such as tone mapping) this process can be made easy and convenient [2].

- **Physical Based rendering**

One of the most widely used applications of HDR imaging is the real life lighting of virtual objects in a mixed reality environment. Real life lighting using HDR images is termed Image Based Lighting

(IBL) [1]. IBL is used in most game engines (including Unity [3]) for enhanced lighting effects and is also widely used in the making of animated films [1]. Images utilizing IBL are produced by taking multiple HDR photographs of a light probe (usually a reflective sphere) placed at every location where an object needs to be rendered. HDR images of the light probe captures incoming light from every possible angle [1]. The information from HDR photographs is then used for lighting and rendering [1]. A recent advance in IBL utilizes a video of a light probe in order to obtain more accurate lighting effects [4].

- **Multispectral Visualization and Security**

Visualization of "light" outside the visible spectrum has many applications. For example, satellite imagery, infrared photography, ultra sound etc. The captured light data is usually represented as an HDR image. For example, visualization of infrared (IR) images is very useful in security and surveillance to detect intruders [5]. HDR imaging is usually used when a camera need to record a sunlit area from inside of a building [5]. HDR imaging has also found applications in visualization of sonar images for underwater navigation [6].

- **Recognition**

One of the most challenging tasks in image based recognition in an outdoor environment is recognition under constant change of illumination. Accuracy of the classifier is dependent on the time of the day the training data were collected. [7] and [8] have shown that recognition tasks using a HDR images are robust against illumination. Robustness against illumination can significantly improve any recognition tasks (for example, lane detection) in an outdoor environment.

1.3 The Pipeline

Figure 1.1 depicts the traditional pipeline used for creating and displaying images with HDR content using multiple exposures. The first stage in an HDR image acquisition pipeline is the capturing of images with multiple exposures. In this stage, images are first subjected to alignment and registration. From the aligned multi-exposure images the camera's response curve is estimated (stage 2). This involves using the exposure time along with corresponding pixel values to estimate the non-linear relationship between a pixel and its radiance value. Multi-exposure images are then linearised using the estimated

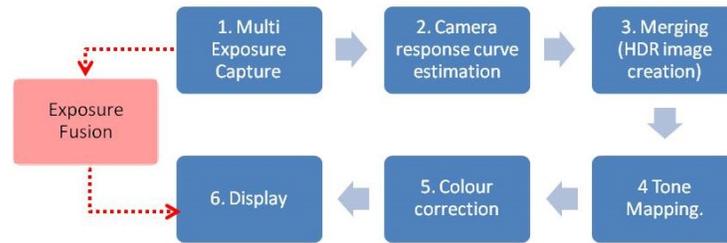


Figure 1.1: HDR Image Acquisition Pipeline

inverse camera response function and fused together in order to generate the HDR image. The fused HDR image is then subjected to de-ghosting and then stored or processed. The HDR image cannot be viewed on conventional monitors since monitors can only display 8-bit images. Therefore the image is subjected to tone mapping and colour correction (steps 4 and 5) before displaying on a monitor.

Most HDR imaging algorithms follow the step 1-6 outlined in Figure 1.1. However, there is a specialized group of algorithms that fuse the images directly into tone mapped images, effectively skipping stages 2-5 of the conventional pipeline. These specialized algorithms, which can directly fuse images into tone mapped images, are termed differently by many authors however they all will be referred as Exposure Fusion (EF), first termed by Mertens et al [9], algorithms in this document for simplicity. The work proposed in this thesis dissertation will improve upon one of these Exposure Fusion techniques. There exist many techniques for Exposure Fusion and one issue that many algorithms share is the accurate displaying of information in bright or dark regions. The Human Visual System (HVS) cannot perceive changes in bright areas as much as dark areas (Weber's law¹ [10]) and therefore, the details in bright regions are not as easily perceived. This problem is amplified when the image is displayed on a large screen monitor due to interpolation used to resize the image.

Another major issue that most Exposure Fusion techniques face is the robustness against a lack of sufficiently exposed images. Current EF techniques are designed to give best results when the exposure sequence contains an equal number of over and under exposed photographs. However, a practical HDR imaging implementation will likely capture either mostly under or over exposed photographs. Having mostly under or over exposed images will yield images which are too bright or too dark when fusing using EF algorithms. Figure ?? displays a typical multiple exposure sequence containing both over and under

¹ Weber's law states that the just-noticeable difference (level at which we can perceive the difference) between two stimuli is proportional to the magnitude of the stimuli. In case of image, Weber's law implies that the small change in bright regions (stimuli) will not be detected as easily as changes in dark regions.

exposed regions for entire scene. On other hand, Figure ?? displays an exposure sequence containing mixed levels of exposures for different regions of the scene. For example, the sky in Figure ?? contains equal number of over and under exposure images whereas the grass region (bottom left) contains only two shots which are slightly over exposed. Using current EF technique on images of Figure ?? will result in an image with a dark grassy region and a bright sky; making the image look unrealistic (Figure ?? (b)-(d)). Figure ?? (d) is the middle exposure shot in the exposure sequence. Figure ?? (d) clearly show that the bottom left grass region appears too dark while the sky is overly bright compared to the grass region. This observation suggests that over or under exposedness is a local attribute and therefore a single exposure shot will be insufficient to capture the whole scene.

This thesis proposes a novel, non-parametric method for directly fusing an exposure sequence into an 8-bit/channel image (skipping the tone mapping step). The proposed method reduces the brightness of a over exposed regions while increases the brightness of a under exposed regions in order that details may be perceived more accurately. The proposed method is an EF technique which is robust against varying exposure times, which may differ from one photographer to next. The proposed method is thus ideal for practical HDR imaging applications.

1.4 Organization

This thesis is organized as follows: Chapter 2 will discuss different techniques of HDR image generation. Chapter 2 will briefly discuss hardware based HDR image generation and focus largely on HDR image generation using multiple exposure shots. Chapter 3 will briefly discuss tone mapping followed by exposure fusion techniques in Chapter 4. Chapter 4 will also present the proposed exposure fusion technique. Chapter 5 will then present the experiment setup and results. Finally Chapter 6 will conclude this thesis and present future work.

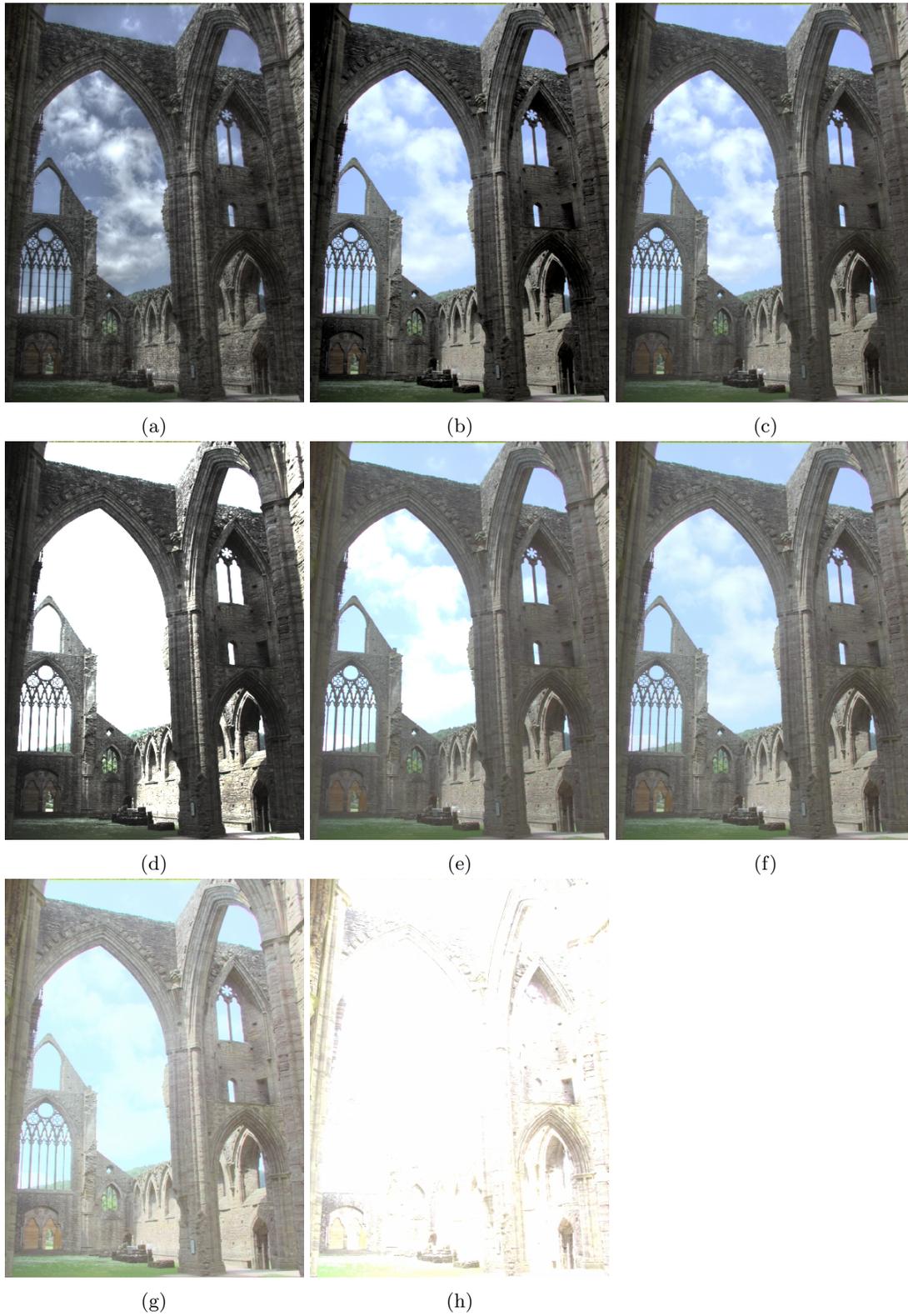


Figure 1.2: Increasing exposure time from (a) to (h)

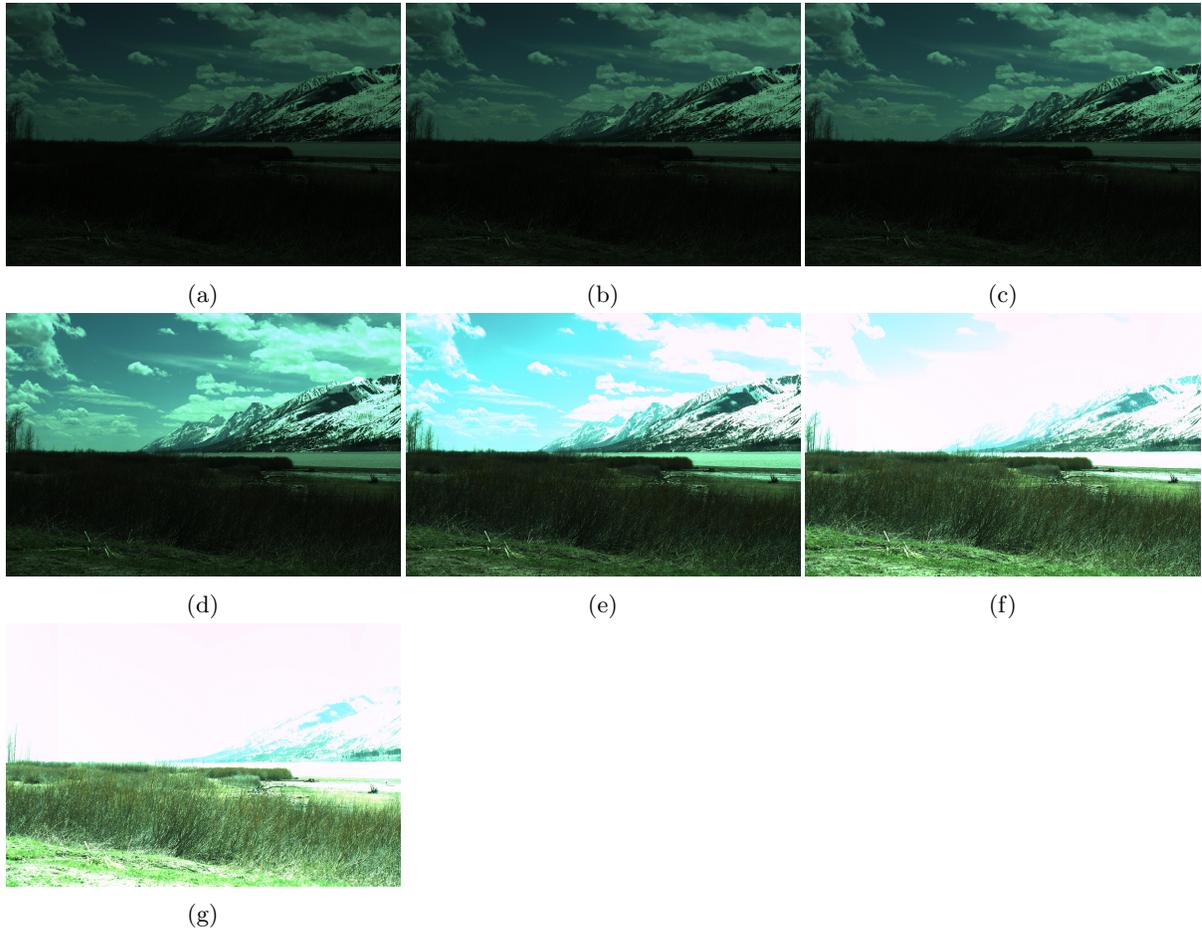


Figure 1.3: Increasing exposure time from (a) to (g)

Chapter 2

HDR Image Generation

2.1 Introduction

A digital camera captures images by refocusing light from a scene through a lens onto an array of light sensors. The lens used for focusing the light is a passive device which does not degrade any incoming information [1] (assuming lens is perfectly clean). The light falling on the sensors then generates electrical signals that get recorded into memory. In non-digital cameras the light falls on a photosensitive film which can then get developed into photographs [1].

Light sensors in general are non linear devices that usually only operates on a specific range of incoming stimuli. Any incoming stimuli outside the operating range of the sensor will over saturate the sensor causing information loss. Cameras control the amount of light falling onto a sensor by adjusting its shutter speed.

Light from a natural scene contains a vast range of light intensities which usually fall outside a sensors' operating range. A single natural scene might contain regions that can over saturate the sensors as well as regions that need large exposure time to be legible. A digital camera cannot locally adjust its shutter speed for capturing light from such a large dynamic range in a single scene.

High Dynamic Range (HDR) imaging tries to solve the problem faced by the conventional cameras by capturing a more extensive range of incoming light. HDR imaging faces two main issues: capture and display. The main issue of capturing the incoming light can be solved at the hardware or software level. Similarly, the representation issue can also be solved at the hardware or software level.

Capturing HDR images at a hardware level involves specialized cameras equipped with light sensors that do not saturate when exposed to large dynamic range of light intensities [1]. Hardware based approach to HDR image generation can directly capture scene with HDR content in a single shot.

A software based approach to HDR image generation involves capturing multiple photographs at different exposures levels and fusing them afterward. The main rationale behind this approach is that HDR scenes that cannot be captured in a single shot can be captured by varying the exposure time thus, ensuring some shots are within the operating range of sensors. The fusion process then locally combines the most relevant part from each exposure shot into one final HDR image. The main disadvantage of using this method is that both the camera and the scene needs to stay perfectly still in all of the exposure shots. A non static scene captured using multiple exposures will cause artifacts, known as ghosting, in the final fused HDR image.

Hardware based HDR image generation is a fairly new technique and it has still not been adopted for traditional photography. Software based HDR image generation is complicated and faces many issues however it can be used by existing devices. Currently, software based HDR image generation is the most popular approach due to its compatibility with existing cameras. However, hardware based HDR image generations is expected to be more in use in the future.

As discussed above, software based HDR image generation is widely popular and therefore, this chapter will mainly focus on that with a brief discussion on hardware based approaches.

2.2 Hardware Based Methods

The hardware based methods of capturing an image with HDR content involves having a specialized cameras that are able to capture the full range of light. Cameras capable of direct HDR image capture are usually equipped with multiple layers of light sensors with different sensitivity. Another methodology used for direct HDR image capture is to use grouping of light sensors with variable sensitivity for each region of the image. For example, [11] used a special camera that uses IcyCAM chips for HDR image acquisition. IcyCAM are SoC (System-on-Chip) chips created by e Centre Suisse d'Electronique et de Microtechnique (CSEM; Neuchtel, Switzerland) that have an optical front end and a processor on the same device. IceCAM chips are able to perform image acquisition, analysis and decision-making on cluster of pixels instead of whole image. Despite the advantages of IcyCAM, it is still in the development

stage and requires vigorous tuning. In [11], IcyCam is used to compensate and reduce the capture of glare.

