# OBJECTIVE IMAGE QUALITY ASSESSMENT BASED TONE MAPPING OPTIMIZATION

by

Xihe Gao

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
July 2017

*To my family for their love and patience.*

# Table of Contents

## List of Tables

## List of Figures

## Abstract

High dynamic range (HDR) images provide the capacity to represent the luminance in real scenes with much higher precision than standard image formats. With advances in hardware and computer graphics technologies, HDR images are rapidly becoming more commonplace. To visualize HDR images on contemporary display devices, the dynamic range needs to be adapted to the much smaller range of the devices. This is accomplished through tone mapping, with the goal of reproducing the visual appearance of HDR scenes. Tone mapping has attracted much attention and several dozens of tone mapping operators have been proposed.

Nevertheless, it remains challenging to objectively evaluate the quality of tone mapped images and optimize tone mapping operators with automated algorithms. Using virtual photographs to bridge the gap of dynamic ranges for feature analysis, we propose two feature-based quality metrics for tone mapped images, which measure the distortion of important image features that affect the perceived quality. We present an image quality metric called visual saliency distortion predictor (VSDP) that measures the distortion in visual saliency for quality assessment. Additionally, by incorporating multiple feature-based measures to predict the quality of tone mapped images, we introduce another quality metric: perceptual distortion predictor (PDP). Subjective and numerical experiments indicate that the proposed feature-based quality metrics can yield more reliable prediction than the alternative approaches.

Once suitable quality metrics are defined, there emerges an opportunity to automate the tuning of existing tone mapping operators. By minimizing the distortion in visual saliency predicted by the quality metric VSDP, we developed an automatic parameter tuning algorithm for tone mapping operators. Moreover, based on the quality prediction of PDP, we propose an automated blended tone mapping algorithm which blends images from multiple operators with varying weights to leverage the strengths of each of operators considered. Experiments with a broad range of HDR images and statistical analysis demonstrate the effectiveness of the tone mapping optimization algorithms.

## List of Abbreviations and Symbols Used

| | |
|---|---|
| **TMQI-II** | Tone Mapped Image Quality Index-II |
| **CSA** | Cumulative Step-size Adaption |
| **DoG** | Difference of Gaussians |
| **EA** | Evolutionary Algorithm |
| **ES** | Evolution Strategy |
| **FOA** | Focus of Attention |
| **HDR** | High Dynamic Range |
| **iCAM** | Image Color Appearance Model |
| **IQA** | Image Quality Assessment |
| **JND** | Just Noticeable Difference |
| **KRCC** | Kendall's Rank Correlation Coefficient |
| **LDR** | Low Dynamic Range |
| **MSE** | Mean Squared Error |
| **PDP** | Perceptual Distortion Predictor |
| **PSNR** | Peak Signal-to-noise Ratio |
| **ROI** | Region of Interest |

**SRCC** Spearman's Rank Correlation Coefficient
**SSIM** Structural Similarity

**TMO** Tone Mapping Operator
**TMQI** Tone Mapped Image Quality Index
**TVI** Threshold Versus Intensity

**VDP** Visible Difference Predictor
**VSDP** Visual Saliency Distortion Predictor

**WTA** Winner Take All

## Acknowledgements

It is my pleasure to take this opportunity to express my gratitude to all the kind people around me.

Above all, I would like to express my deep appreciation to my supervisors, Dr. Stephen Brooks, and Dr. Dirk V. Arnold. Your initial encouragement and continuous support were essential to the coming about of this thesis. I would like to thank you for giving me opportunities to grow as a researcher.

Also, I would like to thank my external reader, Dr. David Mould, for your commitment to my research. For my internal committee members, Dr. Qigang Gao, Dr. Kirstie Hawkey, thanks for your support, not only in the final thesis defense, but also in the aptitude exam, thesis proposal, and my entire PhD study stage. I want to thank all of you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

Dalhousie University offers a productive environment to explore new ideas. I am grateful to have the chance to study in the middle of a greatly supportive community and be surrounded by wonderful colleagues and friends. In particular I would like to thank my colleagues and friends, Yang Yang, Aziz Abdurexit, Xiaoting Hong, Hossein Salimian to name but a few. Thank you for your company and assistance.

Special thanks to my family. Words cannot express how grateful I am to my beloved wife, Kathy, for all of the sacrifices that you've made on my behalf. Thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience. Last but not least, I would like to thank my Mom and Dad for your constant support and endless love, through my PhD study as well as through my life in general. I dedicate this thesis to them.

# Chapter 1

# Introduction

The dynamic range of illumination in a real-world scene is on the order of $10,000$ to 1 from highlights to shadows, and even higher for the scenes including both an outdoor area illuminated by sunlight and an indoor area illuminated by interior light [86, 9]. The development of High Dynamic Range (HDR) images capture a greater dynamic range between the lightest and darkest areas of real-world scenes. Compared with standard dynamic range (SDR) or low dynamic range (LDR) images, HDR images allow more intensity levels which enables them to represent real world visual data in a more accurate way (We don't distinguish between luminance and intensity in the thesis). Usually, HDR images are generated by taking multiple-exposed photographs of the same scene and merging their data with developed algorithms [18]. Because of the limitation of display contrast, HDR images cannot be displayed on regular display devices such as LCDs and CRTs directly, and tone mapping is needed to fit their high dynamic range into the displayable range of conventional devices. The algorithms of tone mapping is called tone mapping operator (TMO) [86].

Figure 1.1 shows an example of HDR imaging. The left image is the photograph sequence with various exposures which are subsequently combined into an HDR image, and the right image is a tone mapped image of the generated HDR scene. As we can see, the tone mapped image can have plausible exposure and detail preservation in both bright and dark areas, such as the sky in the background and the woods in the foreground, while any single exposure in the photograph sequence cannot. A thorough review of HDR imaging can be found in the books by Reinhard et al. [86] and Banterle et al. [9]. In this chapter, we introduce the research problems that will be addressed in current context (Section 1.1), the research objectives (Section 1.2), and the major contributions (Section 1.3), and then we outline the structure of the thesis (Section 1.4).

Photograph sequence Tone mapped image

Figure 1.1: An example of HDR imaging on the HDR image "Sequoia Remains". Left: image sequence with various exposures that can be used to generate the HDR image. Right: tone mapped image from the HDR image. The images are from Mark Fairchild's HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

## 1.1 Research Problems

HDR images provide the capacity to represent the luminance in the real scenes ranging from bright sunlight to faint starlight with much higher precision than standard image formats allow. Typically, HDR images are categorized as scene-referred images whose luminance levels correspond to that of real scenes, while LDR images are display-referred with luminance variants matching the dynamic range of display devices. HDR images are becoming increasingly commonplace, and they have successfully been used for many applications such as digital photography, physical-based rendering, and virtual reality.

Tone mapping is one of the core problems in HDR imaging and corresponding applications, and it is the topic of interest in the thesis. According to Reinhard et al. [86], the ultimate goal of tone mapping is to reproduce the visual appearance of HDR scenes when displayed with low dynamic range. Although an extensive body

of research has focused on tone mapping, it remains an open question to measure and quantify visual appearance reproduction during tone mapping. Due to the huge difference in luminance levels, conventional image analysis algorithms and quality evaluation methods designed for LDR images cannot be applied to HDR images directly. It is fundamental and essential to establish a framework to bridge the gap of dynamic range between LDR and HDR images and build new approaches for objective quality evaluation of tone mapped images, which will provide deeper insight to understand and exert the strength of TMOs.

In this thesis, we mainly focus on three research problems: visual saliency analysis on HDR images, objective image quality assessment of tone mapped images, and tone mapping optimization.

### 1.1.1 Visual Saliency Analysis on HDR Images

The human visual system can obtain a rich stream of visual data ($10^8$-$10^9$ bits per second) from the real world [42]. Rather than processing the data in parallel, the mechanism in our brain prioritizes the important parts over the others to guide our gaze, which is known as selective attention [31]. Based on the mechanism of selective attention, computational visual attention systems have been designed to detect regions of interest in digital images [11]. Although the computational models can lead to satisfying results for LDR images, they do not usually perform well when applied to HDR images [13]. Because HDR images allow a much higher dynamic range, if the saliency detection models are applied to HDR images directly, the dynamic range will be scaled, which can lead to the loss of HDR content that in turn makes salient regions appear not salient or vice versa. Also, Narwaria et al. [74] have found that TMOs "can [...] modify human attention and fixation behavior significantly", thus rendering the approach of applying salience analysis techniques after tone mapping unreliable. In this thesis, we concern ourselves with visual saliency analysis on HDR images. As an important image feature, visual saliency can be used for various applications that involve HDR image understanding and tone mapping, such as visual saliency guided tone mapping, and image quality assessment of tone mapped images. Moreover, visual saliency analysis could be used as an example to explore the possible solutions for image feature analysis on HDR images.

### 1.1.2 Objective Image Quality Assessment

Because any process applied to images may cause information loss or quality reduction, image quality assessment plays an important role in many image processing problems, such as image acquisition, synthesis, enhancement, restoration, and reproduction [12]. Objective image quality assessment provides quantitative measures that can automatically predict the perceived image quality. Compared with subjective assessment, objective assessment is more economic, faster, more consistent, and applicable in optimization of image processing systems. Due to these advantages, it has been widely used in image processing applications [12]. With the development of HDR images and tone mapping algorithms comes a need for image quality evaluation of tone mapped images. Unfortunately, the typical image quality metrics assume that reference and test images have the same dynamic range [100], and they cannot be applied to quality evaluation of tone mapped images. More specifically, the quality metrics calculate visible distortion based on the difference of intensity or contrast values, and thus not applicable for image pairs with significant different dynamic ranges. In comparison with conventional methods, the quality evaluation of tone mapped images should focus on the reproduction of important image features relevant to image quality judgement, rather than the optical match between reference and test images [7].

### 1.1.3 Tone Mapping Optimization

During the last two decades, several dozens of TMOs have been proposed to reproduce the visual appearance of HDR images, ranging from the use of simple sigmoidal functions [88] to more complicated gradient domain operations [13]. Many of the TMOs depend on parameters that significantly impact the quality of the tone mapped images. Choosing an appropriate operator and setting its parameters for a particular HDR image often requires careful tuning, which could be tedious and time-consuming even for knowledgable users. Several user interfaces have been introduced to assist in this manipulation process. Lischinski et al. [57] present an interactive tool for users to indicate regions of interest with brush strokes and make local adjustments of visual parameters. In the work of Chisholm et al. [16], users can iteratively select the best image among a set of blended tone mapped images for parameter optimization. These

interfaces enable rapid and intuitive manual manipulation, but they still need user interaction to achieve satisfying performance. In this thesis, we apply the objective image quality assessment to automatic tone mapping optimization, and problems including parameter tuning and blended tone mapping are addressed.