Hardware based HDR image generation is still in its development stage and thus it is not widely available. However, future cameras will most likely contain direct HDR image capturing capabilities.

2.3 Software Based Methods

As discussed earlier, a major stage in the HDR acquisition pipeline is to capture images at different exposure levels and then fuse them afterward. This section describes some of the techniques used to fuse captured images. Most of the fusion techniques described below requires a tone mapping stage for display purposes. Furthermore, every fusion technique requires that the camera and the scene are perfectly still. As a consequence, Scenes with movement are first subjected to image alignment in order to register the captured images.

2.3.1 Median Threshold Bitmap

Before discussing the fusion of multi-exposure images, it is of great importance to discuss image registration for the exposure sequence. Median Threshold Bitmap (MTB), is by far the most popular image registration technique used for aligning multi-exposure images. MTB is widely used in many HDR related research as well as commercial HDR applications.

MTB was first introduced by Ward [12] and further refined by Pece and Kautz [13]. Ward noted in his paper that registration of multi-exposure images is very different from registration of photographs captured at a single exposure level [12]. Traditional image alignment algorithms rely on edges or other features in an image for matching criteria. However, every photograph in a exposure sequence will generate a different edge map and therefore an image alignment using traditional methods will yield undesirable results. MTB is an image alignment technique which is impervious to varying exposures.

The MTB algorithm initially calculates the median of each exposure shot (grayscale) and thresholding that image to its median value. A pixel will be assigned a value of 0 if it is lower than the median and 1 if it is greater. Each threshold image is then compared with a reference image from the sequence using the XOR operation. The offset of an image is calculated by measuring the difference of XOR values in a region surrounding a pixel. The result of MTB is $(N - 1)I(x, y)$ integer offset at location (x, y) for

each of $N - 1$ exposure images (assuming there are N images in an exposure sequence) from a reference exposure image. Multi-exposure images can then be aligned using the generated offsets.

MTB is a fast image alignment algorithm used prior to any image fusion. MTB also has some noise reduction capability that makes MTB an ideal choice for multi-exposure image alignment.

2.3.2 Image Fusion

Image fusion is the most widely studied and relevant topic in software based HDR image generation. In a multi-exposure image fusion, special care must be taken to retain contrast, natural tone and reduce artifacts. This section will outline the most recent techniques of image fusion.

Pixel-wise image fusion, first proposed by Mann and Picard [14], assumes that the camera and the scene are both completely static. The proposed idea was to do simple weighted average of the corresponding pixels across the exposures.

$$I_d = \frac{\sum \frac{W(v) * f^{-1}(v)}{t}}{\sum W(v)} \quad (2.1)$$

Where:

I_d = Radiance map

\mathbf{v} = captured pixel value at a given exposure

\mathbf{t} = exposure time

$f^{-1}(v)$ = inverse camera response function that maps from pixel value to irradiance

Equation 2.1 is the general framework for a weighted average approach and it requires estimation of camera response function, exposure time and a weight function. The rationale behind Equation 2.1 is that the pixel-wise image fusion is inherently a linear operation thus the argument of the averaging function should also be linear. Applying the inverse camera response function and dividing each pixel by the exposure time will linearize the image. The pixel values are converted into radiance values. After the fusion operation, tone mapping should be applied before displaying or printing. The main challenges in pixel-wise fusion stem from deriving an optimal weighting function.

Mann and Picard initially proposed the weighing function as the derivative of the camera response function arguing that the saturated pixels cannot map to the derivative [14].

Taking motivation from Mann and Picards work, many other weighing functions were proposed. Debevec and Malik [15] proposed a simple hat function arguing that only the mid-range intensities are relevant in each exposure. Mitsunaga and Nayar [16] proposed a weight function to reduce signal to noise ratio. Reinhard et al.[1] combined the work of [15] and [16] by applying a broader hat function to [16]s work. Recently an adaptive Gaussian weighting function was introduced by [17] to accommodate for the case when a great amount of details are in over/under exposed photographs. For example, the Gaussian of over exposed image might have mean closer to 255 while the Gaussian of a under exposed image might have mean closer to 0. The biggest issue of these methods is that the resulting images lose relevant edge information because it is essentially a global process and it does not take spatial information into account.

The problem of weighted average (presented above) was solved by fusing images at different scales and resolutions. Multi-scale image fusion attempts to fuse the sequence by first decomposing the images into different sub images with different resolution (Gaussian pyramid [9] for example). For each location in the transformed image, the value in the pyramid with highest saliency (eg. visual saliency) is selected. The fused image is then reconstructed by taking an inverse transform of the combined images. Xu et al. [18] used this approach to fuse the images in the wavelet domain. The images are first decomposed into wavelet coefficients. The coefficients of each image at each level are then combined using a weighted average method. Finally the fused image is obtained by taking the inverse wavelet transform. Similarly, [19] used wavelet packet transformation for image fusion. The technique used by [20] was very similar to transform domain fusion. In this approach, a gradient of images over multiple resolution is first calculated, the gradient field is then modified in accordance with the multi-resolution decomposition of the original image. Gradients from all the images are then then fused in the multi-resolution domain. Finally, the Poisson equation is used to obtain the fused image and finally the image is linearly stretched to [0-255] range for display.

Finally [21] uses the dictionary learning technique to fuse the image. A dictionary is first trained using K-SVD algorithm [22]. Different outdoor/indoor images were used for training (images were first divided into blocks). Using the trained dictionary, the input images are then decomposed into appropriate dictionary atoms using orthogonal matching pursuit (OMP) [23]. The coefficients are then

merged and recombined with the dictionary to get the final output image.

2.3.3 Ghost removal

Once exposures are aligned and the camera response curve is estimated, it is safe to fuse the images. However if there is motion in the scene, for example, a person walking, the moving object may appear in the final image as ghosts. This is also known as motion blur.

Ghost removal is usually performed after image registration and camera response estimation. Ghost removal consists of two stages: 1) ghost detection and 2) ghost removal.

Motion in a scene can be classified in two ways: 1) moving object on a static background or 2) moving background with static objects. Most ghost detection technique can only detect one of the two types of motion that were described above; very few ghost detectors can detect both.

The ghost detection technique most commonly used, detects ghosts by weighted variance measure [24] described below:

$$VI_{UV} = \frac{\frac{W(Z_{UV}^k) * (E_{UV}^k)^2}{\sum W(Z_{UV}^k)}}{\left(\frac{W(Z_{UV}^k) * (E_{UV}^k)^2}{\sum W(Z_{UV}^k)}\right)^2} - 1 \quad (2.2)$$

Where:

\mathbf{UV} = image pixel location at (U,V)

Z_{UV}^k = pixel value Z of exposure k at position (U,V)

E_{UV}^k = Estimated radiance value of exposure k at location UV

Equation 2.2 describe a general formation for ghost detection by comparing variance in the image. The main purpose of variance calculation is to detect regions in the image that had experienced motion. Variance of a region affected by a motion will be high while variance of a region not affected by motion will be low, therefore the variance map can give an indication of the location of a moving object.

Variance based ghost detection is only one of the many ghost detection algorithm available. There are also algorithms based on entropy and more recently ghost detection is carried out by optical flow algorithms [24].

The final aim of HDR image capture is to remove all ghosting artifacts. One way a Ghost can be removed after ghost detection stage is by applying separate fusion algorithm on the affected areas. This

is done by assigning weights to the affected regions based on its probability of being static or not. The ghost removal stage is usually performed during the fusion process.

2.4 Summary

This chapter outlines the range of approaches to HDR image generation. with each approach having its own advantages and disadvantages. Hardware based HDR image generation involves utilizing specialized cameras while software based image generation can generate HDR images using existing devices. The main advantage of the hardware based approach is that it is fast and it is robust against artifacts caused by motion. However, it requires special hardware which might not be available. On other hand, software based HDR image generation can be employed by existing devices. However, it requires many pre-processing stages (alignment, camera response estimation and ghost removal).

Chapter 3

Tone Mapping

3.1 Introduction

High Dynamic Range (HDR) images cannot be viewed on conventional 8-bit/channel monitors. The process of compressing the HDR content to displayable low dynamic range (LDR) is termed *tone mapping*. The core issue in tone mapping is not only the compression of dynamic range, but compressing the dynamic range in such a way that the information that is "thrown" away is not perceptually relevant. The Human Visual system (HVS) sensitivity range far exceeds the light sensitivity of a camera sensors and thus the tone mapping operation will ultimately degrade the quality of the original HDR image. Degradation due to a tone mapping operation needs to be controlled in order to maximize the outcome of the application at hand.

Tone mapping can also be regarded as a nonlinear quantization process that aims to quantize 32-bit image (HDR image) into an 8-bit image (LDR image) [1]. Every process that performs A/D conversion faces quantization problems. The problem of tone mapping has existed for a long time before HDR imaging came into existence however, it was not a major issue with traditional LDR photographs due to low sensitivity of the sensors [1]. However, it is currently a major challenge in HDR imaging.

The difference in contrast seen in a natural environment cannot be reproduced by conventional cameras and thus in traditional photography it falls to the photographer or the artist to convey this information in a most natural way whilst remaining in the display range of the monitor. For example, a photograph of a person or an object with sun or a major source of light directly behind him/her will cause

a severe darkening effect making the person unrecognizable (the photographer must then compensate by choosing an exposure setting that compromises for a trade-off between detail either in the foreground or background of the captured scene). The most accurate and natural visualization of captured contrast is the key issue of tone mapping.

Various tone mapping operators (TMO) have been introduced thus far and they fall into two main categories: 1) Global TMOs and 2) Local TMOs. Global TMOs are computationally efficient however, they fail to encode subtle contrast difference that make an image appear detailed. On the other hand, a local TMO is able to encode all the details but are computationally expensive. Another issue faced by local TMOs is the halo effect which occurs when the contrast change in a region is too strong (this is not an issue for a global TMO) [1]. Halo effect is an artificial light or dark ring that surrounds an object, like the halo on top of an angel. Due to the nature of the tone mapping problem, the use of various visual models has become the backbone of many TMOs. This chapter will first detail the visual models used by major TMOs and then present the tone mapping operators themselves.

3.2 Stevens' Power Law

Equation 3.1 is the general form of Stevens' law describing the relationship between any physical stimuli and the magnitude of its perceived intensity. Stevens' law states that any physical stimulus (e.g. light, touch, taste etc.) has an exponential relationship between the amount of physical stimuli and its perceived magnitude. Stevens' law models human perception to physical stimuli and therefore many of the tone mapping operators are based on it.

$$S = k * I^\alpha \quad (3.1)$$

Where:

S = Sensation magnitude

k = Adjustable constant

I = Stimulus intensity

α = Power constant dependent on modality

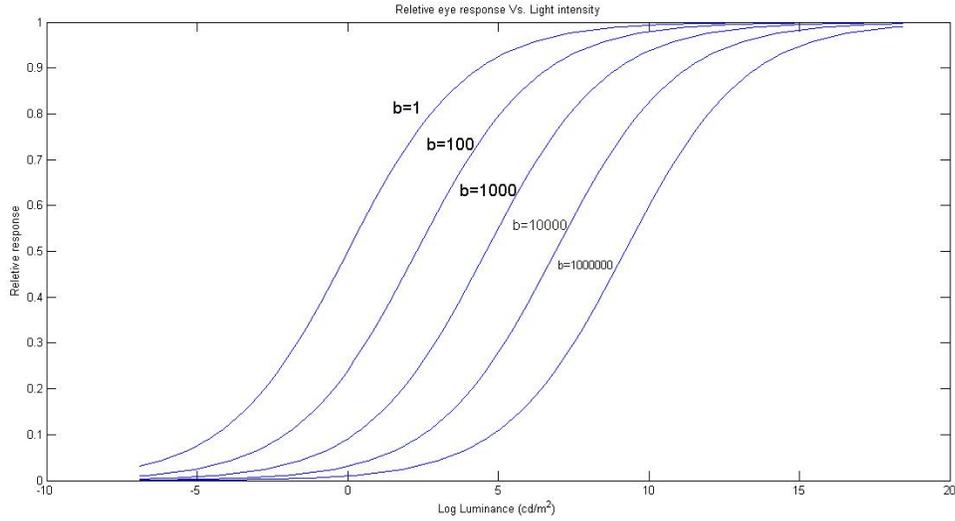


Figure 3.1: Naka-Rushton Model

3.3 Naka-Rushton Model

Psychophysical studies have shown that the photoreceptors response to the incoming light is a sigmoid function when plotted on a log linear graph [1]. Study have shown that the eye response to the incoming light is not only dependent on the light intensity itself but the light hitting the surrounding regions as well [1]. The eye response models the incoming light as a sigmoid function which can effectively model over and under-saturation. The sigmoid function will adjust its position on the light intensity axis based on surrounding light intensities if the receptor stay saturated for too long, ensuring that the eye response does not stay saturated. The eye response can be modeled by the Naka-Rushton equation (Equation 3.2).

$$\frac{R}{R_{max}} = \frac{I^n}{I^n + b^n} \quad (3.2)$$

Where:

I = Radiance image

R = Eye response

R_{max} = Maximum eye response

\mathbf{b} = Semi-saturation constant

\mathbf{n} = Sensitivity control exponent

Parameter b in Equation 3.2 is a constant that is dependent on background intensity. Figure 3.1 shows the plot of the relative eye response curve over a range of background light intensity. Figure 3.1 clearly shows that changing the b value will shift the sigmoid function in the intensity axis and thus allowing the eye receptor to handle large dynamic range of incoming light.

3.4 Retinex

Retinex, the combination of the word retina and cortex, is another simplified model of the HVS. Retinex theory was first proposed by Land in 1971 [25]. Land proposed that the HVS decomposes an image into illuminance and reflectance, latter of which is used in human perception (Equation 3.3). The key challenge in the Retinex model is the decomposition of an image into both its illuminance and reflectance component. Many algorithms have been proposed to do this. Tone mapping operators based on Retinex theory attempt to map the reflectance to display range and are local in nature due to the local nature of calculation involved to predict reflectance.

$$I = R * I \tag{3.3}$$

Where:

\mathbf{R} = Reflectance

\mathbf{I} = Illuminance

Retinex theory is also used extensively in other image processing fields with an alternate name. The Retinex model can also be regarded as a homomorphic filter, where the task is to filter the illuminance from the image. Homomorphic filters are mainly used to filter multiplicative noise. Although the Retinex model and homomorphic filtering share many similar elements, they both use very different methodology to decompose the image into its respective components.

3.5 Global Tone Mapping Operators

Global TMOs are usually logarithmic or S-shaped functions which operate on each pixel individually. Global TMOs do not take spatial relationship of the pixels into account, so they tend to blur the edges of the tone mapped image.

A relatively large class of TMOs uses S-shaped functions with the intention of mimicking the response of the photoreceptors in the eye. A global TMO based on a photoreceptor model is designed to apply a sigmoid or a piece wise logarithmic function on every pixel of the image. On other hand, a local TMO will vary the shape and the operating range of a sigmoid function based on the neighborhood of a pixel.

One of the earliest TMO was introduced by Tumblin and Rushmeier [20]. Tumblin and Rushmeier's TMO is a global TMO based on Stevens' law that aimed to compress the global luminescence to the display range by mapping the perception curve (Stevens' law) with the model of the display device.

$$n = \left[\frac{L_w^{\frac{\alpha(w)}{\alpha(d)}}}{L_{d,max}} * 10^{\frac{\beta(w) - \beta(d)}{\alpha(d)}} - \frac{1}{C_{max}} \right]^{\frac{1}{\gamma}} \quad (3.4)$$

$$\alpha(L) = 0.4 * E(\log(L)) + 2.92 \quad (3.5)$$

$$\beta(L) = -0.4 * (E(\log(L)))^2 - 2.58 * E(\log(L)) + 2.02 \quad (3.6)$$

Where:

γ = Display γ

C_{max} = Maximum display contrast

d = Display luminance (constant)

w = Real world luminance

L_w = Real world luminance

$L_{d,max}$ = Maximum display luminescence

$\mathbf{E}(\log(\mathbf{L}))$ = Expected value of $\log(\mathbf{L})$

Equation 3.4 - 3.6 denote the Tumblin and Rushmeier's tone mapping operator. Tumblin and Rushmeier's operator was specifically designed for grayscale images only and therefore it is not suited for colour images. Furthermore, Tumblin and Rushmeier's operator is a global tone mapping operator that requires display specific parameters which might not be available in many situations.

Ward and Rushmeier [26] later introduced a TMO based on histogram adjustment. Ward and Rushmeier operator introduces a novel histogram adjustment technique based on local adaptation population in the scene. The Ward and Rushmeier operator incorporates models for human contrast sensitivity, glare, spatial acuity, and colour sensitivity to match subjective viewing experience. The histogram adjustment operator is still a global TMO however it is partially local; in the sense that it chooses its parameters based on local attributes.

As discussed earlier, many of the tone mapping operators are based on some aspect of visual model, Naka-Rushton model discussed above is just one of them. Schlick [27] proposed a tone mapping function that is based on the Naka-Rushton model. The tone mapping equation is:

$$F(I) = \frac{I}{I + \frac{I_{max} - I}{p}} \quad (3.7)$$

Where:

I = World luminance

I_{max} = Maximum World luminance

p = Adjustable constant

Equation 3.7 is a special case of Equation 3.2 where "n" = 1 and "b" = $\frac{I_{max} - I}{p}$. The Schlick operator aims to compress the overall dynamic range without any explicit edge or contrast enhancement.

The most popular local TMO used is a TMO based on Reinhard operator [28]. The Reinhard operator first finds how bright a scene is, which is known as the *key* of that image. For example, a night scene will have a low key compared with a day scene. The Key of the scene is estimated using Equation 3.8 which is the log average of the image.

The Reinhard operator first estimates the key of the scene by using Equation 3.8. The second step in Reinhard operator is to use the Equation 3.9 to scale the image appropriately. The final low dynamic range image is computed using Equation 3.10. There are lots of variations of Reinhard's operation such

as variable alpha selection based on local regions in order to make the Reinhard operator a local operator.

$$key = \frac{1}{N} \exp\left(\sum \log(\delta + L_w(x, y))\right) \quad (3.8)$$

$$L(x, y) = \frac{\alpha}{key} L_w(x, y) \quad (3.9)$$

$$L_d(x, y) = \frac{L(x, y) \left(\frac{L(x, y)}{L_{white}^2}\right)}{1 + L(x, y)} \quad (3.10)$$

Where:

N = Total number of pixels

α = Adjustable parameter

L_{white} = White balance value

3.6 Local Tone mapping operators

One of the first local TMO was proposed by Chiu [29]. Chiu's TMO attempts to compress the dynamic range by utilizing a blurred version of the input image. Chiu also incorporates glare removal in its tone mapping algorithm. Chiu's TMO is:

$$L_d(x, y) = \frac{L(x, y)}{k * L_w^{blur}(x, y)} \quad (3.11)$$

Where:

$L_d(x, y)$ = rone mapped image at position x amd y

$L_w^{blur}(x, y)$ = Gaussian filtered image with window size

As with all local TMOs, Chiu's TMO will produce halo artifact in the boundary between light and dark regions.

Retinex based TMOs can be classified as local TMOs that are inspired by the human visual system. They encode the reflectance of an image to a displayable range, enhancing the local contrast in process.

The Original retinex model introduced by Land was named random path retinex. The random path retinex assigns a set of random 1-D paths originating from each pixel location. The ratios of adjacent pixels in a given path are then averaged in order to obtain the reflectance estimation from a given path. The final reflectance value at a given pixel location is then calculated by averaging all the reflectance values obtained from each path. Lands algorithm utilizes random 1-D path for its reflectance calculation and as such resulting image depends on number of paths per pixels and the size of the path. Furthermore, due to the random nature of the paths, the final solution is not always the same for a given image.

The path based retinex was then improved by [30]. [30] introduced the random spray retinex where spray of pixels around the test pixels are used for reflectance calculations. The random spray retinex has many free parameters and it is not always easy to tune for optimal results however unlike path based retinex, the spray retinex yield a constant stable solution for a given image,

Another method of reflectance calculation was introduced by [31]. The new version of retinex, named center surround retinex, estimates the value of reflectance by the ratio of the pixel to its local average.

$$L_d = \log(L_w) - \log(L_w^{blur}) \quad (3.12)$$

Where:

L_d = Tone mapped image

L_w = Gaussian filtered image(HDR)

$L_w Iblur$ = Gaussian filtered image (HDR)

Equation 3.12 shows the general form of center surround retinex. There are many different formulations of blurring filter used in Equation 3.12, each with its own advantages and disadvantages

3.7 Discussion

Figure ?? show a toned mapped image generated by various tone mapping operators. Images shown in Figure ?? have many differences and it is up to the observer to decide the overall quality of the image. Figure ?? (a) and (c) appears "washed out" and dull compared to Figure ?? (b) which seems to be detailed. However, Figure ?? (b) has an overly bright sky region and therefore, the details in the clouds

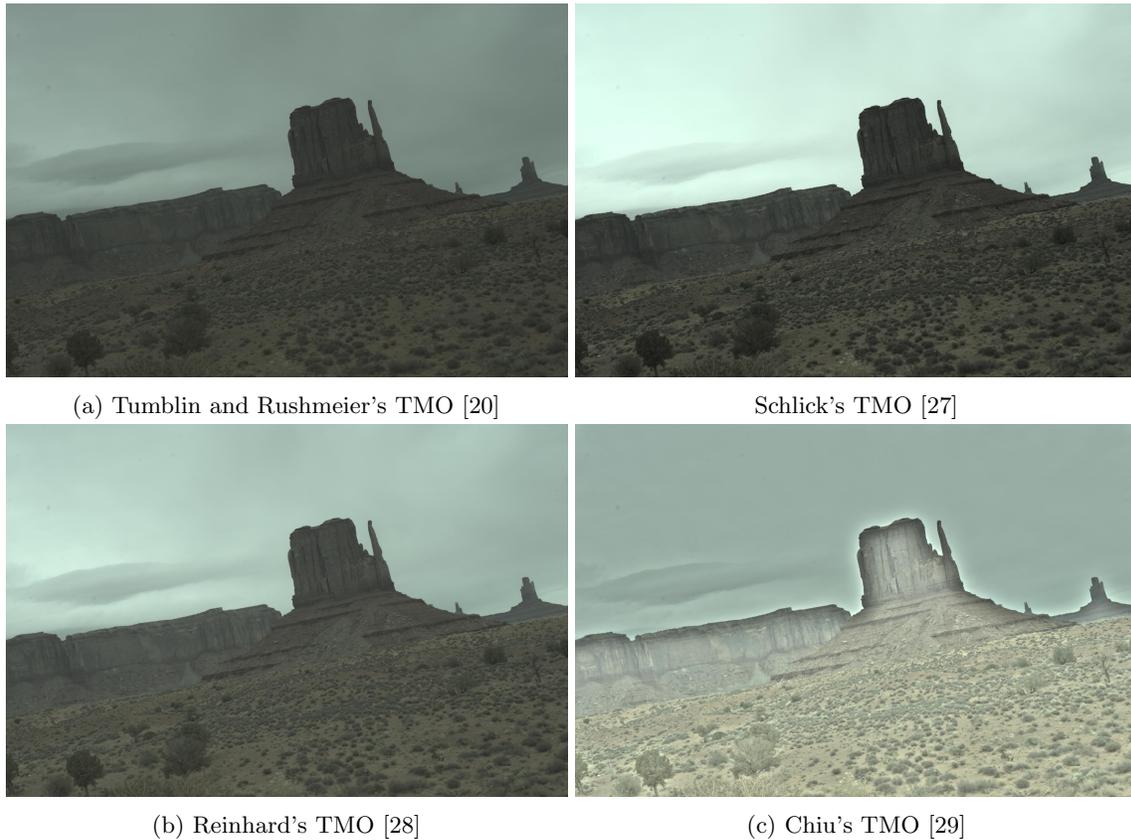


Figure 3.2: Tone mapped image generated using various TMOs

appears attenuated. The ground region in Figure ?? (b) also appears to be darker than that of Figure ?? (a).

Figure ?? (d), which is generated using a local TMO, has overly bright ground regions however, the details do not appear to be lost. The sky region also appears detailed. One major drawback of every local TMO is the halo effect. Figure ?? (d) shows a dark rim of halo around hill where it meets the sky.

Of the TMOs presented in Figures ?? (a)-(d), Reinhard TMO (Figure ?? (b)) appears to generate the best image. However, effectiveness of a TMO is evaluated depends on the particular application on which it is performed and other TMOs might be applicable where Reinhard TMO fails.

3.8 Summary

This chapter outlines the importance of tone mapping operators in the HDR imaging pipeline in that they are required to display the image. There are many tone mapping operators available, each with its own advantages and disadvantages. However, most TMOs are based on some aspect of the human visual system that they try to emulate. TMOs are grouped in two categories, global and local. Global TMOs are computationally efficient, but they tend to dull the edges. Local TMOs can preserve the edges better, but they show halo artifacts where the bright region meets the dark.

Chapter 4

Exposure Fusion

4.1 Introduction

Exposure Fusion is a term given to a group of algorithms that attempts to fuse multi-exposure images into a single displayable image, skipping the tone mapping stage. Furthermore, Exposure Fusion does not require the calculation of the camera response curve as it is not applying tone mapping. There are many different algorithms available for estimating camera response curve, tone mapping and HDR image generation. Each method in each of these categories has its own advantages and disadvantages. And the primary advantage of skipping these three steps in EF algorithms is that finding the optimal method for best results can be avoided. Only one algorithm (EF) will now be needed to be optimized instead of three. Another advantage of Exposure Fusion (aside from no tone mapping) is that resulting images tend to be more perceptually detailed. This can be explained by noticing that Exposure Fusion does not require an intermediate HDR image and therefore, it is not subjected to distortions caused by different fusion methods.

Disadvantages of Exposure Fusion also arise from the lack of intermediate HDR image. Exposure Fusion techniques can only be used for applications that involve image visualization, and are not useful for image based recognition for example. Another disadvantage of the Exposure Fusion is that it does not take exposure time into account during the fusion process. Exposure time, along with camera response curve, is used to process the pixels in the same domain. Without the exposure time, the overall tone of the fused image will differ from the physical scene. For example, an exposure sequence containing

5 under exposed images and 2 over exposed images will most likely generate (from Exposure Fusion) image that is too dark, which might not represent the physical scene well.

This chapter will present the most recent techniques used for Exposure Fusion, its limitations and subsequent motivation for the current work, and a newly proposed methodology for EF.

4.2 Existing Methods

The most influential Exposure Fusion algorithm was introduced by Mertens et al [9]. Mertens' algorithm begins by calculating a series of quality metrics which are *contrast*, *saturation* and *well exposedness* for each of the exposure shot. The quality metrics indicate how important a given exposure shot is in the final fusion process. Mertens, in his paper, stated that the combination of the product of each of the normalized quality metrics should be used as weights to generate the final fused image. Mertens also noticed that fusing images with the combined quality metrics as weights will cause artifacts. Much like the halo artifacts, directly fusing the multiple exposure images will cause *seams* in the boundaries between dark and light regions. In order to solve the *seam* problem, Mertens fuses the exposure sequence in the multi-resolution domain. The quality metrics for each of the image is first decomposed into a Gaussian pyramid while the image itself is decomposed into Laplacian pyramid. Each level of the Gaussian pyramid is multiplied with each level of the Laplacian pyramid. Each level from each exposure shots is then added together to get the final decomposed image. Finally, the fused image is obtained by applying inverse Laplacian to the decomposed image (Algorithm 1).

Laplacian pyramid decomposes the image into a series of band pass images. While the Gaussian pyramid decomposes the image into a series of low pass images. Multiplication of a band pass image with a low pass image is equivalent to a convolution operation in the frequency domain, which results in a band pass image weighted by the low pass image. Combining the images in multi-resolution fashion can be considered as fusing each band pass image with a different weighting function. Mertens' algorithm, which was just described, is summarized below:

Figure 4.1 shows a pictorial representation of Algorithm 1. Figure 4.1 clearly show that each level of the Laplacian pyramid represent a band pass image while each level of the Gaussian pyramid represents a low pass image. The blurring of the quality metric is essential in order to avoid artifacts. Proposed method will improve upon Mertens's EF algorithm.

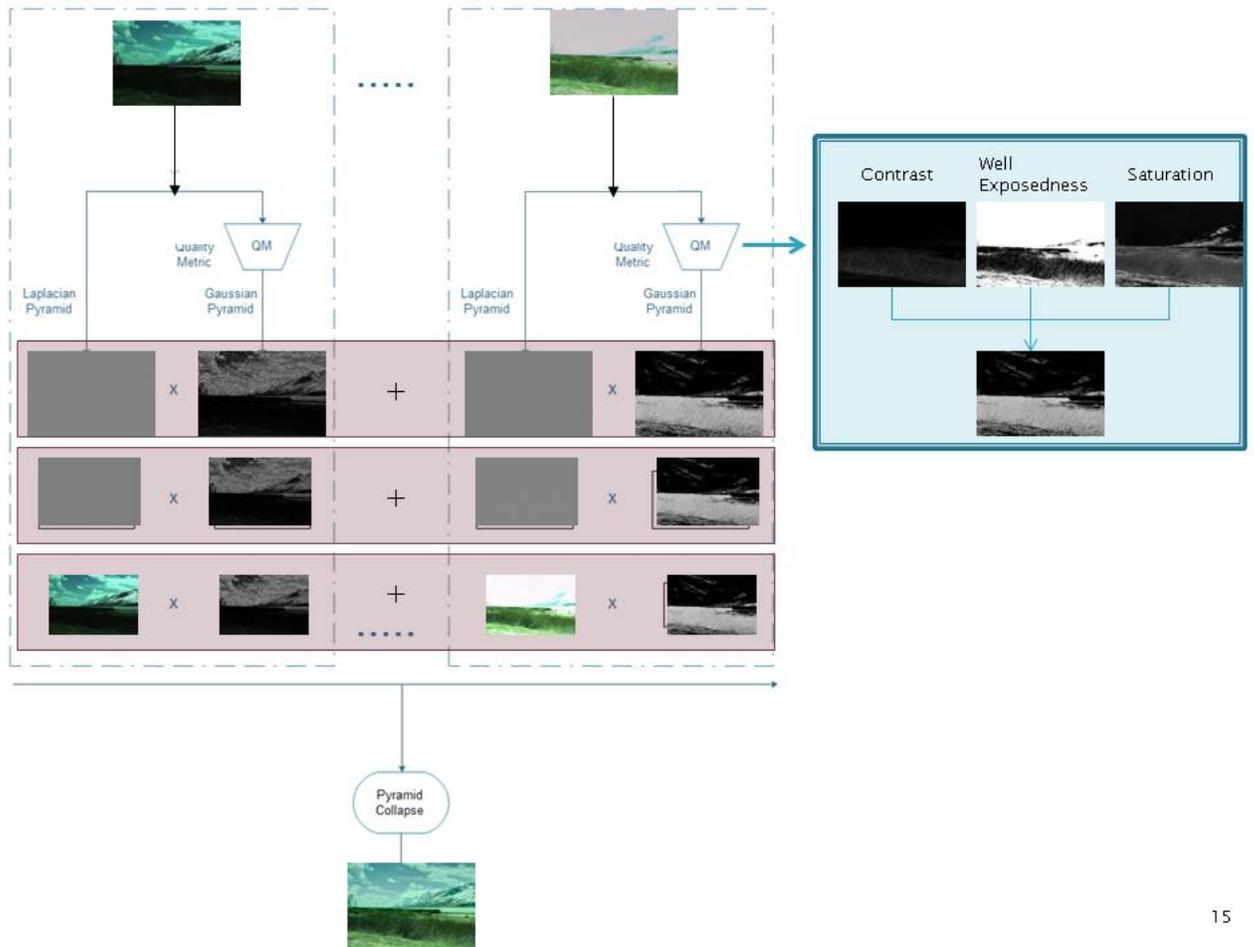


Figure 4.1: Mertens' Exposure Fusion algorithm

A slightly different method of EF was proposed by Raman and Chaudhuri [32]. Raman and Chaudhuri used the difference between the image and the bi-lateral filtered version of the same image. for each exposure shot as the weighting function for image fusion, which aims to preserve weak edges.

A more recent method by Neil Bruce [33] fuses the sequence in log domain using local entropy as a weighting function. The main rationale behind Bruce's method is that a local region with maximum information (entropy) should be weighted more strongly, thus representing a non-linear as opposed to linearly weighted blending of images.