## 1.2 Objectives

The main objectives are to develop objective image quality metrics for tone mapped images and apply the image quality metrics for automatic tone mapping optimization in terms of parameter tuning of TMOs and blended tone mapping. More generally, we would like to explore the solutions for image feature analysis and comparison across different dynamic ranges, which could serve as the basis for various applications in HDR imaging.

## 1.3 Contributions

The main contributions of the thesis are summarized as follows:

- We introduce a novel algorithm for visual saliency analysis of HDR images [33]. The algorithm decomposes HDR images into multi-exposed LDR images, which are referred to as virtual photographs, and then incorporates them for visual saliency analysis. We demonstrate that our method can produce more consistently reliable results than existing methods. The algorithm is general and can be easily tailed for other image feature analysis on HDR images (Chapter 3).

- Based on the assumption that regions of interest predicted by bottom-up visual attention models should be preserved during tone mapping, we propose a new feature-based quality metric called visual saliency distortion predictor (VSDP) [34, 35]. The quality metric measures the distortion in visual saliency for quality predictions of tone mapped images. We have evaluated the quality metric by applying it to a number of test images and found that it yields more accurate evaluation than prior work (Chapter 4).

- Derived from perceptual studies, we then present a more comprehensive feature-based quality metric called perceptual distortion predictor (PDP) [38, 32]. The

quality metric measures the distortions of tone mapped images in terms of brightness, visual saliency, and detail reproduction in light and dark areas, and assigns an overall distortion value to each image. Validation using a subject-rated image database indicates the proposed metric is more consist with subjective evaluation results than alternative approaches (Chapter 4).

- We employ the proposed quality metrics to tone mapping optimization, and develop an automatic parameter tuning algorithm that can optimize the parameters of arbitrary TMOs by minimizing visual saliency distortion [34, 35]. The minimization is accomplished by employing an evolutionary algorithm (EA). Experiments using several TMOs demonstrate the effectiveness of our parameter tuning algorithm. Statistical analyses are conducted to assess the improvement over default parameter settings and previous methods (Chapter 5).

- With the perceptual distortion predictor, we develop an automated blended tone mapping algorithm to leverage the strengths of different operators for any particular HDR image [36]. The blended tone mapping is solved as an optimization problem, where the operators' parameters and the weights are tuned with an EA to generate the optimal solution. With a variety of HDR images, we demonstrate its superiority over the conventional global and local TMOs (Chapter 5).

- Using parameter space of TMOs as solution space and quality metrics as objective, EA provides a universal solution for objective quality assessment based tone mapping optimization. We conduct a comparison between EA and an alternative gradient-based optimization under a common platform, which shows that EA results in significantly reduced computational effort [37] (Chapter 6).

## 1.4 Structure of the Thesis

The remainder of the thesis is organized as follows. Chapter 2 describes the research topics and the existing methods that serve as the background of the thesis which covers human visual attention, high dynamic range imaging, quality assessment of tone mapped images, and evolution strategies. Chapter 3 presents the visual photograph

based saliency detection method for HDR images. Chapter 4 focuses on objective image quality metrics for tone mapped images, and two feature-based quality metrics VSDP and PDP are introduced and validated against prior works. Chapter 5 presents the application of the proposed quality metrics for tone mapping optimization. Tone mapping optimization problems, including automated parameter tuning of TMOs, and blended tone mapping, are addressed. Chapter 6 describes the comparison between EA and gradient-based optimization for tone mapping optimization, and Chapter 7 concludes the thesis and discusses the directions for future research.

# Chapter 2

# Background

This chapter provides an overview of the various fields related to research topics found throughout the thesis. First of all, we describe the concept of human visual attention, computational attention models, and improved method for application to HDR images (Section 2.1). Then, we discuss high dynamic range imaging, including visual adaptation models, TMOs, and recent directions (Section 2.2). After that, we outline the objective image quality assessment of tone mapped images and the state-of-the-art methods (Section 2.3). Finally, we discuss evolution strategies in terms of the optimization algorithm and parameter control (Section 2.4).

## 2.1 Visual Attention Models

Visual attention can be defined as the process of selecting a subset of the all the available information for further processing. Since human needs to deal with a large amount of information at each moment and the amount can be too high to be completely processed in its entirety at once, the brain tends to allocate the processing resources to certain regions to obtain the most significant information, which is known as the mechanism of selective attention [31]. By understanding and simulating the selection mechanism, computational attention models analyze the regions of interests (ROIs) which can attract more visual attention for their distinctive features compared with others in the images. The computational models have a wide variety of applications in computer vision and image processing, such as object detection [97, 70], human-robot interaction [41, 73], image and video compression [43], and image resizing [6, 99].

Visual attention can be categorized into two types: bottom-up and top-down attention [19]. Bottom-up attention, also known as saliency-based attention [46], is driven by purely visual data. The regions with sufficiently discriminative features with respect to surrounding features can attract visual attention in a bottom-up

manner. On the other hand, top-down attention is driven by cognitive phenomena, such as knowledge, expectations, and the current task. For instance, it is more likely for car drivers to see gas stations in a street than other targets. Because data-driven stimuli can be easier to control than cognitive factors, bottom-up attention are more thoroughly investigated than top-down attention [31]. We focus on bottom-up visual attention in this thesis.

### 2.1.1 Related Concepts and Theory

This section introduces some concepts of visual attention and psychological theory which are related to computational attention models.

- **Visual Saliency:** Visual saliency (or visual salience) "is the distinct subjective perceptual quality which makes some items in the world stand out from their neighbors and immediately grab our attention" [44]. To address the inability of our brain to fully process all locations in parallel, visual attention is attracted to visually salient stimuli, and processes one region at one time. This raises a question: how to select the targets of attention? Visual saliency, which is a bottom-up and stimulus-driven perceptual quality in the early stage of visual processing, is capable of helping the brain to make reasonable and efficient selections. In visual attention models, computing saliency is the detection of the regions whose visual features such as intensity, orientation, and color are discriminative with respect to surrounding regions.

- **Saliency Map:** A saliency map is an explicit two-dimensional map that represents visual saliency of any location in the corresponding visual scenes. The concept of saliency map was first introduced by Koch and Ullman [50] for visual attention deployment. Once the saliency map is established, the most salient targets that attract visual attention can be calculated by a Winner-Take-All (WTA) network. A saliency map can be generated by analyzing the saliency map of each individual feature (called conspicuity maps) [31] in parallel and then integrating the maps together into a single map. Since saliency map is a computationally tractable representation of visual saliency, it has been widely used in computational visual attention models.

- **Feature Integration Theory:** Feature integration theory developed by Treisman and Gelade [92, 93] has been one of the most influential psychological theories of human visual attention. The theory claims that "different features are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage, which requires focused attention" [92]. A demonstration of feature integration theory is shown in Figure 2.1. In the pre-attentive stage, primitive features, such as color, and orientation, will be automatically and unconsciously analyzed resulting in topographical maps, which highlight the conspicuities from various features. In the focused attention stage, the conspicuity maps of features are collected in a master map of location, and serially scanning the master map focuses visual attention on the selected salient regions.

### 2.1.2 Computational Model

Because visual attention provides a selection mechanism to determine the most relevant targets within visual data, there is a significant interest in visual attention models in computer vision and robotics. In the last two decades, many computational attention models have been developed, which greatly improve existing vision systems [31]. Most models are built on feature integration theory introduced by Treisman and Gelade [92], whose core idea is to extract several types of features and combine their saliency to generate a saliency map.

A milestone is the computational model developed by Itti et al. [47]. Based on the behavior and the neuronal architecture of the primates' early visual system, Itti et al. [47] proposed a visual attention model to predict bottom-up attention in static color images. Their model uses a center-surround mechanism to generate the feature maps for color, intensity, and orientation, and then combine the maps into a unified saliency map. Later on, the model was revised by Itti and Koch [46], who introduced a within-feature competition scheme for feature combination. Itti et al.'s model [47, 46] is the best-known attention system, and it serves as the basis for many research into visual attention [31]. Most recent visual attention models share a similar structure but suggest various improvements. Some of the models focus on the selections of visual features, and other features such as skin color [56, 41], motion [62, 45, 94],

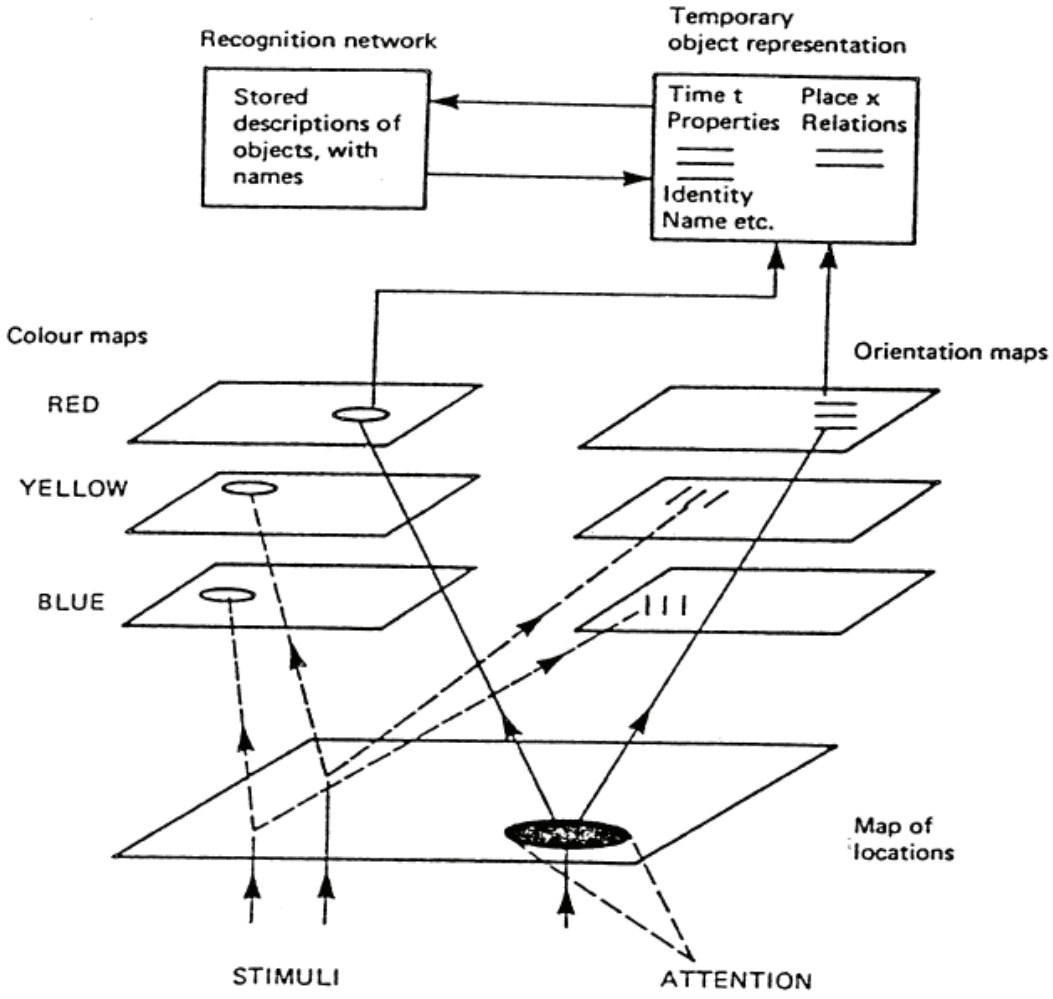



Figure 2.1: Demonstration of the Feature Integration Theory [93].