Algorithm 1 Mertens' Exposure Fusion

```

C = scale(contrast(LDR_images);
S = scale(saturation(LDR_image));
W = scale(well_exposedmess(LDR_images))
Quality_measure = normalize(C*S*W);
for Each image I in the input sequence do
  pyr = Laplacian_pyr(I)
  pyrG = Gaussian_pyr(Quality_measure)
  for L - number of levels do
    R[L] = pyr[L]*pyrG[L]
  end for
end for
reconstruct_oyr(R)

```

4.3 Adaptive Exposure Fusion

4.3.1 Motivation

There are two primary methods used to display HDR images to regular monitors. The first method involves the creation of a HDR image and then displaying that HDR image using a tone mapping operation. The second method used to display an image with HDR content is by directly fusing the exposure sequence into a displayable image (Exposure Fusion (EF)). The main difference between these two methods is that the first method, which generates the HDR image using Equation 2.1, utilizes exposure time information (among other things) for HDR image generation. While the second method does not use any exposure time information. The variables introduced during final image generation can result in very extreme differences in the final image when there are no guidelines to what exposure time settings to use (currently it is left to user to decide). Figure ?? shows the extreme difference between tone-mapped vs. exposure fused image is when the exposure time for shots is not calibrated correctly. The tone mapped image in Figure ?? appears overly bright while the exposure fused image appears dark. There is no way to correctly identify the tone of the scene. The image in Figure ?? (a) appears to be taken in mid-afternoon where the sun is bright whereas Figure ?? (b) appears to be taken at dusk. Main reason for the bright tone mapped image is that it was generated from HDR image while the image generated using Exposure Fusion was not. The rationale behind the previous statement is that the division of every pixel by its exposure time in Equation 2.1 (used to generate HDR image) will yield a larger value for pixels corresponding to shorter exposure time than pixels corresponding to longer exposure time. For example, an exposure sequence containing five under exposed images and only two

over-exposed images will yield a dark image if generated by Exposure Fusion algorithms whereas the same sequence fused by the traditional HDR imaging pipeline will not. The issue discussed above is an issue faced by all of the Exposure Fusion methods and is yet to be solved.

The following section thus proposes a novel non-parametric pre-processing stage to Exposure Fusion image generation which aims to mitigate the effect of images taken with improper exposure time. The Exposure Fusion algorithm introduced so far do not employ any pre-processing before fusion. The pre-processing needs to be applied before the fusion stage because the images themselves need to be improved before any further processing.

4.3.2 Proposed Method

As discussed earlier, Exposure Fusion by definition does not utilize exposure time information when fusing multiple images taken at a multiple exposure levels. This issue of exposure time in EF techniques is not as noticeable for exposure sequences containing equal (roughly) number of over and under exposed images thereby giving final fused image that is somewhere between the two extremes. However, the issue of exposure time is much more noticeable in a practical HDR imaging scenarios. In a practical HDR imaging scenario, without any guidelines for best the exposure time settings, an exposure sequence will most likely contain unequal numbers of over and under exposed images causing the fusion result to favor the set of images which are dominant in term of exposure time. The aim of the proposed method is to mitigate the negative effect of exposure time and, at the same time, yield more natural appearing images.

Another issue with most of the EF techniques is that these algorithms do not consider how natural the resulting image appears. In [34], the authors modified the Mertens' algorithm to work in $L^*a^*b^*$ colour space in order to generate more natural appearing image. $L^*a^*b^*$ is a non-linear colour space whose aim is to linearize the HVS perceived colour difference in the image. The main issue with Mertens method (and its modification in [34]) is that details in the darker regions appears *washed out*. Details in the bright regions are severely attenuated when image is resized for larger display. The main goal of the proposed method is to increase the contrast of the most relevant region of the image by darkening the overly bright regions while brightening the dark regions of an image.

The proposed algorithm first converts the exposure sequence into $L^*a^*b^*$ space as [34]. The proposed algorithm then applies *Exposure Adjustment* algorithm (proposed method) to the L^* channel of the



Figure 4.2: a) Tone mapped image; Exposure Fusion Image

exposure sequence. As per the original EF algorithm by Merten (shown in Algorithm 1), *contrast*, *well exposedness* and *saturation* metrics are then computed (from exposure adjusted images). [34] has shown that the retinal-like-subsampling-contrast (RSC) metric [35] is more perceptually accurate measure of contrast in an image and it has been used by [34] to generate more natural appearing exposure fused image. The RSC uses the difference of Gaussian (DoG) filter mask to measure the edges instead of Laplacian operator used by Merten's method. The present implementation of the proposed algorithm replaces RSC filter mask with Laplacian of Gaussian (LoG) filter because LoG is more computationally efficient and yield similar results to RSC.

The fusion of images is then performed in a multi-resolution fashion. Each image is decomposed into Laplacian pyramid while the *saturation*, *well exposedness* and *contrast* metrics (after multiplication and normalization) are decomposed into a Gaussian pyramid. At each pyramid level the Laplacian and the Gaussians pyramid levels are multiplied together. The resulting product at each pyramid level of each of the input images is then added together. The pyramid is then collapsed and converted back to RGB colour space to obtain the final image. The proposed algorithm is summarized in the following page.

4.3.3 Exposure Adjustment

As mentioned earlier, current EF technique do not employ any pre-processing before the image fusion and this has caused many issues with regards to exposure time. The goal of the *Exposure Adjustment* is to pre-process the exposure sequence in such a way that every image in the sequence have equal

Algorithm 2 Proposed Adaptive Exposure Fusion

```

LDR_images = Exposure_Adjust(LDR_images);
C = scale(contrast(LDR_images);
S = scale(saturation(LDR_image));
W = scale(well_exposedmess(LDR_images));
Quality_measure = normalize(C*S*W);
for Each image I in the input sequence do
  pyr = Laplacian_pyr(I)
  pyrG = Gaussian_pyr(Quality_measure)
  for L - number of levels do
    R[L] = pyr[L]*pyrG[L]
  end for
end for
reconstruct_oyr(R)

```

perceptive contrast. The proposed *Exposure Adjustment* equation is shown in Equation 4.1 [36].

$$I(x, y) = I(x, y) * \exp(k - I_{ave}(x, y)) \quad (4.1)$$

Where:

k = brightness level

I = Exposure adjusted image

I_{ave} = input image filtered with 3x3 averaging mask

Equation 4.1 assumes that pixel values ranges between 0-1. Equation 4.1 applies a simple exponential scaling to the original image. The local average is a good measure of amount of brightness in a region surrounding the given pixel, (*I_{ave}*), therefore the local average is used to determine the local exposure adjustment levels. The *k* parameter of Equation 4.1 determines when a pixel should be amplified or attenuation and by how much. For example, any value of *I_{ave}* larger than *k* will attenuate the pixel while any value of *I_{ave}* smaller than *k* will amplify the input pixel value. Any *I_{ave}* values equal to *k* will not affect the pixel values at all. The resulting pixel value is not limited to 0-1 range and therefore the output of the *Exposure Adjustment* needs to be scaled to a values between 0 and 1 using simple contrast stretching equation before further processing. Figure 4.4 shows the results after setting *k* parameter of Equation 4.1 to 0,0.2,0.4...1. The most important aspect to note from Figure 4.4 is that *k* value of Equation 4.1 determines how dark or bright the resulting image will appear. For example, setting *k*=0

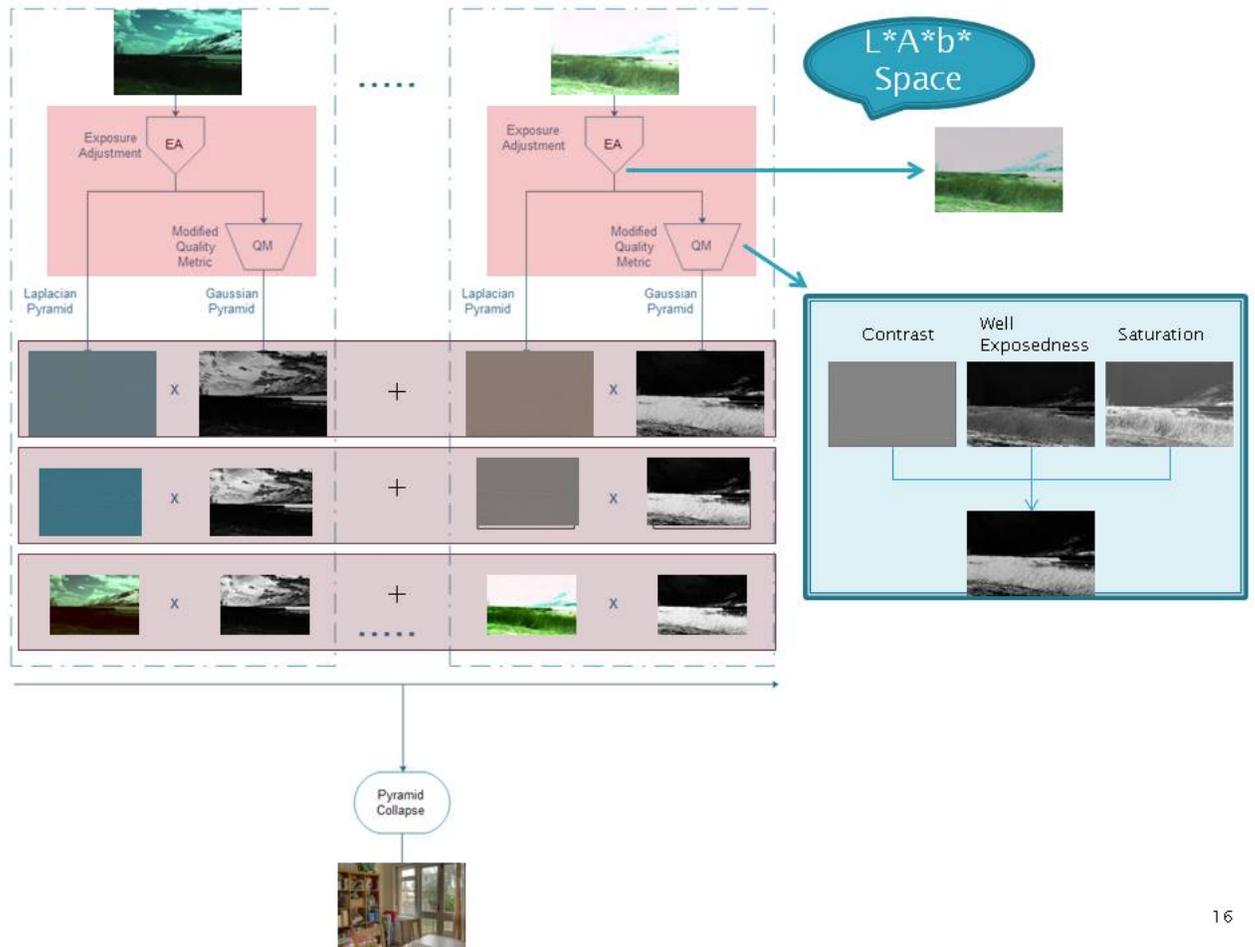


Figure 4.3: proposed Exposure Fusion algorithm

will result in very dark image compared to setting $k=1$. The k parameter, in effect, will determine the tone of the image.

I_{ave} is an image filtered by 3x3 averaging mask and the filter size was chosen based on trial and error basis. A large filter mask causes unwanted artifacts in the image as it blurs the edges severely. A small filter kernel can effectively predict the background luminance value without deteriorating the edges. The main drawback of having small filter kernel is that it cannot predict the global tone of the image giving undesirable effects in some cases, which will be discussed later.

Equation 4.1 is also an approximation of how the HVS operates. Equation 4.1 combines the local averaging aspect of the Naka-Rushton function and the exponential nature of Stevens' power law. Unlike Stevens' power law, Equation 4.1 applies an exponential weighting to the image as opposed to exponential

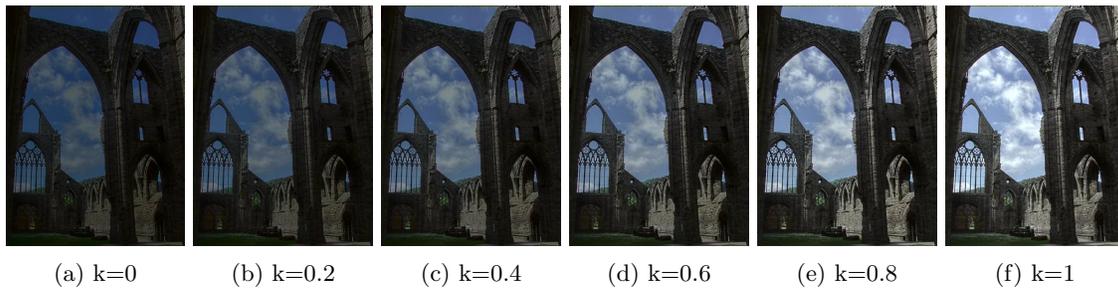


Figure 4.4: Result of equation 4.1 after setting various k values

mapping and this change was adopted from Naka-Rushton function. Equation 4.1 was not designed to mimic the functionality of the HVS and therefore it does not have any experimental evidence confirming its correlation with the HVS. However, it does indicate some possible correlation with known HVS behavior and thus, future research can be aimed at checking Equation 4.1's correlation with HVS.

4.3.4 Adaptive k

Equation 4.1 is the general equation that describes how the exposure of a given image should be adjusted. As discussed earlier, the k parameter of Equation 4.1 determines the overall tone of the image and it can be left as a parameter that the user can set. In order to make the *Exposure Adjustment* algorithm more robust and non-parametric, this paper also proposes a possible scheme to set the k values adaptively based on intuitive reasoning that can compensate for Weber's effect and enhance details.

An image usually contains many types of regions but these regions can be classified into two major groups: 1) light or dark regions and 2) detailed or no-detailed regions. In order to overcome the limitation of the HVS, it is natural to increase the contrast by darkening the overly bright regions and similarly the dark regions need to be amplified. However, Weber's law and other visual phenomenon suggest that only informative regions need to be adjusted and the information free regions of the image can remain un-altered. Rationale behind the above statement is that humans tend to focus on edge rich informative region more than flat information free regions. For example, it is much easier to focus on a small black dot on a piece of white paper than a single spot on a blank paper. The HVS then increases the contrast of object of our focus by darkening bright areas and brightening the dark areas to perceive the details more accurately. Hence, the contrast of the informative regions needs to be increased in order to better perceive the details.

Based on the rationale presented above the k value needs to satisfy the following:

- bright and high informative region - low k value
- dark and low informative region - low k value
- bright and low informative region - high k value
- dark and high informative region - high k value

The first 2 points will insure that only the informative region in the bright area of the image will darken (to compensate for Weber's effect). Similarly, the last 2 points will insure that only informative regions in the dark area will get brighter.

As discussed earlier, an image filtered using an averaging mask is a good indication of how bright or dark the region surrounding a pixel is. Edges in an image constitute for the high frequency region of the image, therefore it can be considered as highly informative. An image filtered using a Laplacian of Gaussian (LoG) mask can be used as an indication of informative region in the image. Using an image filtered by an averaging mask and a LoG mask to determine the k values can provide an image of localized k values instead of a constant over the entire scene.

Equation 4.2 is a simple scheme to calculate the k value satisfying all four points discussed above.

$$k = (1 - L_{edge}) * (L_{ave}) + (L_{edge}) * (1 - L_{ave}) \quad (4.2)$$

Where:

L_{ave} = L channel (of L*a*b* space) filtered with 3x3 averaging mask

L_{edge} = L channel filtered with 3x3 LoG mask

The first part of Equation 4.2 satisfies the first 2 points while the second part of Equation 4.2 satisfies the last 2 points

Figure 4.5 clearly shows that Equation 4.2 satisfies all of the 4 points. Furthermore, Figure 4.5 shows that the k values are smooth over its operating range. Figure 4.7 shows the k values corresponding the exposure sequence of Figure ???. Figure 4.7 clearly shows that the dark grassy region (bottom left) of the scene appears much brighter in under-exposed images than the corresponding image in Figure ???.

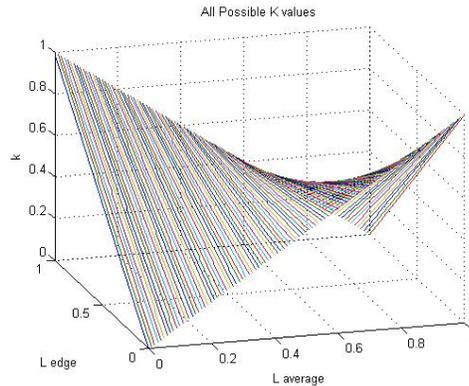


Figure 4.5: Graphical representation of equation 4.2

Similarly, the sky region in Figure 4.7 is darker in over exposed images than its corresponding image in Figure ???. The k map of Figure ??? in Figure ??? also emphasizes details more clearly than its corresponding originals.

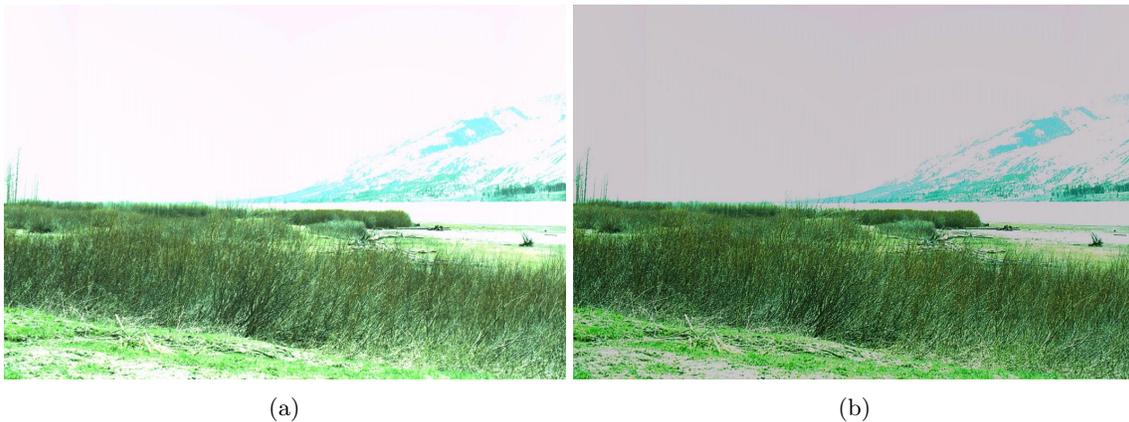
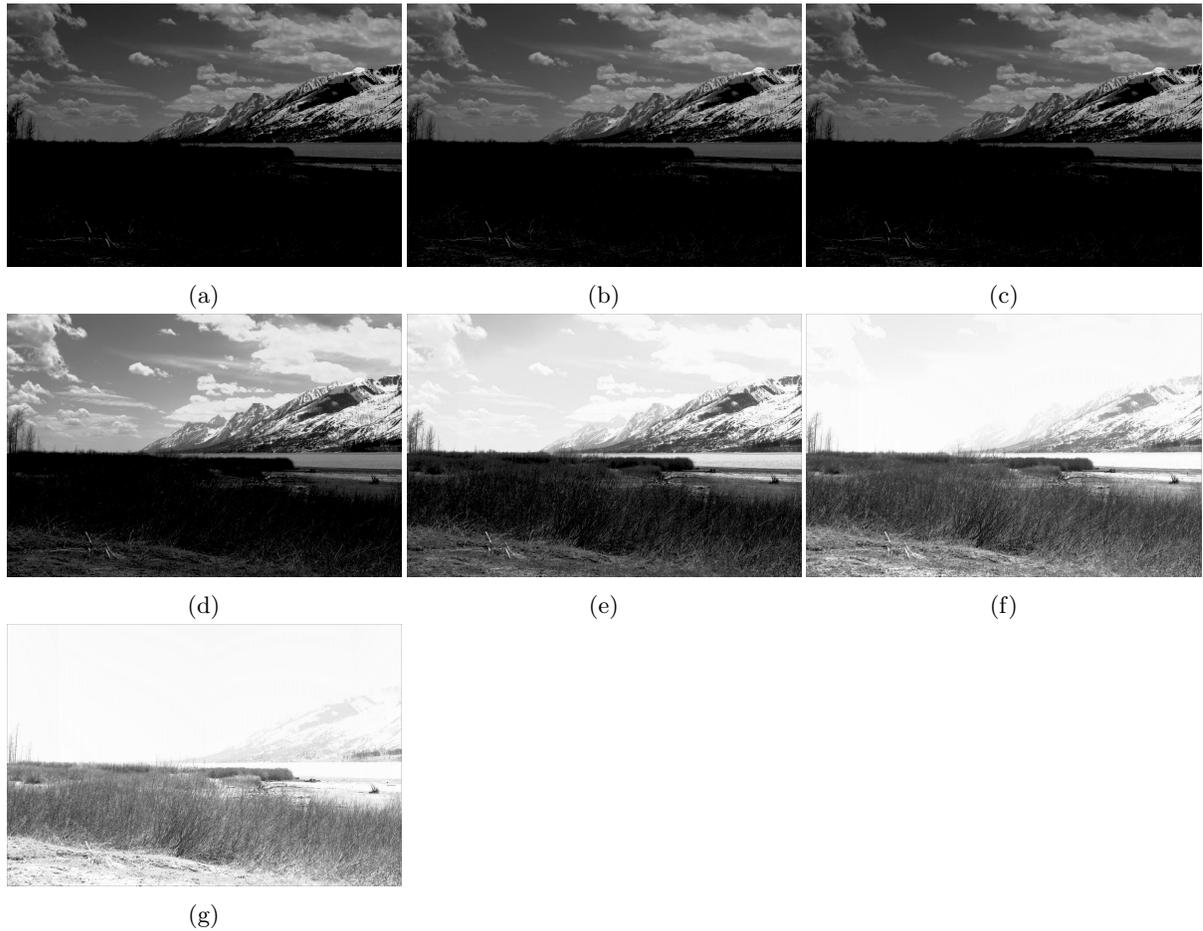


Figure 4.6: (a) Comparison of original exposure image before and after Exposure Adjustment a) Original exposure image; (b) After application of Exposure Adjustment stage

Figure 4.6 shows an image before and after *Exposure Adjustment*. Figure 4.6 clearly shows that the dark twigs in the bottom left hand corner, for example, of Figure 4.6 (b) is much brighter than the same region of Figure 4.6 (a). Figure 4.6 (b) appears to be enhancing edges in the dark regions by brightening them, and at the same time, the edges in the bright regions (sky) are darkened. The overall effect of brightening the dark regions and darkening the bright regions is that overall contrast of the image is increased.

Figure 4.7: Adaptive k value for an exposure sequence

4.3.5 Contrast, Saturation and Well Exposedness

The *contrast* is calculated by filtering the image with a Laplacian of Gaussian (LoG) filter mask. The LoG filter was chosen for the current implementation based on the results of [34] and [35] which showed that a Difference of Gaussian (DoG) filter can yield a better indication of a perceived contrast in an image. The DoG filter is a close approximation of LoG filter and therefore a LoG filter was used in the present implementation.

Figure 4.8 (a) and (b) shows the resulting image with and without *Contrast* metric respectively. The main purpose of the *Contrast* metric is to give high weights to the edge pixels in the final averaging stage. The edges in Figure 4.8 (a) so appear slightly brighter than of Figure 4.8 (b) which lacked the *Contrast* metric. The overall effect of the contrast metric do not appear to be significant to the final



Figure 4.8: (a)Image generated using all three metric;(b) Resulting image with the absence of Contrast metric

image however small detail enhancement is necessary for some image visualization applications.

Saturation is defined as the measure of colourfulness in the picture. The *saturation* is calculated as the absolute difference of channels a^* and b^* . The rationale behind it is that for a region containing no colour, the a^* and b^* component of the $L^*a^*b^*$ colour space will have the same value (0 saturation) while a colourful region would have different a^* and b^* component (high saturation).



Figure 4.9: (a)Image generated with all three quality metrics; (b) Resulting image with the absence Saturation metric

Figure 4.9 (a) and (b) shows the resulting image with and without *Saturation* metric respectively. The image in 4.9 a) appears to have slightly more vivid colours then that of 4.9 b)however the overall feel of the image is not affected.

The *Contrast* and *Saturation* metric has been modified from the original methods introduced by

Mertens et al. The reason for this change was simply to add slightly enhanced naturalness and realism to the image.

Well Exposedness is calculated as a Gaussian weighting of the L channel with mean of 0.5. Figure



Figure 4.10: (a)Image generated with all three quality metrics;(b) Resulting image with the absence Well Exposedness metric

4.10 (a) and (b) shows the resulting image with and without *Well Exposedness* metric respectively. The image in Figure 4.10 (a) appears to be bright and more pleasing to the eye than image in Figure 4.10 (b). And this indicates that the *Well Exposedness* metric has a significant impact on the final image. Figure 4.10 (a) also appear to be more colourful than that of Figure 4.10 (b) and this can be explained by noting that the *Well Exposedness* metric will brighten colours that were previously too dark to notice. This significant effects of *well exposedness* on the final results is due to the proposed *Exposure Adjustment* stage. The *Exposure Adjustment* stage proposed in this paper aims to bring the edge pixels closer to the mean of the Gaussian, giving them higher weights in the final fusion/averaging process. The *Exposure Adjustment* stage allows more pixels to have impact on the final merging/fusing process.

Figure 4.11 (a)-(c) shows an example of resulting quality metric of the exposure shown in Figure 4.6 (b). Figure 4.11 (d) shows the resulting quality metric after merging and normalization (across all the exposures) of Figure 4.11 (a)-(c). The quality metric is calculated across the entire exposure sequence and then decomposed into a Gaussian pyramid (for each exposure image). Each level in the Gaussian pyramid is multiplied (pixel wise) by its corresponding level in the Laplacian pyramid of the original image. Due to the normalization of the quality metric, the quality metric acts similar to weights in the averaging process. The exposures are combined in a multi-resolution fashion in order to ensure that

artifacts caused by sudden spatial change in weights is avoided.

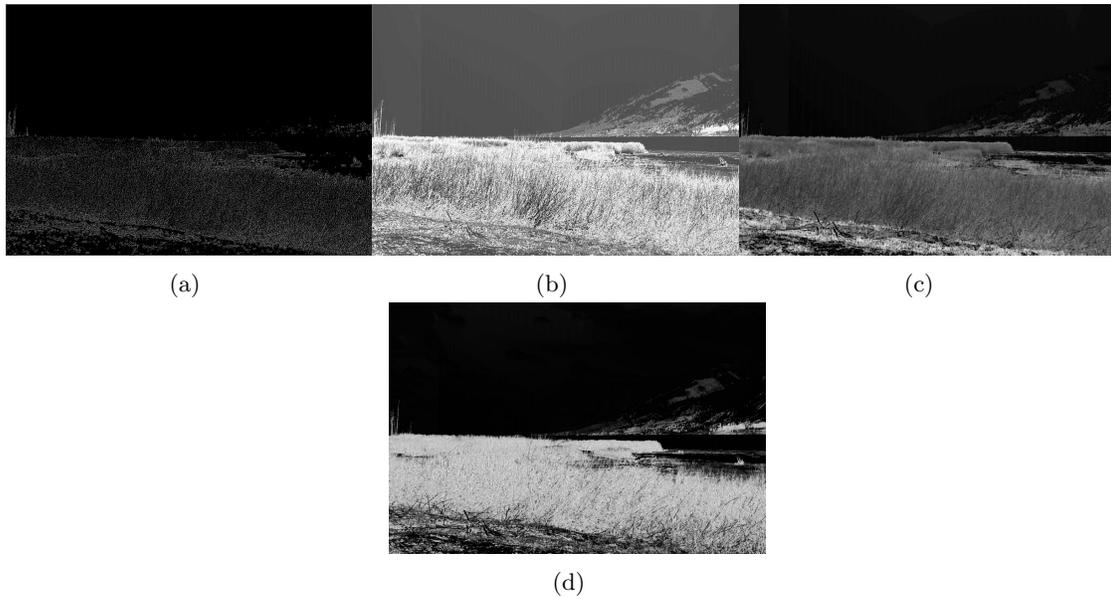


Figure 4.11: visualization of the quality metric: (a) Contrast; b) Saturation; c) Well Exposedness d) combined quality metric

Merging the exposure sequence in a multi-resolution will yield a Laplacian pyramid, which is then needed to be collapsed in order to generate the final fused image. Figure 4.12 (a) shows an example of weight metric decomposed into Gaussian and pyramid. While Figure 4.12 (b) shows an example of original image decomposed into Laplacian pyramid. Figure 4.12 (c) is the resulting pyramid after multiplication of Figure 4.12 (a) and (b). Finally Figure (d) is obtained by adding the multiplied (Gaussian and Laplacian) pyramids of all exposures. The entire fusion process is shown in Figure 4.3.

4.4 Summary

This chapter presented Exposure Fusion techniques which directly fuse the exposure sequence into displayable image. The main advantage of Exposure Fusion is that it does not require tone mapping stage however Exposure Fusion techniques are limited to visualization applications only.

This chapter also proposes a novel pre-processing algorithm for images generated using Exposure Fusion. The proposed method aims to brighten the dark regions while darkening the bright regions. Proposed method is also robust against varying nature of exposure time used to capture the images

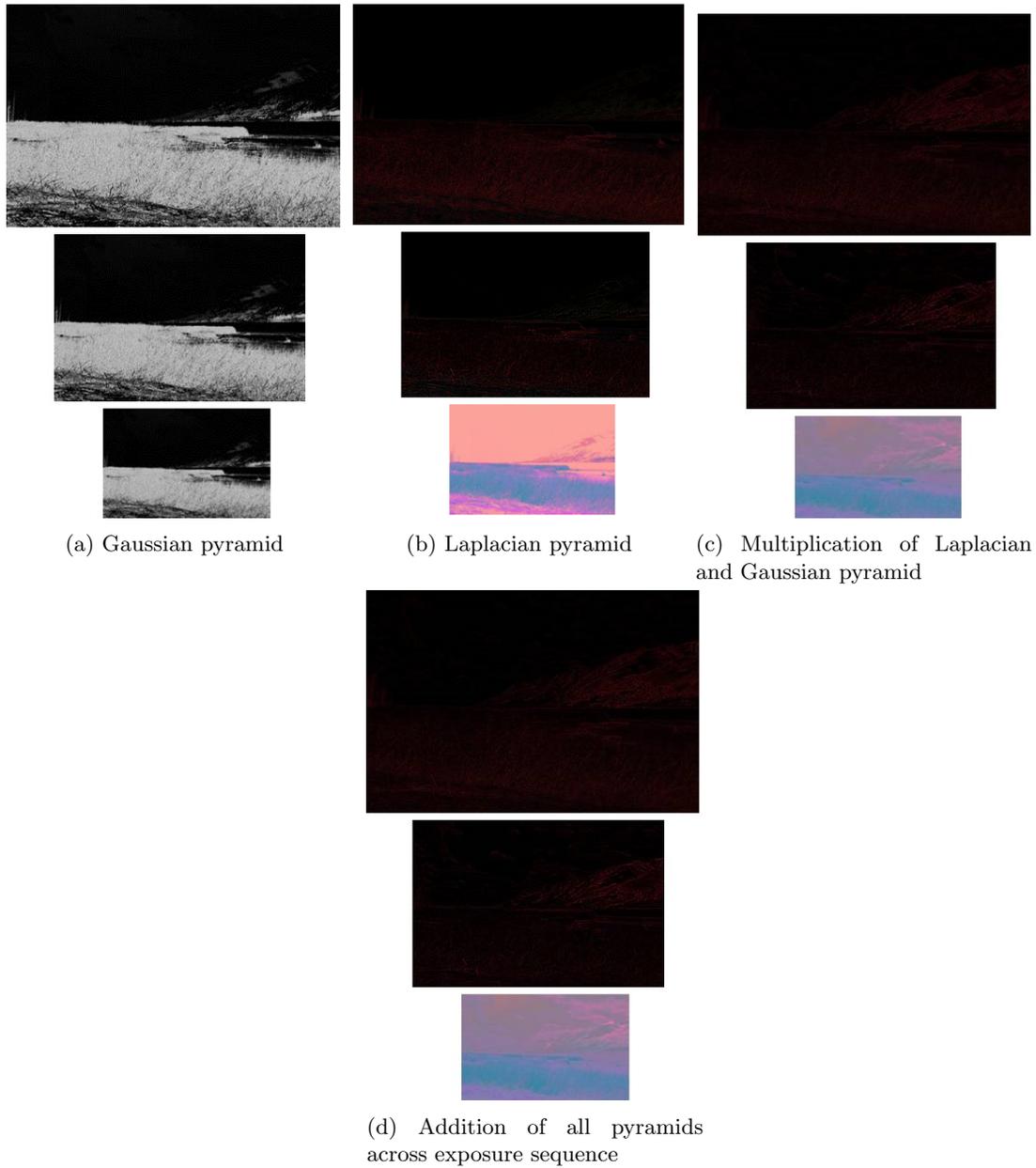


Figure 4.12: Resulting pyramids decomposition of one exposure shot

Chapter 5

Experimental Setup and Results

5.1 Introduction

Every Image enhancement technique is best evaluated by human subjects. In order to test the effectiveness of proposed algorithm, multi-exposure images of each scene (full dataset 1 and 20 select images from dataset 2) were fused using state of the art EF algorithms and compared with proposed method. Images generated using proposed algorithm was not compared with a tone mapped images because the EF technique can only be compared with other EF method namely the algorithm by [34], [33] and [32].