depth [14, 30], optical flow [96], flicker [45], corner [29, 78], and symmetry [41] are adopted for visual attention analysis. Also, some works address the representation of visual attention. More advanced approaches that integrate image segmentation on feature [98] or saliency maps [30] are developed to determine irregularly-shaped attention regions. A thorough review of current computational attention models can be found in the survey paper by Borji and Itti [11].

Most of computational visual attention models share a very similar structure originally adopted from psychological theories such as the feature integration theory [92, 93]. We use Itti et al.'s model [47, 46] as an example to introduce the general structure of visual attention models. Since the revised model [46] is employed in the

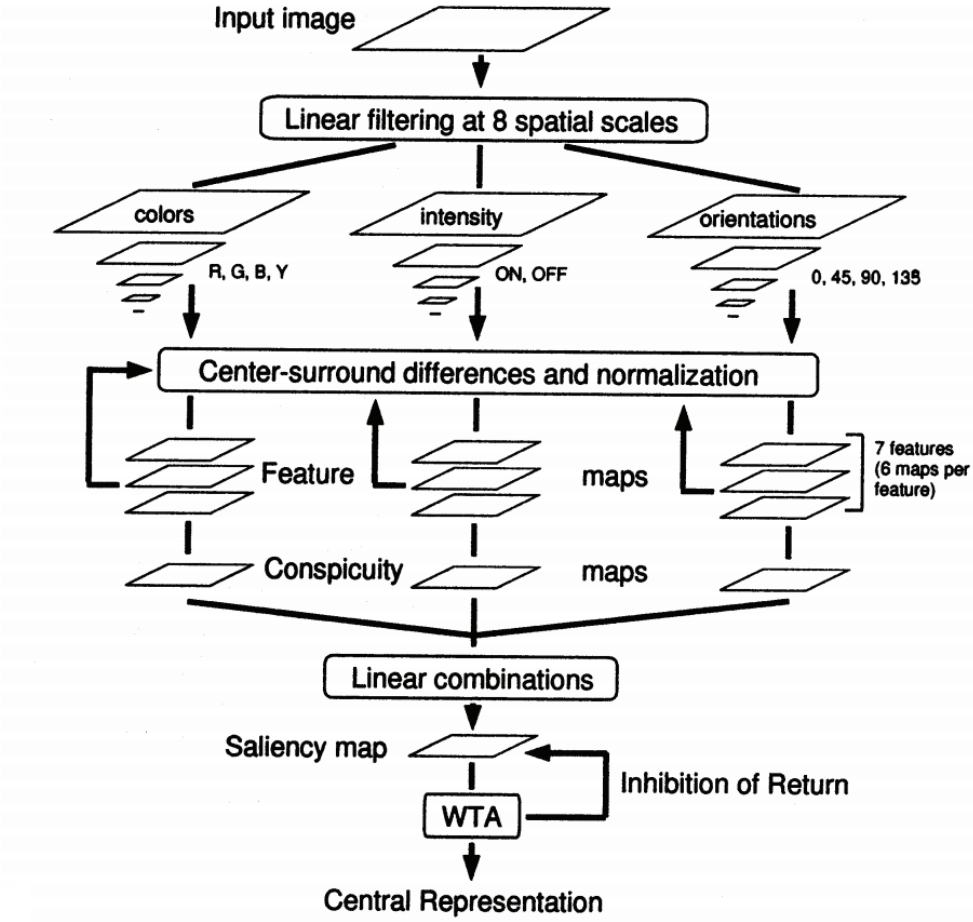



Figure 2.2: Architecture of Itti and Koch's model [46].

thesis, we will refer to the model as "Itti and Koch's model" in the following context. The architecture of their model is demonstrated in Figure 2.2.

First, Itti and Koch's model extracts low-level visual features including intensity, color, and orientation from the original images at several spatial scales. The different scales are generated using Gaussian pyramid to avoid explicitly applying large filters that could be slow. Each feature is computed with a center-surround operation. Inspired from retinal ganglion cells in the visual receptive fields [80], the operation compares the intensity values in center regions with those of surrounding regions. It is implemented by calculating the difference between fine and coarse scales. Usually, the finest scale of pyramids is ignored to reduce the influence of noise. Three commonly used features are used in Itti and Koch's model: intensity, color, and orientation.

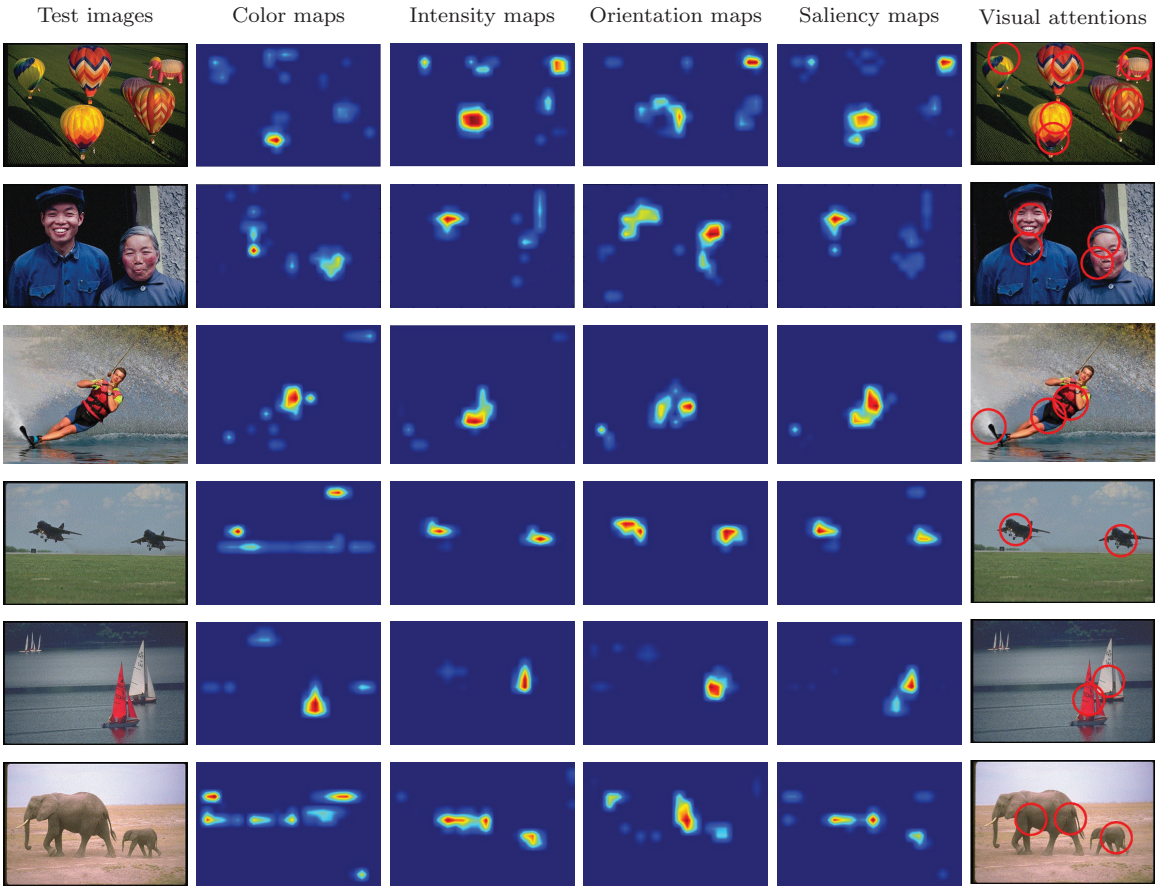



Figure 2.3: Results from Itti and Koch's model [46]. First column: the test images. Second column: the saliency maps calculated from the color feature. Third column: the saliency maps calculated from the intensity feature. Forth column: the saliency maps calculated from the orientation feature. Fifth column: the combined saliency maps. Sixth column: the predicted visual attention (red circles).

The intensity is calculated as on/off intensity contrast, color is computed on double-opponent red/green and blue/yellow, and orientation is measured on the local contrast of orientations of 0°, 45°, 90°, and 135°.

Afterwards, the cross-scale feature maps are combined into the conspicuity map for each feature. To solve the signal-to-noise problem during combination, Itti and Koch [46] introduced a within-feature spatial competition scheme. According to optical imaging [105] and human psychophysics [110], the interaction of long-range cortico-cortical connections are thought to result from a balance of excitation and inhibition between neighboring neurons. By simulating the structure of the interaction, the within-feature spatial competition scheme is realized by a two dimensional

difference-of-Gaussians (DoG) approach. Each feature map is iteratively convolved with the DoG filter, which yields local excitation at each visual location counteracted by broad inhibition from neighboring location.