5.2 The Dataset

The proposed algorithm was tested on two publicly available datasets. The first dataset on which the proposed algorithm was tested on was provided by Yeganeh and Wang [37]. The photographs in this dataset are a compilation of images used in papers (HDR imaging related) dating as far back as 1997. Furthermore, the photographs in this dataset were captured by an expert photographer and contains sufficient number of over and under exposed images. The dataset consists of photographs of 15 scenes taken with a standard camera. Each scene was photographed using 8 different exposure levels (which are also available). The dataset also provide a single HDR image (fusion of the 8 multi-exposed images) in a .hdr format.

The second dataset that the proposed algorithm was tested on was provided by Funt et al. [38]. The

dataset consists of 105 scenes taken with Nikon D700 digital still camera. Each scene in the dataset was photographed using between 3-9 exposure levels. Furthermore the capture photographs do not contain equal number of over and under exposed images.

5.3 HDR Image Evaluation Site

A website was created for the sole purpose of evaluating the proposed method with other similar methods. Figure 5.1 shows the general setup for the survey. The site presents the user with a random scene generated using four different algorithms (proposed method,[34], [33] and [32]). All 4 images for a particular scene are displayed to the user on a single page. The user then assign each image with number between 1 and 4 (with 1 being the best and 4 the worst) using the draggable numbers also available on the page. Note: the position of the 4 images is also randomized. The user presses next button for the next scene. Figure 5.1 (b) shows the web page used for the survey. The larger image at the bottom displays a zoomed version of the selected image. There are 35 scenes in total, 15 scenes from dataset 1 and 20 selected scenes from dataset 2. The survey ends when the user have rated all 35 scenes.

Readers are encouraged to take the survey at hdreval.artform.ca.

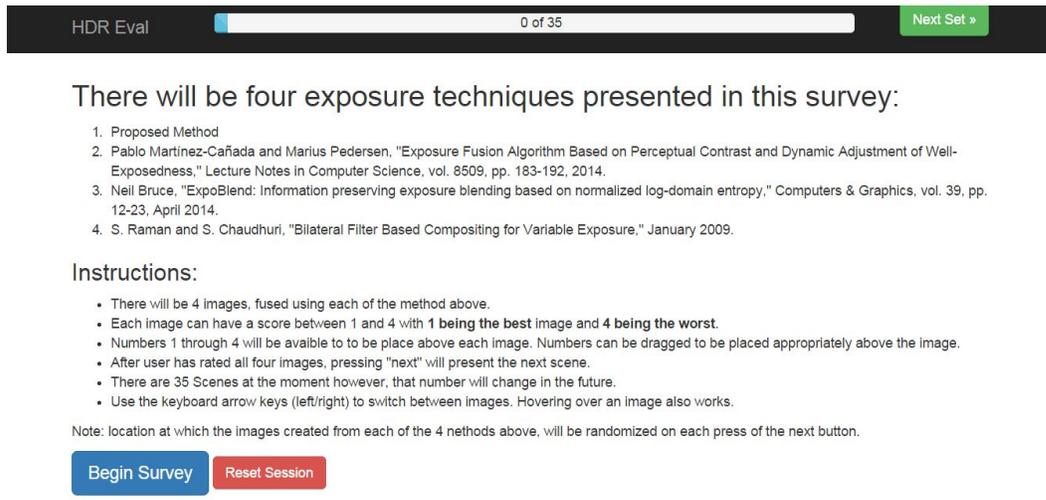
5.4 Mean Reciprocal Rank (MRR)

Mean Reciprocal Rank (MRR) (Equation 5.1) is a statistical tool used to evaluate any process that is evaluated relatively. The evaluation procedure described above rate the images relatively, meaning it evaluates a given algorithm compared with other. The inverse in Equation 5.1 implies that the each ranked image is inversely "better" then the next image. For example Equation 5.1 implies that the image with rank of 1 is 50% better than the image with rank of 2. This method is not the most accurate measure of comparison, however it is a good approximation.

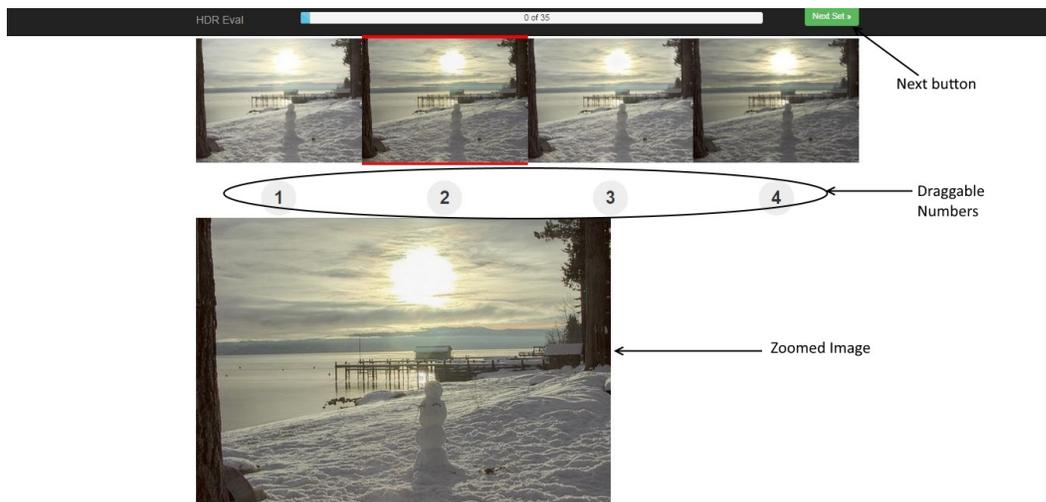
$$MRR = \frac{1}{N} \sum \frac{1}{Rank} \quad (5.1)$$

Where:

N = Number of different algorithms (4 in this case)



(a) Front Page



(b) Survey Page

Figure 5.1: evaluation site

Rank = Rank of the image (with 1 being the best)

5.5 Results

Figure 5.2 shows the evaluations results using the most recent exposure fusion algorithms. The proposed algorithm was evaluated by 25 observers (non-expert) using the evaluation site discussed previously. Users were asked to rank in order of most visually appealing image to least visually appealing image.

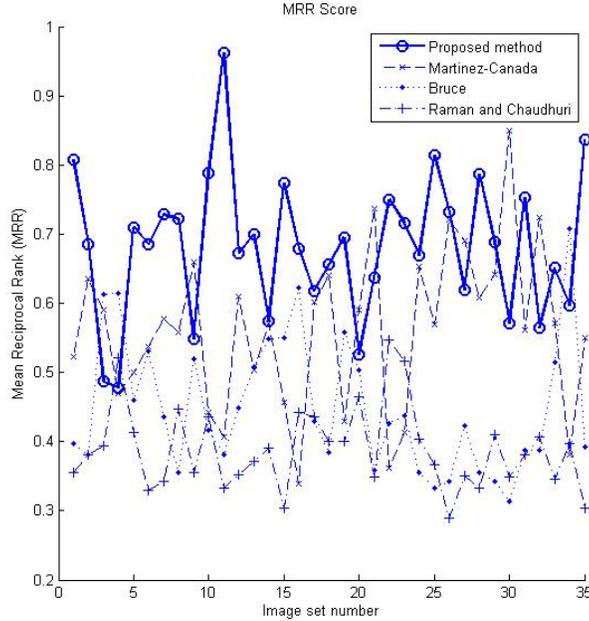


Figure 5.2: Mean Reciprocal Rank (MRR) score of the survey

Users were not asked to change any display settings of their monitors and as such, this was a uncontrolled experiment. Figure 5.2 shows the MRR score of the 35 scenes used for the evaluation. Image set number 1-15 comes from dataset 1 which contained ideal number of over and under exposed images. Image set 16-35 comes from dataset 2 which does not contain fixed number of exposures (per scene) and the exposure times are not evenly distributed. The MRR score for each dataset is shown in Figure 5.2.

The magnitude of MRR score in Figure 5.2 is not as important as its relative placement.

5.6 Comparison with other HDR image evaluation metric

Most algorithms for HDR image processing are evaluated subjectively by the author but there have been some attempts at an objective evaluation. Tone Mapping Quality Index (TMQI) [10] is an evaluation tool used by many recent HDR researchers to rate a tone mapped image objectively. Tone mapping, as discussed earlier, is a term used for compression of HDR images into LDR images.

TMQI evaluates the tone mapped image by comparing its structural information (local standard deviation) with that of HDR image. HDR image is taken as a reference [10]. TMQI is not a suitable

metric to evaluate images based on EF techniques because there is no intermediate HDR image used and thus, reference is not the HDR image but the exposure sequence itself. Figure 5.3 shows the TMQI score for the images in dataset 1 and 2 as discussed above.

Another popular evaluation tool used by many algorithms is HDR-VDP [39]. HDR-VDP stands for High Dynamic Range-Visual Difference Predictor and is a modified version of the popular method VDP [40]. Visual Difference Predictor evaluates an image based on its similarity with a reference (HDR image). HDR-VDP employ complex visual pathway models and other part of the HVS to determine how likely a human observer will be able to visualize the difference (error). As with TMQI, HDR-VDP uses the generated HDR image for reference. Figure 5.4 shows the HDR-VDP score for completeness.

5.7 Discussion

The results of the survey along with the results of widely used evaluation metric (TMQI and HDR-VDP) are presented in Figure 5.2-5.4. Figure 5.2 clearly shows that the proposed algorithm is widely favored for exposure sequence containing equal number of over and under exposed images (image set 1-15) and also for exposure sequence containing uneven number of over and under exposed images (image set 16-35). A noteworthy aspect of Figure 5.2 is that the image fusion using proposed algorithm consistently performs better than other algorithms. Specifically, the EF method by Bruce's shows significantly decreased performance while method by Martinex-Canada significantly improve after image set 15. As mentioned earlier, image set 1-15 consists of scenes photographed with sufficient number of over/under exposed images while image set 16-35 does not. Figure 5.2 indicates that Bruce's method does not perform well when number of exposure images vary significantly.

Figure 5.3 shows the TMQI results of fused images that were used in the survey. The TMQI does not correlate with the results of the survey. The magnitude of the TMQI score is not as important as its relative placement however, TMQI score still does not correlate with the survey results. The TMQI score indicate that method 2 ([34]) is preferable for image set 1-15 which is in sharp contrast to the survey results. The TMQI score also suggests that the proposed method's performance is only slightly better than other method for image set 16-35, which is also a sharp contradiction to the survey results.

Figure 5.4 shows the HDR-VDP results of the fused images. The HDR-VDP score clearly show that it is in very sharp contrast to the survey results. HDR-VDP score indicates that the proposed method

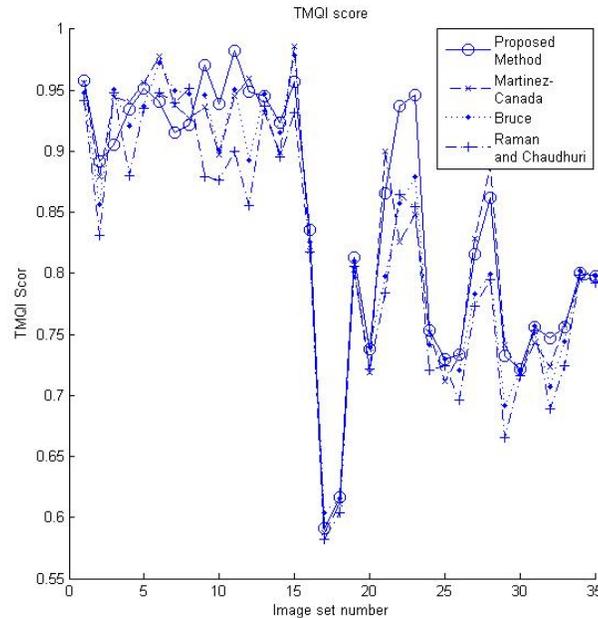


Figure 5.3: TMQI Score

performs poorest of all, which the survey shows otherwise. Results of applying image evaluation metric do not correlate with the result of the survey. The discrepancy between the image evaluation metric and the survey is most likely caused because the images that were being evaluated were not tone mapped images but exposure fused images. Which also suggests that this metric is not entirely reflective of the image quality itself, but rather, as a metric, is quite dependent on the fusion process/mechanism employed. TMQI and HDR-VDP evaluates the quality of an image by comparing it with a reference HDR image. Tone mapped images are a processed HDR images and therefore the HDR image can be used for reference. The exposure fused images are result of processing of multiple images which cannot be used in TMQI or HDR-VDP for a reference.

As discussed earlier, HDR-VDP measures the visible difference between the reference image and the image in question. The reference image (HDR image) represents the income radiance while the tone mapped image represent the response of a light sensor to incoming light. The tone mapping will warm an image for better visibility and therefore the difference between the tone mapped image and an HDR image will be significant. The HDR-VDP score appears to drop significantly after image set 15 and that can be explained by noting that image set 16-35 contains images that were generated from mostly under

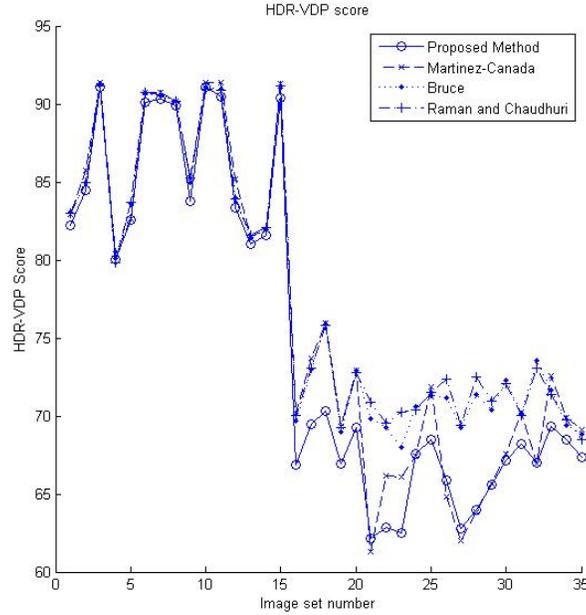


Figure 5.4: HDR-VDP Score

exposed images (dark images). The HDR image associated with each exposure sequence will be bright because they are not dependent on exposure time. While the image generated using the exposure fusion will be dark because they are dependent on exposure time, thus the difference between the HDR image and the exposure fused image will be high. There are no images in the current database whose MRR score matches the HDR-VDP score.

TMQI evaluate an image based on its structure similarity with the reference (HDR image). As discussed earlier, the HDR image corresponding to image set 16-15 will be bright and therefore the TMQI will reward a bright image over a dark image. Image set 16-35 generated mostly dark images and therefore these images will be more perceptually pleasing if they are bright. RMQI score also reward bright images and therefore some level of correlation is expected with the MRR score. There were 6 images from the dataset whose TMQI score matched the MRR score and five these images belonged to dataset 2 (image set 15-35). Although there was some correlation between MRR score and TMQI score, TMQI is still not suitable for evaluating images generated using Exposure Fusion. TMQI failed to give accurate results for majority of the images.

Figure ??(a)-(d) show the fusion results using the most recent exposure fusion algorithms. Exposure

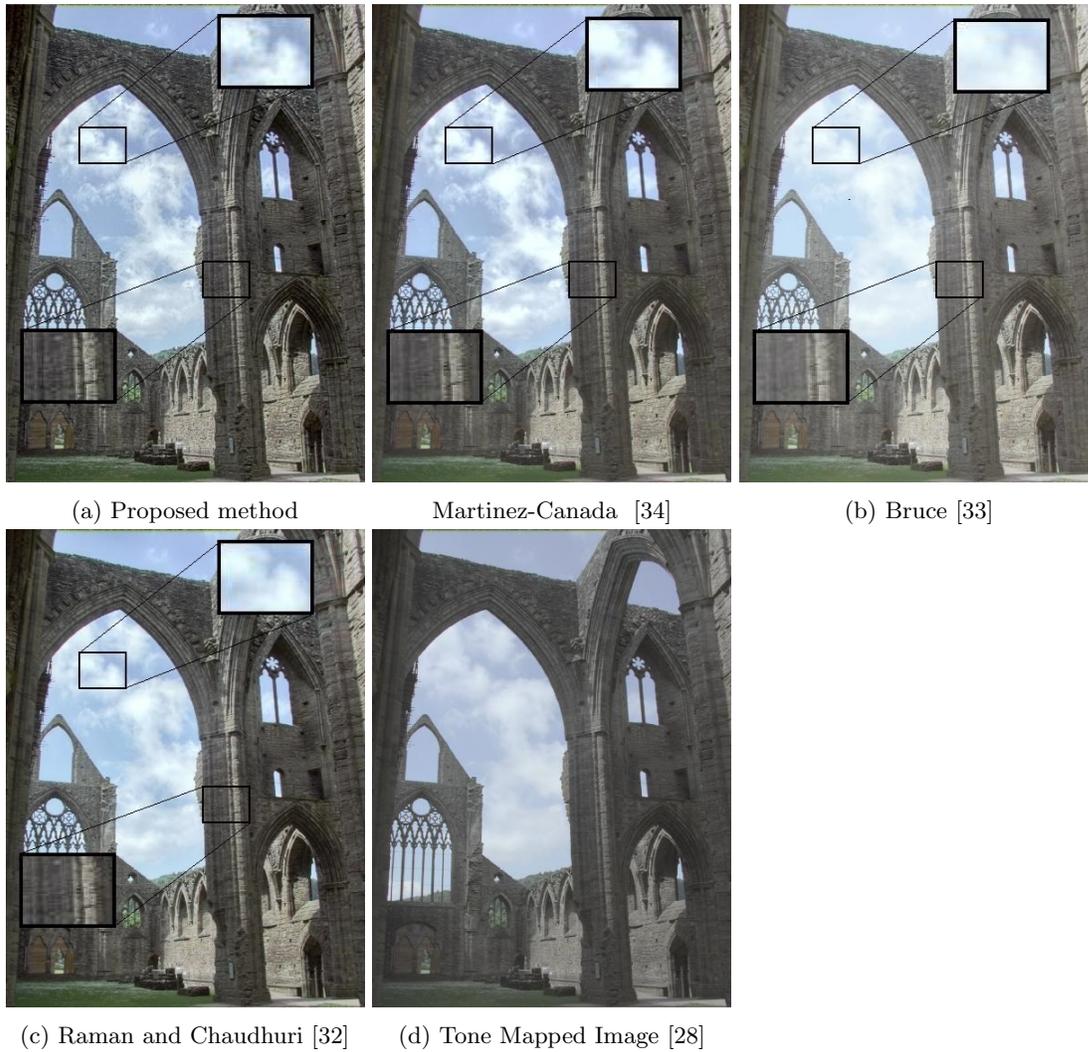


Figure 5.5: Image set 15 generated using various EF methods

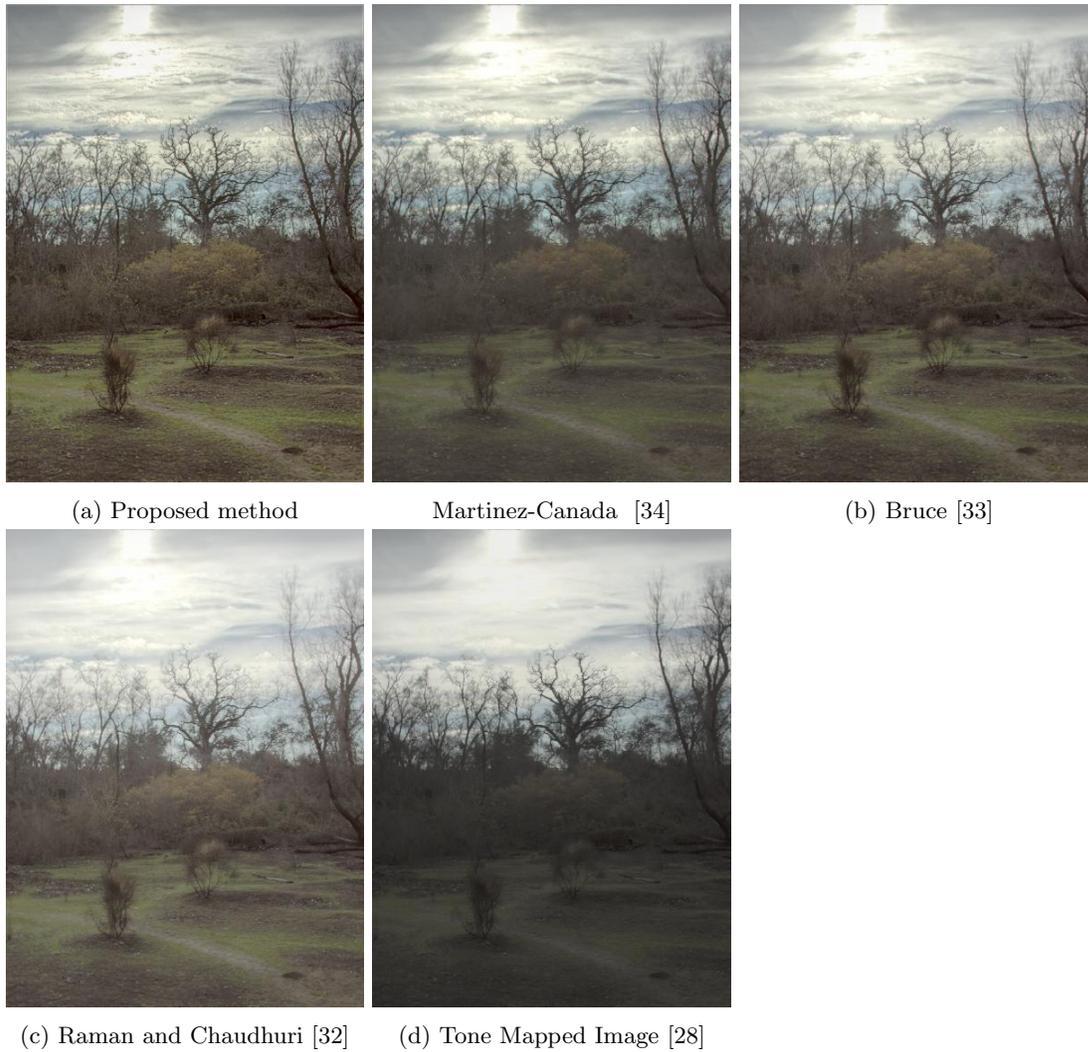


Figure 5.6: Image set 11 generated using various EF methods

Table 5.1: Score for image set 15

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.77	0.95	90.41
Martinez-Canada	0.46	0.99	91.28
Bruce	0.55	0.98	90.9890
Raman and Chaudhury	0.30	0.93	91.15
Tone mapped Image		0.97	88.8

Table 5.2: Score for image set 11

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.96	0.98	90.42
Martinez-Canada	0.40	0.94	91.37
Bruce	0.38	0.95	90.98
Raman and Chaudhury	0.33	0.90	90.87
Tone mapped image		0.90	88.64

sequence used to generate Figure ?? contained equal number of both over and under exposed images. Furthermore, MRR score in Table 5.1 indicate that the image generated using the proposed method is highly preferable.

The key difference between all four of the images in Figure ?? is the amount of details that can be perceived and how natural the resulting image appears. The proposed algorithm aims to increase the details in both bright and dark regions, whilst retaining the naturalness of the image which other methods fail to capture.

Figure ??(c) and (d) shows the result of using Bruce’s and Raman and Chaudhuris algorithm respectively [33][32]. The bright regions (sky) in these images are overly bright which have led to loss of details. The cloud region in Figure ??(c) and (d) shows where this loss of detail has occurred and the same region in Figure ??(a) clearly show that the proposed algorithm has more perceived details. Because Figure ??(c) and (d) are bright, the regions in the dark area are visible but they appear ”washed out”. Figure ??(b) is based on a method that is most similar to proposed method. The cloud regions in fig ??(b) appear detailed however dark regions (e.g. enclosed by the lower rectangle), show that these areas shows a loss of details and appear ”washed out”. The same regions in Figure ??(a) show more visible structural details.

The TMQI score shown in Table 5.1 indicate that the proposed method performs poorly compared to method of Bruce and Martinez-Canada. However, the user study indicates otherwise. One possible

reason for poor TMQI score is that the HDR image that was used for comparison might be a bright image therefore TMQI results preferred the brighter image. As discussed earlier HDR images cannot be displayed directly and require tone mapping for display. The bright tone mapped image in Figure ??(e) indicate that the HDR image was indeed bright and therefore, the tone mapped image also gained a higher TMQI score even though the details in the cloud region is not visible. The HDR-VDP scores Table 5.1 also ranked other methods higher then proposed method and that is another indication that the HDR image used for the reference might be the cause of this discrepancy.

Figure ?? is another example of images generated using Exposure Fusion techniques (same techniques used to generate Figure ??). Exposure sequence used to generate Figure ?? contains equal number of over and under exposed images. Furthermore, the MRR score in Table 5.2 indicates that the proposed method is highly preferable over other method.

Figure ?? (a) clearly shows that the image generated using the proposed method appears bright and detailed. Figure ?? (d) also appear bright however the details in the sky region appears "washed out". The grass and the tree region in Figure ?? (b) and Figure ?? (c) appear "washed out" as well (more in Figure ?? (b)) making the imaged appears dull. Figure ?? a) also appears more colourful, making it highly preferable.

TMQI score in Table 5.2 indicate that the performance of the proposed method is far greater than other methods. However, TMQI score does not correlate with MRR score of remaining methods. The TMQI appears to be favoring brighter image over darker one. The tone mapped image appears dark and therefore the TMQI score was also at the bottom.

Figure ?? and Figure ?? are an examples of fusion results from a sequence containing mostly under exposed images. The MRR scores in Table 5.3 and 5.4 indicates that the proposed method generated the most favorable image.

The proposed algorithm aims to bring the relevant edge pixels to the mid display range (0.5) and therefore they are not attenuated by the fusion step. In an ideal case where equal number of over and under exposed images are used for the fusion, the fusion result will come somewhere between the two extremes. However, when the images used for fusion do not contain equal number of over and under exposed shots, the resulting image will be dominated by a dark or bright pixels and thus, the image will lose its natural feel (Figure ??(b)-(d)). The proposed method aims to mitigate this effect by pre-processing the images before the fusion. Figure ?? (a) clearly shows that the overall tone of the image

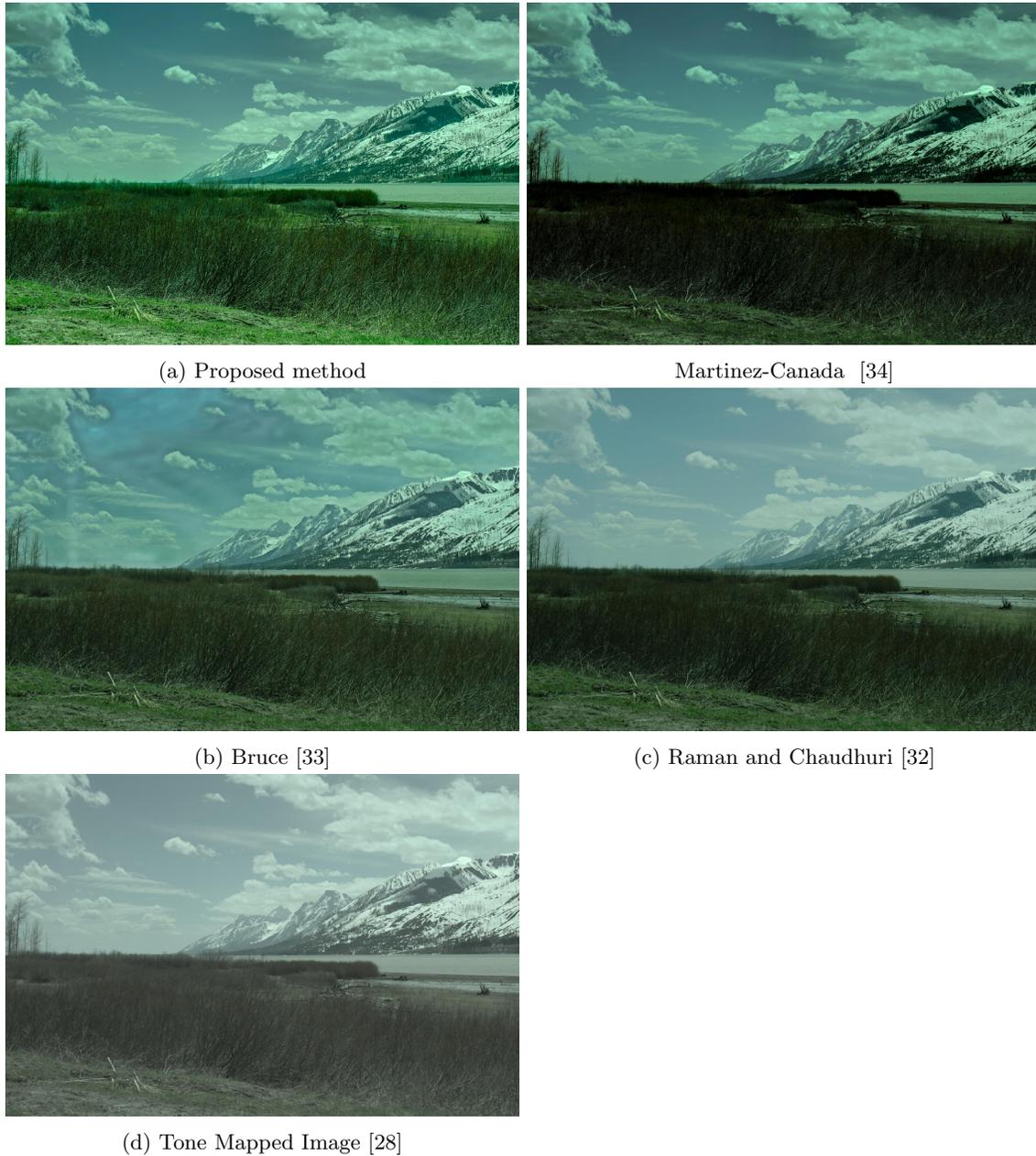


Figure 5.7: Image set 23 generated using various EF methods



Figure 5.8: Image set 35 generated using various EF methods

Table 5.3: Score image set 23

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.72	0.95	62.53
Martinez-Canada	0.41	0.85	66.09
Bruce	0.44	0.88	68.00
Raman and Chaudhury	0.52	0.86	70.24
Tone mapped image		0.84	52.53

Table 5.4: score fpr image set 35

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.83	0.80	67.33
Martinez-Canada	0.55	0.80	69.15
Bruce	0.39	0.80	68.83
Raman and Chaudhury	0.30	0.79	68.45
Tone mapped image		0.88	50.29

is bright. The grass region of the image (bottom left) clearly indicates the location and the angle of the sun. On other hand, Figure ?? (b)-(d) gives mixed feeling of the overall tone of the image. The region with tall grass (center) for example, appears too dark in Figure ?? (b)-(d). TMQI results in 5.3 gives higher ranking to proposed method and this is an indication that TMQI prefers bright image more (proposed method yielded bright image).

Figure ?? is an example of a scene where the image generated by tone mapping and exposure fusion differs greatly. The TMQI score in Table 5.4 highly favors the tone mapped image while the MRR score favors the image generated using proposed method. High TMQI score of tone mapped image indicate that the HDR image is also overly bright.

As discussed earlier, the proposed method aims to bring the edge pixels near the mean of the Gaussian (0.5) and thus, images littered with edges will be greatly improved. Scene in Figure ?? does not contain any smooth regions and therefore, the proposed method is expected to perform better for this particular scene. Figure ?? (b)-(d) appears overly dark due to fusion of mainly under exposed images. However, the ?? a) is a bright, colourful image that is not only favored by majority of people but it also appear highly natural.

Figure 5.2 shows that the proposed method gives superior result for majority of the images however there are few cases where the proposed method is not favored. For example, the proposed method is highly unpopular for the image set 3 and 4.