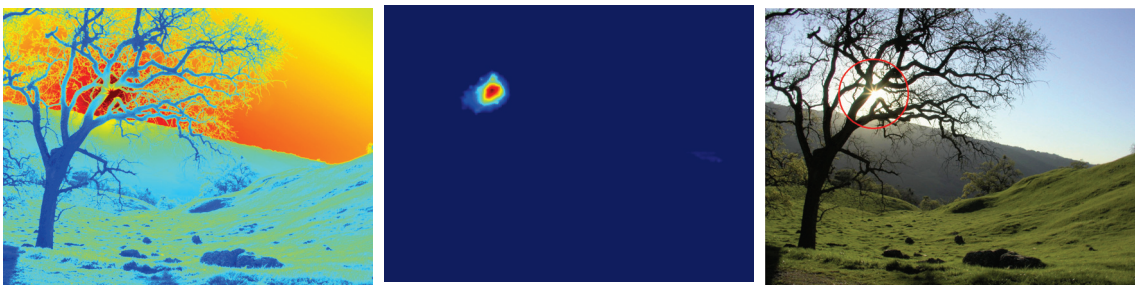
Finally, the conspicuity maps of intensity, color, and orientation are linearly summed into a single saliency map. The saliency map could already be regarded as the final output of computational models since it is capable of showing the saliency of each location in the input scene. Nevertheless, computational models usually compute the trajectory of visual attention as well to mimic human saccade (the movement of eyeball) which starts with the target with the highest saliency value, in which a winner-take-all (WTA) network can be used to select the image regions with local saliency maxima. A fixed size disk is applied to represent the focus of attention (FOA) since visual attention is usually on a region rather than a single point [46]. When a winner is found, the FOA shifts to the winning location. The shifting activates inhibition which can prevent the network from returning to its initial state: the inhibitory center is at the location of the winner; the winner and its neighbors are inhibited in the saliency map.

Figure 2.3 demonstrates several examples of saliency detection from Itti an Koch's model. As shown in the examples, the model is capable of recognizing the salient targets which are discriminative for their unique visual features relative to the surrounding fields, such as the balloons (First row), faces (Second row), water skiing person (Third row), airplanes (Forth row), boats (Fifth row), and elephants (Sixth row).

### 2.1.3 Improved Model for HDR images

Because of the huge difference in the dynamic range of luminance, conventional models using contrast-based feature analysis is not suitable for saliency detection on HDR images. When applying the models for HDR images directly, the significant reduction of contrast will cause inaccurate predictions, in which salient regions may appear not salient or vice versa. As shown in the example of Figure 2.4, only the sun can be detected while other eye-catching targets such as the tree and stone are missed.

Brémond et al. [13] proposed a visual attention model for HDR images. Based on the observation that saliency can be better preserved for color features than the



HDR scene | Saliency map | Focus of attention

Figure 2.4: Applying Itti and Koch's model on the HDR image "Sunol8". Left: false color map of the HDR image. Middle: saliency map generated from Itti and Koch's model [46]. Right: focus of attention (red circle) predicted from the saliency map. HDR image from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2006 Elsevier Inc.

other features when applying Itti and Koch's model [46] directly for HDR images, they hypothesize that the better preservation of saliency for color features is due to the normalization of color features. Therefore, they modify Itti and Koch's model by adding normalization to other visual features as well. More specifically, the cross-scale differences in terms of intensity and orientation are normalized over the intensity, for which the authors suggest that normalization may be seen as a gain modulation, which is the physiological mechanism of visual adaptation.

In eye tracking experiments with human subjects, they find that their approach provides more accurate saliency maps than that of Itti and Koch [46] when the latter is applied either directly to the HDR image or after dynamic range compression with tone mapped images from six TMOs. Before our approach was proposed [33], Brémond et al.'s model was the only method for visual attention analysis on HDR images.

## 2.2 HDR Tone Mapping

The problem of HDR tone mapping is to reproduce real-world scenes on existing display devices with lower dynamic ranges. It was first recognized by early artists to faithfully depict natural scenes on canvas. Because the light intensity levels in the environment may be completely beyond the levels that the pigments can provide, artists make use of some techniques such as drawing all objects with middle range

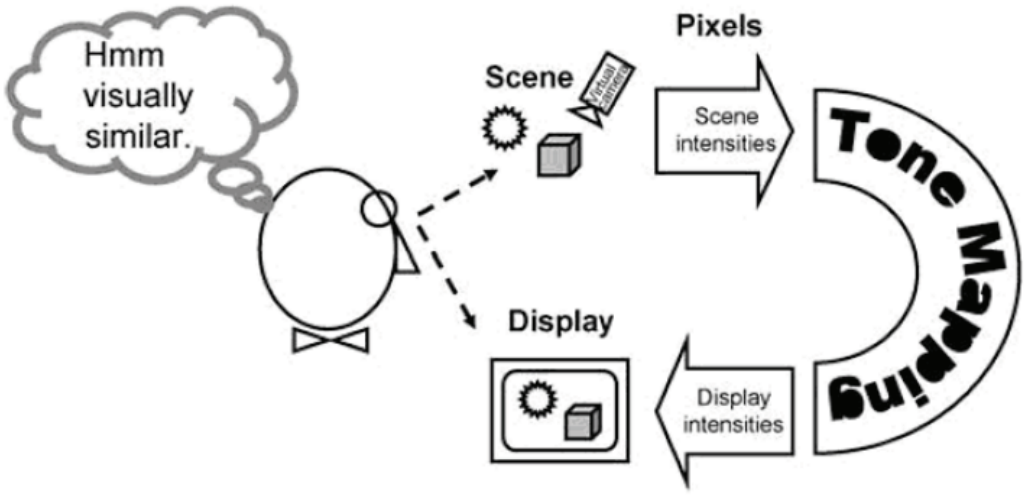



Figure 2.5: Pictorial outline of HDR tone mapping. The image is from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2006 Elsevier Inc.

colors [60] to overcome the limited dynamic range. Today, the focus of the problem has shifted to rendering HDR images on conventional display devices or media with limited luminance levels. HDR tone mapping addresses the significant contrast reduction from the scene intensities to the display intensities with the goal of achieving a visual match between the observed scene and the tone mapped images on the display (see the illustrated outline in Figure 2.5). Since Tumblin and Rushmeier [95] formally introduced the "display range" problem and used human visual models to solve the problem, HDR tone mapping has become an active research area in the community of computer graphics and many TMOs have been proposed.

### 2.2.1 Visual Adaptation Models

The human visual system addresses a similar problem to HDR tone mapping. The signal-to-noise of in individual channel in the visual pathway (from retina to brain) is less than 2 orders of magnitude [20]. In spite of the limited dynamic range, the human visual system enable us to perceive the detailed contrast under a wide range of illumination. Therefore, it would be informative to understand the mechanisms of the human visual system when solving the problem of tone mapping. Two of the most relevant visual adaptation models, including threshold versus intensity function, and photoreceptor responses relation, are briefly discussed. These visual adaptation

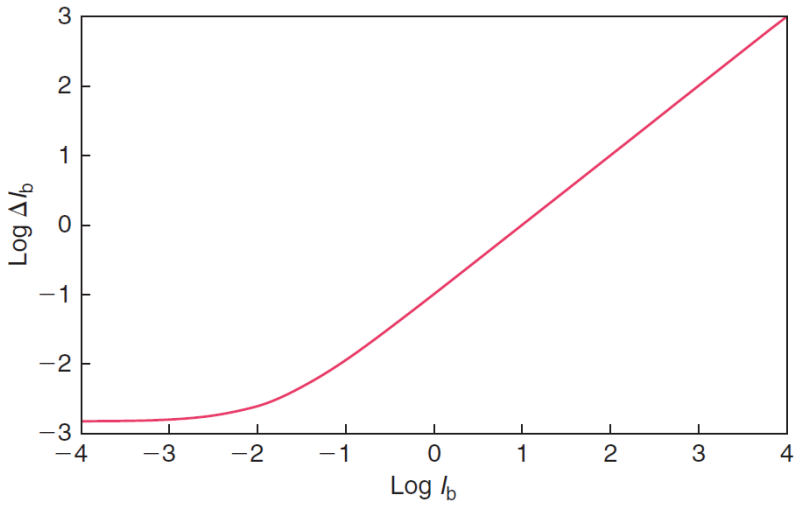



Figure 2.6: Threshold versus intensity (TVI) function. The plot illustrates the noticeable threshold $\Delta I_b$ at various background intensity $I_b$. The image is from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2010 Elsevier Inc.

models are used as theoretical supports by most existing TMOs [86].

In psychophysics, human visual adaptation is studied by measuring the minimum light increment that can be noticed by observers from the background intensity. The minimum amount of increment is called just-noticeable difference (JND). The threshold versus intensity (TVI) function, which provides the relation between JND denoted by $\Delta I_b$ and background intensity given by $I_b$, is illustrated in Figure 2.6. As shown in the function curve, the ratio of $\Delta I_b / I_b$ over much of the intensity range of background is roughly constant. Since the rule was first discovered by Ernst Heinrich Weber, it is known as the Weber's law, and the ratio is called Weber constant. Weber law indicates that visual adaptation scales the scene intensity according to that of the background in order to preserve our ability for contrast perception within a large range of intensity levels.

Human visual adaptation to various illumination conditions is accomplished with the coordinated action of the pupil, the rod-cone cells, and the photoreceptor mechanism [86]. After light passes through the pupil and reaches the retina, it will be absorbed by photoreceptor cells including rods and cones, which function in varying lighting conditions. The photoreceptor cells convert the absorbed light energy into

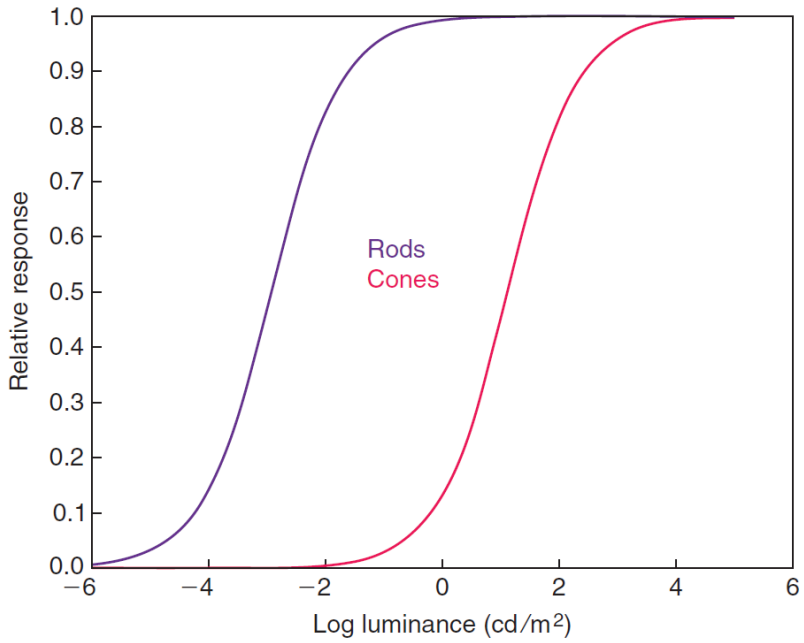



Figure 2.7: Response curve of dark-adapted rod and cone cells to various intensities in arbitrary units. The image is from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2010 Elsevier Inc.

neural responses, and the process is referred as the photoreceptor mechanism. Even though the human visual system performs over a wide range of background intensities, it always maintain its log-linear property for about 3 log units of intensity range. The photoreceptor response curves of dark-adapted rod and cone cells are given in Figure 2.7. The response curve of rods cells appears in the left position because of its higher sensitivity to light. Independent measures have verified that the S-shaped response curve remains the same for different background intensities [86]. The position of the curve will shift horizontally along the intensity axis with varying background intensities, which means that human visual system adapts to new environments and retains local contrast.

### 2.2.2 Tone Mapping Operators

TMOs aim to compress the dynamic range of HDR images to fit into the available display range of a particular device or medium. TMOs can be roughly classified into two categories: global operators [104, 88, 23, 83], and local operators [24, 27, 5, 85].

Global operators handle images as a whole and apply the same transformation to every pixel. The transformation could be realized with logarithmic curve, sigmoid curve, or others derived from characteristics of the human visual system. On the other hand, local operators simulate the visual adaptation mechanism and apply spatially variant transformation for each pixel. Since local operators reduce scene contrast according to neighborhood intensities, they can better preserve local contrast than global operators. Nevertheless, local operators are computationally more expensive, and may introduce artifacts such as halos, which impair the perceived naturalness of tone-mapped images. A good review of existing TMOs can be found in the books by Reinhard et al. [86] and Banterle et al. [9]. Although this is not exhaustive, we briefly describe a number of TMOs that are frequently used [55, 15, 52, 16]. Some of the operators will be adopted in the following chapters.

- **Linear Mapping:** A simple method for tone mapping is to linearly scale the contrast levels of HDR images into the displayable ranges. Linear tone mapping usually causes significant content loss, making it insufficient to accurately reproduce the visual appearance of the original scenes.

- **Drago Logarithmic Mapping [23]:** Drago's logarithmic operator is derived from the log-linear curve of photoreceptor response function. To preserve details while compressing contrast, the algorithm improves the logarithmic compression by introducing an adaptive adjustment of logarithmic bases for different pixel values.

- **Schlick Uniform Rational Quantization [88]:** Schlick's operator applies an uniform rational quantization for the purpose of dynamic range reduction. Compared with other more complete perceptually-based operators, the algorithm provides a simple yet efficient solution to generate realistic looking images on display devices.

- **Ashikhmin Spatially Variant Operator [5]:** Ashikhmin's operator is a local operator based on human visual adaptation. For local contrast preservation, the algorithm calculates the local world adaptation as the average luminance of neighboring pixels, and then employs a perceptual capacity function to compute the local display adaptation and display values.

- **Durand and Dorsey Bilateral Filtering [24]:** The bilateral operator uses an edge preserving smoothing filter, known as a bilateral filter, to separate the HDR scenes into different frequency components. Then, the low-frequency component is compressed and recombined with the high-frequency component to generate the final output results.

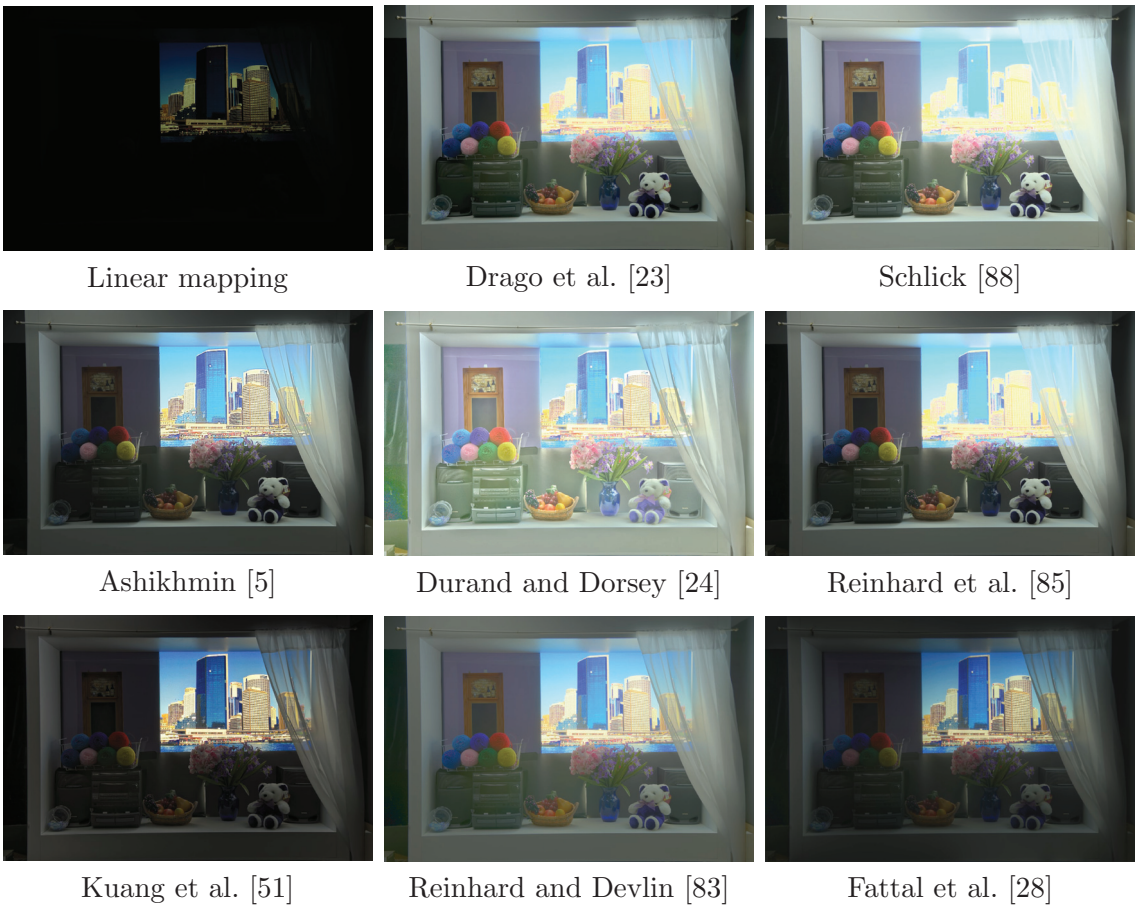
- **Reinhard Photographic Tone Reproduction [85]:** The photographic operator mimics the techniques developed in conventional photography for dynamic range compression. The method employs a sigmoid function to compress the contrast of HDR scenes, and replicates the photographic dodging and burning to increase pixel contrast relevant to the surrounding areas.

- **Image Color Appearance Model (iCAM06) [51]:** iCAM was originally proposed as an image color appearance model [27]. Later, Kuang et al. incorporated an edge preserving spatial filter and light adaptation functions into the color appearance model, and utilized the revised model for HDR images rendering.

- **Reinhard and Devlin Photoreceptor Model [83]:** The photoreceptor operator simulates the mechanisms of photoreceptor adaptation to solve tone mapping problem. Several user parameters are provided that allow control of intensity, contrast, and the adaptation level in terms of light and color.

- **Fattal Gradient Domain Compression [28]:** By performing dynamic range compression in gradient fields, Fattal's operator identifies gradients at varying spatial scales, attenuates their magnitudes with a compressive function, and then integrates the compressed gradients by solving a Poisson equation.

Figure 2.8 illustrates tone mapped images generated by the above-mentioned TMOs with default parameter settings. Since the operators address tone mapping with a variety of methods and goals, they usually lead to images that look quite different from each other. Because different TMOs even parameters yield different tone mapped images, one of most challenging tasks in the domain of tone mapping is to select an appropriate operator and parameter settings for a particular HDR image.



Linear mapping | Drago et al. [23] | Schlick [88]

Ashikhmin [5] | Durand and Dorsey [24] | Reinhard et al. [85]

Kuang et al. [51] | Reinhard and Devlin [83] | Fattal et al. [28]

Figure 2.8: Tone mapping images from various operators on the HDR image "Lab Window". The HDR image is from Mark Fairchild's HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

### 2.2.3 Recent Directions

More recent approaches to HDR tone mapping offer several new directions, such as user-assisted tone mapping [57, 16], styled rendering [8, 2], display conditions aware tone mapping [65, 84], and tone mapping optimization based on objective image quality assessment [108, 34, 35, 36]. Since the development of objective equality metrics and tone mapping optimization algorithms will be discussed in the following context, here we focus on the other three directions.

The TMOs with default parameter settings usually cannot guarantee good results and manual adjustment is required for further improvement. Several user interfaces are introduced to assist in the manipulation process, which allows interactive control. Lischinski et al. [57] present an interactive tool for users to indicate regions

of interest with brush strokes and make local adjustments of tone values and other visual parameters in an image. In the work of Chisholm et al. [16], users are able to iteratively select one of several alternative tone mapped images provided in the interface for parameter optimization rather than tweaking parameters directly.

While conventional TMOs target a natural and faithful rendering of real-world scenes, some algorithms are proposed to learn artistic styles from the predefined examples and generate the results with personal taste and preference. Based on a two-scale non-linear decomposition of an image, Bae et al. [8] adjust visual qualities such as the tonal balance and detail amount to explore various styles on tone mapped images. Also, Akyüz et al. [2] propose an algorithm to learn the style from a set of manipulated images and transfer the learned style for new images.

Display devices can differ dramatically in their peak brightness, contrast, and black level. It can be expected that the same tone mapped image shown on various devices will have different appearances. In order to have accurate representation on a particular device, Mantiuk et al. [65] introduce a novel algorithm which takes ambient illumination and display characteristics into consideration for tone mapping. Moreover, inspired from color appearance model, Reinhard et al. [84] propose an appearance reproduction method to produce HDR images and video for display under specific viewing conditions of environment and devices.

## 2.3 Image Quality Assessment of Tone Mapped Images

Tone mapped images vary across different TMOs and parameter settings. Therefore, a natural question is which tone mapped image most faithfully reproduces the visual appearance of an HDR scene. Generally speaking, the image quality of tone mapped images can be evaluated with both subjective and objective methods. Subjective evaluation can be carried out with psychophysical experiments, where human subjects are asked to make judgements of image quality. On the contrary, objective evaluation is implemented with computational models based on image processing theories or assumptions, which can make quality prediction in an automated manner.