Figure 5.9: Image set 3 generated using various EF methods



Figure 5.10: Image set 4 generated using various EF methods



(a) Proposed method



Martinez-Canada [34]



(b) Bruce [33]



(c) Raman and Chaudhuri [32]



(d) Tone Mapped Image [28]

Figure 5.11: Image set 30 generated using various EF methods



Figure 5.12: Image set 32 generated using various EF methods

Table 5.5: Score for image set 3

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.49	0.90	91.11
Martinez-Canada	0.59	0.94	91.36
Bruce	0.61	0.95	91.21
Raman and Chaudhury	0.39	0.95	91.16
Tone mapped image		0.92	89.26

Table 5.6: Score for image set 4

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.48	0.93	80.00
Martinez-Canada	0.47	0.94	80.39
Bruce	0.61	0.92	80.10
Raman and Chaudhury	0.52	0.88	79.82
Tone mapped image		0.93	74.65

Table 5.7: Score for fig 30

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.57	0.72	67.17
Martinez-Canada	0.84	0.72	67.56
Bruce	0.31	0.72	72.29
Raman and Chaudhury	0.34	0.72	72.10
Tone mapped image		0.81	51.98

Table 5.8: Score for image set 32

Method	Survey Result (MRR)	TMQI	HDR-VDP
Proposed Method	0.56	0.75	67.03
Martinez-Canada	0.72	0.72	66.92
Bruce	0.38	0.71	73.53
Raman and Chaudhury	0.41	0.69	73.04
Tone mapped image		0.83	50.22

Proposed method aims to bring the edge pixels to the middle of the display range in order to increase the contrast. The main reason an image might not be favored (image generated using proposed method) might be because of the existence of large smooth areas in a scene as well as relatively small edge rich area. The propose method will not bring any change to the smooth regions however, the small edge rich regions gets darker if it's bright and brighter if it's dark. This will reduce the overall contrast in the image. For example, a scene with a bright edge rich area on a large dark background (smooth)

will generate a dull appear image. In the case described in previous sentence, the proposed algorithm will darken the edge rich area (since its bright) and instead of increasing the contrast, the contrast will be decreased due the presence of large smooth dark background and ultimately the image with poor contrast will not be favored.

Figure ?? shows a scenario which contains large smooth background with small bright edge rich area. The images in Figure ?? were generated from exposure sequence containing equal number of over and under exposed images. Table 5.5 shows that MRR score for the proposed method is much lower than the other methods. This is due to the presence of large dark background and small edge rich area (face). The proposed method darkens the bright edge area thereby reducing the overall contrast (making the image undesirable). The tone mapped image show a bright Figure on a dark back ground, which indicates that the HDR image also contain a bright Figure on a dark background, giving TMQI score that correlate the MRR score.

Figure ?? is another example where the image generated by the proposed method was undesired. The scene in Figure ?? was generated from an exposure sequence containing equal number of over and under exposed images. The scene contains a bright background with a dark foreground which the proposed method might not enhance. Figure ?? (a) shows that the proposed method generated an image with a bright foreground (room) with a bright background. Because the dark edge rich area was enhanced while the bright smooth background was left unaltered the contrast was reduced, making it undesirable. For the scene depicted in Figure ??, the TMQI score does not correlate with the survey results.

Figure ?? and ?? are an example of images generated by exposure sequence containing uneven number of over and under exposed images and the proposed method fails to generate a pleasing output. The bright stone in middle of the pond in Figure ?? is an edge area that the proposed method will attenuate thus, causing loss of contrast and making the image undesirable. On other hand, Figure ?? shows a smooth bright area (the sky) behind a dark foreground (forest). The "washed out" effect of ?? b) is desirable in this situation.

The HDR image of scene depicted in ?? and ?? appears to be a bright image as indicated by a bright tone mapped image.

5.8 Summary

This chapter presents the results of a user survey, which was specifically designed to measure the effectiveness of the proposed method with other methods of similar nature. Images used for the evaluation were generated using two distinct types of exposure sequences. The first set of exposure sequences contained both over and under exposed images, representing an ideal situation. While the second set of exposure sequences did not contain fixed number of exposure shots, representing a practical scenario. The survey show that the proposed method's performance was consistently superior to other methods in both, practical and ideal, situations. Where as the performance of other competing methods showed different level of performance for the two different types of exposure sequences.

Chapter 6

Conclusion and Future Work

HDR imaging is an emerging technology whose aim is to capture the vast range of natural light intensity that a conventional camera cannot capture. This is achieved by utilizing the existing hardware or using new cameras specialized for HDR image capture. Cameras that are able to capture HDR image directly are still not commercially available due to cost and the experimental technology involved, however, HDR image generation using existing cameras is widely popular.

Existing camera can capture a HDR image by capturing sequence of photographs each captured at a different exposure level. The HDR image is created by fusing all the multi-exposure images. HDR image generation using this method has several other processes involved. For instance, image alignment is applied to the exposure sequence and then camera response curve is estimated from the aligned images. De-ghosting is then applied before the final fusion process. De-ghosting is a process of removing artifacts caused by an object motion between the exposure shots.

The generated HDR image cannot be displayed on a conventional monitors and therefore the dynamic range compression (tone mapping) is performed before displaying any HDR content. The tone mapping process is a non-linear quantization process that quantize the image into displayable range while keeping the human visual system and perception into consideration.

There also exist algorithms to directly fuse the image into displayable image know as Exposure Fusion (EF). EF algorithms do not require the intermediate HDR image generation stage and yield more natural appearing image. However EF techniques are not robust against varying nature of exposures used for the fusion.

The proposed method introduces a novel EF technique that is robust against the varying nature of exposure shots. The proposed method aims to increase the contrast by adaptively darkening the bright regions and vice versa. The proposed method was assessed in form of a user survey which proved the effectiveness of the proposed method. In support of these (subjective) evaluations, a number of metrics used to assess HDR were also considered, however it was found that these metrics (whilst informative), do not reflect the perception of quality of users as evidenced in the user survey. This implies that such metrics may not be adequate for use in evaluating HDR image quality. The need to explore more adequate metrics is thus warranted.

6.1 Future Work

Future work will be focused on 2 major issues regarding Exposure Fusion: 1) evaluation metric and 2) better pre-processing algorithm. The evaluation metrics introduced so far are only suitable for tone mapped images (as discussed in this chapter) and therefore an evaluation metric capable of assessing images generated by Exposure Fusion techniques needs to be researched.

Table 5.7 and 5.8 shows that the MRR scores for the scenes is not the highest. As discussed earlier, the main reason for low MRR score is the presence of dark smooth region behind a small bright edge rich region. The proposed method aims to pre-process the images for best results for all condition, however, there are some cases where the proposed method fails. Future research will be aimed at improving the pre-processing algorithm introduced in this chapter.

Appendix 1

Gaussian Pyramid

The Gaussian pyramid is an algorithm to reduce image resolution. Gaussian pyramid iteratively decimate an image (usually by factor of 2) by first filtering the image using a Gaussian mask and then down sampling. the aim of the Gaussian pyramid is to generate a sequence of low pass images, each with stopband reduced by factor of 0.5. The Gaussian mask can be generated using equation 1.1.

$$f(x, y) = Ae^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} \quad (1.1)$$

Where:

A = Scale factor

(x, y) = x and y coordinate

(x_0, y_0) = x and y coordinate of center of the Gaussian

(σ_x, σ_y) = Adjustable parameter to set the width of the Gaussian

The σ in equation 1.1 should be set to have $\pi/2$ stopband. The Gaussian pyramid effectively "throw" away the high frequency component of the image in each iteration. Therefore, the original image cannot be recovered from the Gaussian pyramid. Algorithm 3 summarizes the steps needed to compute the Gaussian pyramid.

Algorithm 3 Gaussian Pyramid

```
N = NumberOfLevels);  
I = InputImage);  
for L = 1 to N do  
    If = Gaussian_filter(I)  
    Id = DownSample(If)  
    pyr(L) = Id  
end forreturn pyr
```

Appendix 2

Laplacian Pyramid

Laplacian pyramid also is an algorithm to reduce image resolution. However, the laplacian pyramid decomposes the image into series of band pass images. The Laplacian pyramid is calculated similarly to Gaussian pyramid with the added step of interpolating the decimated image and then subtracting that image from the original. This step will preserve the details that a Gaussian pyramid "throws" away. The Laplacian pyramid will yield a single base level (low pass image) and series of detail levels (band pass images). Algorithm 4 summarizes the steps needed to compute the Laplacian pyramid.

Algorithm 4 Laplacian Pyramid

```
N = NumberOfLevels);  
I = InputImage);  
for L = 1 to N do  
  If = Gaussian_filter(I);  
  Id = DownSample(If);  
  Iu = UpSample(If);  
  Ifd = Gaussian_filter(Iu);  
  Idetail = I - Ifd;  
  I = Id;  
  pyr(L - 1) = Idetail;  
  pyrL = I;  
end for return pyr
```

2.1 inverse Laplacian

The Laplacian pyramid does not "throw" away any information and therefore, the original image can be re-constructed.

Inverse of the laplacian pyramid, also known as collapsing the pyramid, is obtained by iteratively interpolating the base layer and adding the adjacent detail layer. Algorithm 5 summarizes the steps needed to compute the inverse Laplacian pyramid.

Algorithm 5 Inverse Laplacian Pyramid

```
N = NumberOfLevels);  
pyr = Laplacianpyramid);  
for L = 1 to N do  
  Iu = UpSample(pyr(N))  
  If = Gaussian_filter(Iu);  
  pyr(N - 1) = pyr(N - 1) + If;  
  N = N - 1  
end for return pyr(1)
```

References

- [1] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.
- [2] Leon Kanelovitch, Yaakov Itzhak, Arie Rundstein, Miri Sklair, and Hedva Spitzer. Biologically derived companding algorithm for high dynamic range mammography images. *Biomedical Engineering, IEEE Transactions on*, 60(8):2253–2261, 2013.
- [3] URL <http://scriptvegas.com/blog/posts/image-based-lighting-without-skyshop>. Accessed: 2015-08-30.
- [4] Jonas Unger, Stefan Gustavson, Joel Kronander, Per Larsson, Gerhard Bonnet, and Gunnar Kaiser. Next generation image based lighting using hdr video. In *ACM SIGGRAPH 2011 Talks*, page 60. ACM, 2011.
- [5] Alessandro Rossi, Nicola Acito, Marco Diani, and Giovanni Corsini. Dynamic range reduction and contrast adjustment of infrared images in surveillance scenarios. *Optical Engineering*, 52(10):102002–102002, 2013.
- [6] V Singh and Krishnan Nallaperumal. Application of hdr imaging techniques for sonar images. In *Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on*, pages 1–5. IEEE, 2012.
- [7] Fei Kou, Weihai Chen, Jianhua Wang, and Zhiwen Zhao. A lane boundary detection method based on high dynamic range image. In *Industrial Informatics (INDIN), 2012 10th IEEE International Conference on*, pages 21–25. IEEE, 2012.

- [8] Sukhan Lee, Khanh Trong Mai, and WoongJi Jeong. Virtual high dynamic range imaging for robust recognition. In *Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication, ICUIMC '12*, pages 22:1–22:6, New York, NY, USA, 2012. ACM.
- [9] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer Graphics Forum*, volume 28, pages 161–171. Wiley Online Library, 2009.
- [10] Hojatollah Yeganeh and Zhou Wang. Objective quality assessment of tone-mapped images. *Image Processing, IEEE Transactions on*, 22(2):657–667, 2013.
- [11] Apiparn Borisuit, Mirjam Münch, Laurent Deschamps, Jérôme Kämpf, and Jean-Louis Scartezini. A new device for dynamic luminance mapping and glare risk assessment in buildings. In *SPIE Optical Engineering Applications*, volume 8485, pages 84850–84850, 2012. doi: 10.1117/12.929845.
- [12] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of graphics tools*, 8(2):17–30, 2003.
- [13] Fabrizio Pece and Jan Kautz. Bitmap movement detection: Hdr for dynamic scenes. In *Visual Media Production (CVMP), 2010 Conference on*, pages 1–8. IEEE, 2010.
- [14] S Mann and R Picard. *Being "undigital" with digital cameras*. MIT Media Lab Perceptual, 1994.
- [15] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 Classes, SIGGRAPH '08*, pages 31:1–31:10, New York, NY, USA, 2008. ACM.
- [16] Tomoo Mitsunaga and Shree K Nayar. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999.
- [17] Takao Jinnō and Masahiro Okuda. Multiple exposure fusion for high dynamic range image acquisition. *Image Processing, IEEE Transactions on*, 21(1):358–365, 2012.
- [18] Jianbo Xu, Youjun Huang, and Jianli Wang. Multi-exposure images of wavelet transform fusion. In *Fifth International Conference on Digital Image Processing*, volume 8878, pages 88780F–88780F. International Society for Optics and Photonics, 2013.

- [19] Qi Wang, Zongxi Song, and Wei Gao. A multi-exposure image fusion method based on wavelet packet transform. In *ISPDI 2013-Fifth International Symposium on Photoelectronic Detection and Imaging*, volume 8907, pages 89070X–89070X. International Society for Optics and Photonics, 2013.
- [20] Jack Tumblin and Holly Rushmeier. Tone reproduction for realistic images. *Computer Graphics and Applications, IEEE*, 13(6):42–48, 1993.
- [21] Jinhua Wang, Hongzhe Liu, and Ning He. Exposure fusion based on sparse representation using approximate k-svd. *Neurocomputing*, 135:145–154, 2014.
- [22] M. Aharon, M. Elad, and A. Bruckstein. k -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov 2006. ISSN 1053-587X. doi: 10.1109/TSP.2006.881199.
- [23] Yagyensh Chandra Pati, Ramin Rezaiifar, and PS Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, volume 1, pages 40–44. IEEE, 1993.
- [24] Abhilash Srikantha and Désiré Sidibé. Ghost detection and removal for high dynamic range images: Recent advances. *Signal Processing: Image Communication*, 27(6):650–662, 2012.
- [25] Edwin H Land and John McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971.
- [26] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *Visualization and Computer Graphics, IEEE Transactions on*, 3(4):291–306, 1997.
- [27] Christophe Schlick. Quantization techniques for visualization of high dynamic range pictures. In Georgios Sakas, Stefan Mller, and Peter Shirley, editors, *Photorealistic Rendering Techniques*, Focus on Computer Graphics, pages 7–20. Springer Berlin Heidelberg, 1995.
- [28] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 267–276. ACM, 2002.

- [29] Ken Chiu, Michael Herf, Peter Shirley, S Swamy, Changyaw Wang, Kurt Zimmerman, et al. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface 93*, pages 245–245. Canadian Information Processing Society, 1993.
- [30] Edoardo Provenzi, Massimo Fierro, Alessandro Rizzi, Luca De Carli, Davide Gadia, and Daniele Marini. Random spray retinex: A new retinex implementation to investigate the local properties of the model. *Image Processing, IEEE Transactions on*, 16(1):162–171, 2007.
- [31] Edwin H Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10):3078–3080, 1986.
- [32] Shanmuganathan Raman and Subhasis Chaudhuri. Bilateral filter based compositing for variable exposure photography. In *Proceedings of Eurographics*, 2009.
- [33] Neil DB Bruce. Expoblend: Information preserving exposure blending based on normalized log-domain entropy. *Computers & Graphics*, 39:12–23, 2014.
- [34] Pablo Martinez-Caada and Marius Pedersen. Exposure fusion algorithm based on perceptual contrast and dynamic adjustment of well-exposedness. In Abderrahim Elmoataz, Olivier Lezoray, Fathallah Nouboud, and Driss Mammass, editors, *Image and Signal Processing*, volume 8509 of *Lecture Notes in Computer Science*, pages 183–192. Springer International Publishing, 2014. ISBN 978-3-319-07997-4.
- [35] Alessandro Rizzi, Gabriele Simone, and Roberto Cordone. A modified algorithm for perceived contrast measure in digital images. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2008, pages 249–252. Society for Imaging Science and Technology, 2008.
- [36] Sidhharthkumar Patel, Dimitri Androustos, and Matthew Kyan. Adaptive exposure fusion for high dynamic range imaging. In *International Conference on Image Processing*, 2014.
- [37] Hojatollah Yeganeh and Zhou Wang. URL <https://ece.uwaterloo.ca/~z70wang/research/tmqi>. Accessed: 2015-08-30.
- [38] Funt et al. hdr dataset. URL http://www.cs.sfu.ca/~colour/data/funt_hdr/. Accessed: 2015-08-30.

-
- [39] Rafat Mantiuk, Kil Joong Kim, Allan G Rempel, and Wolfgang Heidrich. Hdr-vdp-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on Graphics (TOG)*, volume 30, page 40. ACM, 2011.
- [40] Scott J Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*, volume 1666, pages 2–15. International Society for Optics and Photonics, 1992.