A number of subjective studies have been conducted to compare the existing TMOs and analyze the features that contribute to good image quality. Drago et

al. [21, 22] compare six TMOs by asking subjects to judge the similarity and dissimilarity of pairs of tone-mapped images. The statistical analysis reveals that image quality is most related to apparent naturalness (the degree of resembling realistic scenes) and apparent level of details (the visibility of scene content). Ledda et al. [55] validate six TMOs with HDR scenes displayed on a HDR device. In their experiments, human subjects are asked to make quality assessment based on overall similarity and detail reproduction respectively. Kuang et al. [52] perform a series of experiments to evaluate seven TMOs for their performance in overall preference and reproduction accuracy. Also, Čadík et al. [15] conduct a more comprehensive assessment in which observers are asked to rank the tone mapped images from fourteen operators. Their study suggests an approximation of overall image quality based on the measurement of image features including brightness, contrast, detail reproduction and color appearance.

Subjective methods can have reliable performance for quality evaluation. Nevertheless, they suffer from some fundamental drawbacks. First of all, subjective assessment is built on psychophysical experiments which could be expensive and time-consuming. Secondly, it cannot be incorporated in an automatic framework for tone mapping optimization. Although the limitations could be addressed by objective methods, the conventional image quality metrics commonly assume that the compared image pairs have a similar dynamic range [100], and they are not suitable for quality evaluation of tone mapped images where the assumption does not hold. In an attempt to overcome the drawbacks of conventional methods, several objective quality assessment methods for tone mapped images have been proposed. Using a new definition of visible distortion, Aydin et al. [7] compute an image quality metric for image pairs with arbitrary dynamic ranges. Also, by measuring structural fidelity and statistical naturalness, Yeganeh and Wang [108] propose an objective assessment method called tone mapped image quality index (TMQI). More recently, Ma et al. [61] modify the measurement used in TMQI and introduce TMQI-II as an improved variant.

In this section, we focus on the objective quality assessment of tone mapped images. The current methods, such as dynamic range-independent quality assessment [7], TMQI [108], and TMQI-II [61], are discussed.

### 2.3.1 Dynamic Range-Independent Image Quality Assessment

Based on advanced models of human visual system, contrast distortion metric such as visible difference predictor (VDP) [17] can capture the near threshold differences and scale them in just noticeable difference units. Mantiuk et al. [66] propose an HDR extension of VDP (HDR-VDP) to make prediction of perceivable difference in the full luminance range of HDR images. Taking two HDR images as reference and test images, HDR-VDP uses threshold fidelity measure to analyze the visibility of distortion and output a probability map of detection. Based on the measurement of HDR-VDP, Aydin et al. [7] propose a dynamic range-independent quality evaluation methods for tone mapped images shown on displays. The central idea of the metric is a new definition of visible distortion based on the classification of structural changes, in which three types of distortion can be detected (illustrated in Figure 2.9):

- Loss of visible contrast: the contrast that is visible in the reference image become invisible in the test image, which commonly happens during tone mapping.
- Amplification of invisible contrast: the contrast that is invisible in the reference image become visible in the test image, and it can be caused by contrast stretching in inverse tone mapping.
- Reversal of visible contrast: the contrast can be seen in both reference and test images but with different polarity. This strong distortion is usually related to noticeable artifacts.

The three types of distortion are visualized as an in-context distortion map [17] with an arbitrary color, and the magnitude of detection probability is denoted by the scale of saturation. A sample of the distortion visualization is provided in Figure 2.10, where the loss of visible contrast, amplification of invisible contrast, and reversal of visible contrast are represented with color green, blue, and red respectively. Aydin et al. [7] conduct perceptual experiments to validate the proposed quality metric, and the results show that the maps of contrast distortion have good correlations with the subjective assessment of image distortion types. However, since it is not obvious how the quality maps could be condensed into a single score for an entire image, the metric is not immediately applicable for tone mapping optimization.

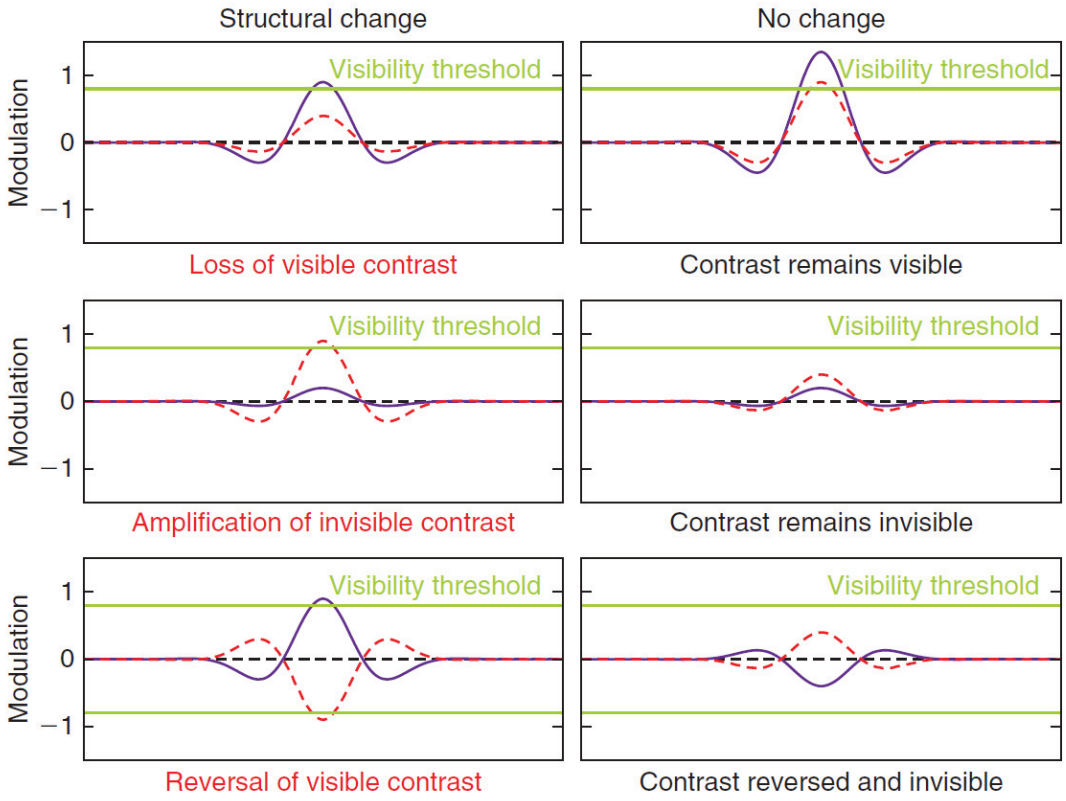



Figure 2.9: Three types of contrast distortion that the quality metric classifies as a structural change (left) or a lack of structural change (right). The solid and dashed lines denotes the reference and test signals, and the horizontal lines depict the visibility threshold [7]. The image is from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2010 Elsevier Inc.

### 2.3.2 Tone Mapped Image Quality Index

A milestone in the development of objective quality assessment is the introduction of the Tone Mapped Image Quality Index (TMQI) by Yeganeh and Wang [108]. TMQI measures the image quality in terms of structural fidelity and statistical naturalness, and assigns a single quality score to each image, which makes it suitable for the application of quality improvement in an optimization framework.

The structural fidelity is an improved SSIM index [102, 101] for the comparison of image pairs with various dynamic ranges. Based on the assumption that the perceived changes in structure information can be used as predictor of image quality,

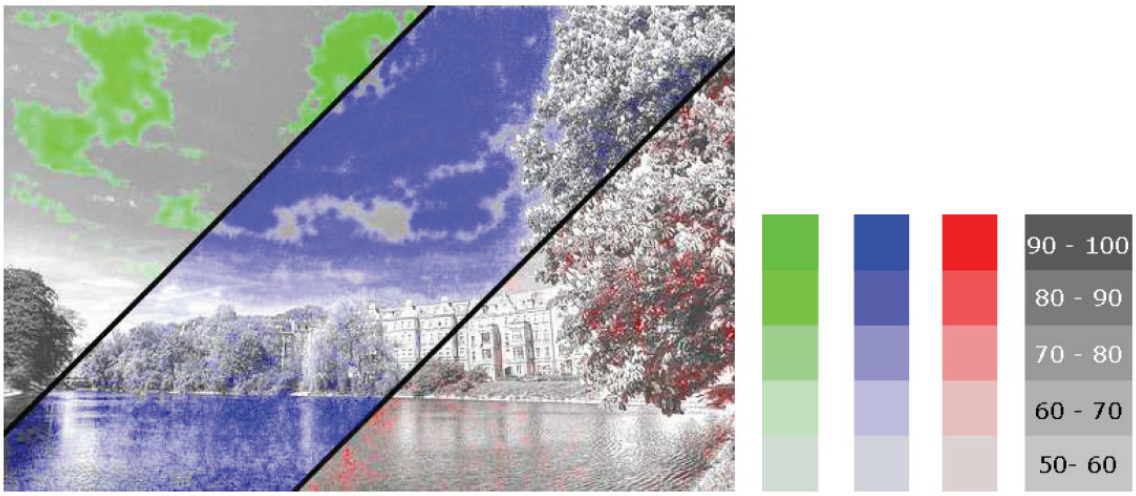



Figure 2.10: Distortion maps that are partially shown and saturation scales that indicate the magnitude of detection probability [7]. Green depicts loss of visible contrast; blue denotes amplification of invisible contrast; red represent reversal of visible contrast.

SSIM calculates the similarity between two images based on their spatial dependencies of pixels. The original SSIM algorithm consists of three comparison components including luminance, contrast, and structure. Since tone mapping causes significant changes in luminance and contrast, the direct comparison of these components is inappropriate for quality assessment of tone mapped images. To address that, the structural fidelity modifies the SSIM algorithm by discarding the luminance component and redefining the contrast component. In the new definition, the difference of signal strength will not be penalized if the signal strengths of the HDR and LDR images are both significant (above a visibility threshold) or both insignificant (below a visibility threshold). The measurement of structural fidelity is performed on two local image patches from the HDR and the tone mapped LDR images and a sliding window is adopted to run across the entire image space. Inspired by multi-scale SSIM [103], the structural fidelities are analyzed at multiple scales, where the images are iteratively processed by low-pass filtering and downsampling to create an image pyramid. A quality map that reflects the variation of structural fidelity across the image space is generated at each scale, and the maps of different scales are averaged into a score $S_l$ where $l$ denotes the number of the scale. The quality score $S_l$ can be described as follows,

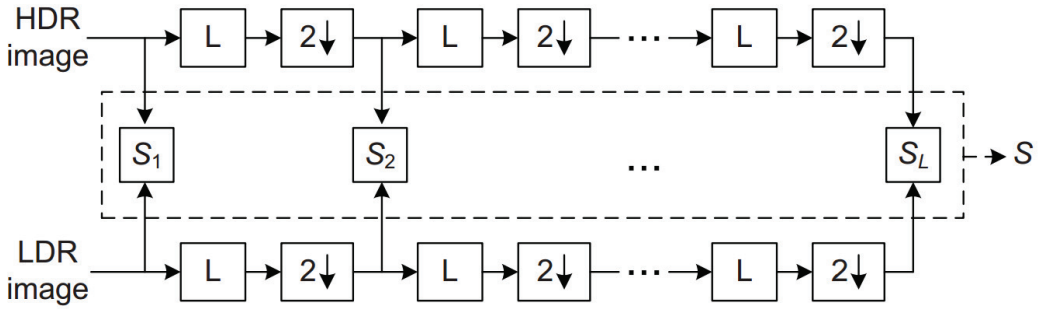



Figure 2.11: Framework of multi-scale structural fidelity measurement [108].

$$S_l = \frac{1}{N_l} \sum_{i=1}^{N_l} S_{local}(x_i, y_i). \tag{2.1}$$

In the equation, $x_i$ and $y_i$ denote the $i-th$ patch of the HDR reference and LDR test images respectively, and $N_l$ denotes the number of patches. Afterwards, the overall structural fidelity $S$ is computed by combining the scores of different scales

$$S = \prod_{l=1}^{L} S_l^{\beta_l}, \tag{2.2}$$

in which $L$ denotes the total number of scales and $\beta_l$ denotes the weight assigned to the $l-th$ scale. Based on the psychophysical experiment [103], $L$ is set to 5 and $\{\beta_l\}$ is set to $\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$. The framework of the multi-scale structural fidelity measurement is demonstrated in Figure 2.11. It is worth mentioning that the measure is performed only on the luminance channel for color images.

Assuming that good quality tone mapped images should should look as "natural" as possible, Yeganeh and Wang [108] also introduce a scene-independent measurement named as statistical naturalness, which calculates the similarity of brightness and contrast between the tone mapped images and natural images. They assert that the measurement of statistical naturalness can best complement the structural fidelity where brightness modeling and evaluation are missing. The statistical naturalness is built upon a large data set consisting of about 3000 8 bits/pixel gray-scale images that represent many different types of natural scenes. As shown in Figure 2.12, the means and standard deviations of the images are analyzed to indicate the overall brightness

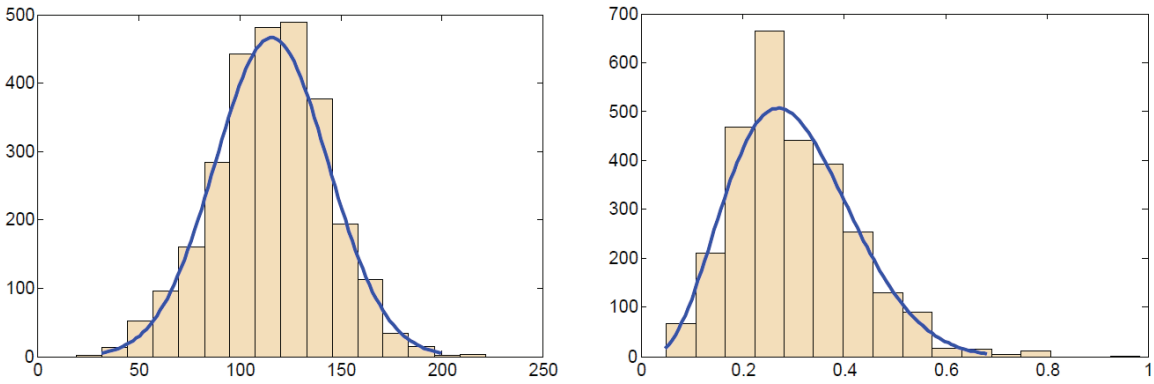



Figure 2.12: Histograms of means fitted by Gaussian PDF (left) and standard deviations fitted by Beta PDF (right) of natural images [106].

and contrast of "natural scenes", and a penalty is given to tone mapped images with different brightness and contrast. In order to quantify the distortion, the histograms of means $m$ and standard derivations $d$ are fitted with a Gaussian probability density function $P_m(m)$ and a Beta probability density function $P_d(d)$ respectively, which are formulated as follows,

$$P_m(m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\{-\frac{m-\mu_m}{2\sigma_m^2}\} \tag{2.3}$$

$$P_d(d) = \frac{(1-d)^{\beta_d - 1} d^{\alpha_d - 1}}{B(\alpha_d, \beta_d)} \tag{2.4}$$

where $B()$ denotes the Beta function. The parameters used in the formula are estimated with regression: $\mu_m = 115.94$, $\sigma_m = 27.99$, $\alpha_d = 4.4$, and $\beta_d = 10.1$. The fitted curves of these two functions are given in Figure 2.12. With the estimations of distortions in brightness and contrast, the statistical naturalness measure $N$ is defined as the product of $P_m(m)$ and $P_d(d)$:

$$N = \frac{1}{K} P_m P_d \tag{2.5}$$

where $K$ is a normalization factor computed as $K = max\{P_m P_d\}$.

With the structural fidelity measure $S$ and the statistical naturalness measure $N$, the TMQI measure $Q$ is defined as a three-parameter function to combine these two measures:

$$Q = aS^{\alpha} + (1-a)N^{\beta} \tag{2.6}$$

In this function, $a$ controls the relative importance of the two components, while $\alpha$ and $\beta$ adjust their sensitivities respectively. Since both $S$ and $N$ are upper-bounded by 1, the overall quality measure $Q$ is upper-bounded by 1 as well. With an iterative learning process, the parameters are tuned to best fit the subjective evaluation data provided by Song et al. [90], and the settings are $a = 0.8012$, $\alpha = 0.3046$, and $\beta = 0.7088$.

The TMQI was then validated by comparing the objective quality assessment results with subject-rated image data, and the results show good correlation between the TMQI measure and subjective ranking scores. Nevertheless, the TMQI suffers from several drawbacks which could lead to inaccurate prediction. First of all, the visibility threshold used in the structural fidelity is too sensitive, and any tiny changes in local patch in HDR images could cause a significant difference in quality measure [61]. Secondly, the statistical naturalness measurement always favors the images with "average" brightness and contrast, and that bias towards "average" images is too coarse to make accurate quality prediction for HDR scenes under a broader range of illuminations [34, 35, 61]. The drawbacks of the statistical naturalness in the quality evaluation and application of tone mapping optimization will be further discussed in the following chapters.

### 2.3.3 Tone Mapped Image Quality Index II

Most recently, the limitations of TMQI have been recognized by Ma et al. [61], and they propose an improved variant of the quality measure named TMQI-II, in which the original measurement of structural fidelity and statistical naturalness are updated for better accuracy.

Using a sliding window across the image space, the structural fidelity of TMQI can generate a quality map that indicates the preservation of local structural information. Compared with the original SSIM index [102, 101], the measurement modifies the contrast comparison by suggesting that the HDR and tone mapped images patches should have the same visibility of local contrast. To access the contrast visibility, they apply a nonlinear function derived from the contrast sensitivity model for the

Figure 2.13: Structural fidelity map on the HDR image "Belgium house" (brighter indicates higher quality) [61]. Left: tone mapped image from photographic operator [85]. Middle: structural fidelity map generated by TMQI. Right: structural fidelity map generated by TMQI-II.

contrast which is calculated as the local standard deviation. However, as mentioned earlier, the threshold determined by the function suffers from the problem of over sensitivity, which in turns mistakenly marks the invisible contrast as visible contrast. The phenomenon is illustrated in Figure 2.13. The structural fidelity measure of TMQI incorrectly recognizes the homogeneous wall areas as contrast visible in HDR images, and then applies quality penalties for these areas of the tone mapped image resulting in inaccurate predictions. Assuming that the contrast visibility function should be adapted to the local luminance levels of HDR images, Ma et al. employ the coefficient of variation as an approximation of the local contrast in HDR image patches rather than the standard deviation. They assert that the modification is consistent with Weber's law and remains invariant to linear contrast stretching. The quality map generated by the updated structural fidelity is shown in Figure 2.13 (right). In the same fashion with TMQI, the single quality score for the entire image is computed as the weighted average of the quality maps across multiple scales.

The statistical naturalness in TMQI is built on the statistical average of brightness and contrast derived from a large set of natural images. Using a Gaussian density function $P_m$ and a Beta density function $P_d$ is used to indicate the naturalness qualities of brightness and contrast respectively, the overall quality score of statistical naturalness is calculated as the normalized product of these two density functions. The measure of statistical naturalness is completely independent of the image content of reference HDR scenes, which is an over simplification. The model suggests

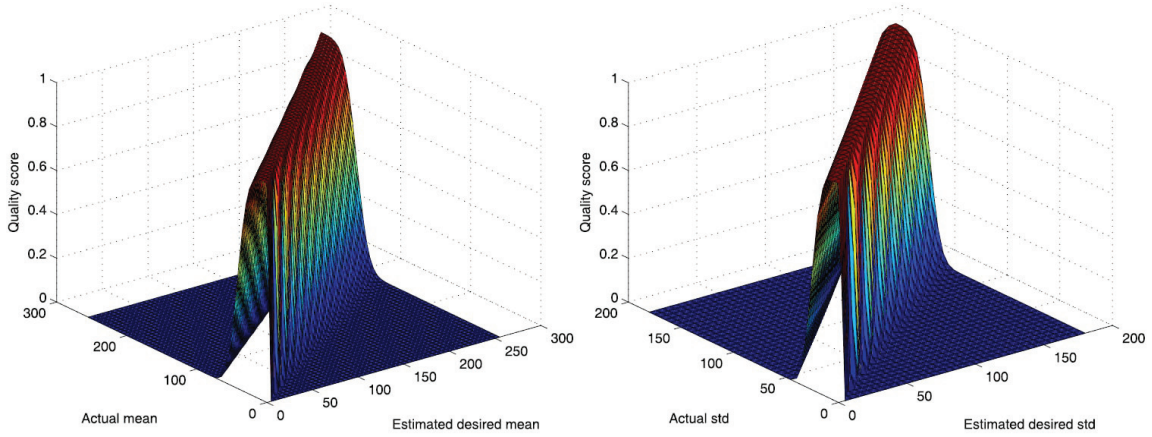



Figure 2.14: Surfaces of the functions $P_m$ (left) and $P_d$ (right) [61].

that statistically natural tone mapped images of dynamic range $[0, 255]$ should have an average brightness around 116 and a contrast around 65 which correspond to the peaks in the density functions. However, each tone mapped image should have different brightness and contrast values to look natural depending on the original scenes. In order to overcome the limitation, Ma et al. propose an image dependent naturalness model. They compress the luminance of HDR images based on the logarithmic average, and calculate the mean $\mu_e$ and standard deviation $\sigma_e$ of the compressed luminance as the estimation of brightness and contrast for naturalness. Asserting that the measures should remain in certain ranges to be "natural", they introduce acceptable boundaries $[\mu_l, \mu_r]$ and $[\sigma_l, \sigma_r]$ for brightness and contrast respectively. Afterwards, with the approximated values and boundaries of the two components, the functions $P_m$ and $P_d$ in statistical naturalness are redefined using Gaussian cumulative distribution functions, which are described as follows,

$$P_m = \begin{cases} \dfrac{1}{\sqrt{2\pi}\theta_1} \displaystyle\int_{-\infty}^{\mu} exp\left(-\frac{(t-\tau_1)^2}{2\theta_1^2}\right) dt & \mu \leq \mu_e \\ \dfrac{1}{\sqrt{2\pi}\theta_2} \displaystyle\int_{-\infty}^{2\mu_r - \mu} exp\left(-\frac{(t-\tau_2)^2}{2\theta_2^2}\right) dt & \mu > \mu_e, \end{cases} \tag{2.7}$$

$$P_d = \begin{cases} \dfrac{1}{\sqrt{2\pi}\theta_3} \displaystyle\int_{-\infty}^{\sigma} exp\left(-\frac{(t-\tau_3)^2}{2\theta_3^2}\right) dt & \sigma \leq \sigma_e \\ \dfrac{1}{\sqrt{2\pi}\theta_4} \displaystyle\int_{-\infty}^{2\sigma_r-\sigma} exp\left(-\frac{(t-\tau_4)^2}{2\theta_4^2}\right) dt & \sigma > \sigma_e \end{cases} \tag{2.8}$$

where the parameters $\tau_i$ and $\theta_i$ ($i \in [1, 2, 3, 4]$) are determined by points on the curve; see [61] for details. Figure 2.14 illustrates the surfaces of the two functions. It can be observed that heavy penalties will be applied for tone mapped images whose brightness or contrast is outside of the acceptable boundaries.

The overall quality of TMQI-II is defined as the weighted average of the measures of updated structural fidelity and statistical naturalness. Instead of learning weights from subjective data, TMQI-II emphasizes the equal importance and employs the same weights for these two components. The comparison of TMQI and TMQI-II is conducted with a subject-ranked database, and the experiment shows that the updated measurements can have improved correlation with respect to subjective evaluations. As the improved variant of TMQI, TMQI-II has been used as a baseline to compare our feature-based quality metric against.

### 2.3.4 Other Methods

Other than the state-of-the-art methods discussed in the previous sections, a few other image quality assessment approaches for tone mapped images have been proposed, and most of them use partially or completely scene-independent measures.

Derived from the TMQI [108], Nasrinpour and Bruce [75] propose a saliency weighted tone mapped quality index that use visual saliency to further align the human adjustments of image quality. They modify the measurement of structural fidelity in TMQI by introducing visual saliency in the pooling strategy to compute the overall score. Since the statistical naturalness component remains the same, their method inherits the limitations of TMQI regarding the over-simplification of quality judgement in terms of brightness and contrast. A similar saliency-based modification of TMQI has also been presented by Liu et al. [59] which suggests a new pooling method for the structural fidelity using visual saliency while having the statistical naturalness unchanged. Based on the local phase information of images, Nafchi et al. [72]

use the locally weighted mean phase angle map for quality evaluation, and incorporate the phase-based measure in TMQI as an improved quality metric. This approach still uses the scene-independent statistical naturalness to make explicit constraints on brightness and contrast, which could introduce artifacts in quality predictions. Also, asserting that higher quality tone mapped images should maintain much more details, Gu et al. [39] propose a totally blind quality metric by estimating the amount of local details, in which the details are quantified as the entropies in the brightened and darkened tone mapped images.

Although no-reference measures provide an easy solution when reference images are not immediately available for comparison, they are too coarse to make accurate prediction for the breadth of real world images. Some discussion of the limitations can be found in the prior works [35, 61]. Considering the ultimate goal of tone mapping is to reproduce the visual appearance of HDR scenes [86], we mainly focus on the full-reference objective quality assessment for tone mapped images, which calculates the distortion between compressed LDR and original HDR images.

## 2.4 Evolution Strategies

Evolution strategies (ESs) [81, 89], also referred to as evolutionary strategies, are optimization techniques based on the principles of biological evolution. They belong to a general class of evolution methodologies that take the candidate solutions to the optimization problem as individuals in a population and select the better individuals in each generation.

ESs are most commonly used for black-box optimization problems where derivative information is not available. They are relatively robust with regard to the ruggedness of the objective function (rugged fitness landscape). In the present context, evolution strategies are useful as the quality of a tone mapped image can only be evaluated through application of the TMOs, and no convenient mathematical assumptions regarding the objective can be made. This section briefly introduces the main principles of the optimization algorithm and parameter control. A thorough survey of evolution strategies was conducted by Hansen et al. [40].

### 2.4.1 Algorithm Description

From the algorithmic viewpoint, ES are stochastic optimization algorithms that iterate variation and selection in populations of candidate solutions. The search space is the continuous domain $\mathbb{R}^n$ and the solutions is search space are $n$-dimensional vectors. We assume a population of individuals. Each individual represents a candidate solution, and consists of a parameter vector $x \in \mathbb{R}^n$ and an associate fitness value $f(x)$. In some cases, there is only one individual in the populations. Depending on the roles of individuals in a generation, they can also be denoted as parents or offspring.

For each generation, one or several parents are selected from the current population, which is called mating selection, and offspring are generated from the selected parents by duplication and recombination. Then, the offspring undergo mutation which introduces small, random, and unbiased changes to these individuals. The average size of the changes are adapted over time, and the parameters that allow the control of adaption is called control parameters. For instance, the step-size $\sigma$ determines the notion of "small". Although the method of parameter control is not always directly inspired by biological evolution, it can be a central feature for evolution strategies. After mutation, the offspring become new members, and the size of the population grows accordingly. Next, based on the fitness value, environment selection reduces the population to its original size, in which only the best individuals can survive and become the parents for next generation. There exist two basic strategies for the environment selection: plus- and comma- selection. In plus-selection, individuals' age is not considered, and the best individuals are selected from both parents and offspring. Comma-selection takes age into account, and only the offspring can survive to the next generation.

### 2.4.2 Parameter Control

Controlling the parameters of mutation is key to the design of ES [40]. The step-size $\sigma$ is a scaling factor for the random vector perturbation, and it plays an important role in parameter control. To a large extent, the step-size controls the convergence speed of the algorithms. When larger step-sizes lead to larger expected improvement, the step-size control techniques should increase the step-size to meet the target. Conversely, when smaller step-sizes contribute to better performance, the techniques

should decrease its value instead. The control of the step-size can be performed on different levels: the step-size can have different values for various individuals, or a single step-size value can be applied to all individuals.

**Algorithm 1** The $(1+\lambda)-ES$ with the Chisholm et al.'s Strategy

1: **given** $n, \lambda \in \mathbb{N}_+, \alpha \in [0,1]$
2: **initialize** $P = \{(\mathbf{x}, \sigma, f(\mathbf{x})), \mathbf{x} \in \mathbb{R}^n, \sigma > 0$
3: **while** not happy **do**
4: **for** $k \in \{1, ..., \lambda\}$ **do**
5: $\mathbf{x}_k = \mathbf{x} + \sigma \times \mathbf{N}(\mathbf{0}, \mathbf{I})$
6: $P \leftarrow P \cup \{(\mathbf{x}_k, \sigma, f(\mathbf{x}_k)) \mid 1 \leq k \leq \lambda\}$
7: $P \leftarrow$ **select_1_best**$(P)$
8: **if** $P$ doesn't change **then**
9: $\sigma \leftarrow \sigma \times \alpha$

A number of ES have been proposed with various parameter control methods, such as the $1/5th$ success rule [81], self-adaption [40], and cumulative step-size adaption (CSA) [4]. The $1/5th$ success rule [81] is established in the early stage of the development of evolutionary strategies. It uses 1/5 as the success value to switch between increasing and decreasing the step-size: step-size increases if the success probability is larger than 1/5 and decreases otherwise. The parameter control via self-adaptation is seminal in the domain of evolution strategies. In self-adaption [40], the generation of new step-size is similar with that of offspring, which is conducted by recombination and mutation. As for the cumulative step-size adaption [4], it introduces a cumulative path to guide the adaption of step-size. The cumulative path is the combination of all steps made by the algorithm with their importance decreasing exponentially over time.

Here, we use a simple method proposed by Chisholm et al. [16] as an example to introduce the procedure of parameter control. The method was proposed for interactive evolution of tone mapping, and it is efficient to yield good converge within a relatively small number of iterations. In their approach, the step-size is updated with a strength factor when the parent is superior to all its offspring. The algorithm 1)

demonstrates the implementation of $(1 + \lambda) - ES$ with the strategy. Given a population $P$ of 1 individual $(\mathbf{x}, \sigma, f(\mathbf{x})), \mathbf{x} \in \mathbb{R}^n$ (Lines 2), where $\mathbf{x}$ is a solution vector, $\sigma$ is the step size, and $f(x)$ is the fitness function from $\mathbb{R}^n$ to $\mathbb{R}$. Firstly, $\lambda$ offspring are generated from the single parent $\mathbf{x}$ with mutation defined by the step-size $\sigma$ (Lines 5). Afterwards, the new offspring are added to $P$ (Line 6), and the best individual in $P$ that survives the environmental selection will be parent for next generation (Line 7). If parent remains the same with the previous generation, the algorithm updates the step-size $\sigma$ by multiplying a strength factor $\alpha$ (Line 9). This process will continue until satisfying results are arrived.

# Chapter 3

# Visual Saliency Analysis on HDR Images

## 3.1 Introduction

By simulating the selective-attention mechanism in the human visual system [31], visual saliency analysis aims to detect the salient targets that attract eye fixation. Because saliency analysis can predict the regions of interest and allow the prioritized allocation of resources in subsequent applications, it plays an important role in computer vision and image understanding.

However, typical visual attention models are built on LDR images, and they are not suitable for saliency analysis on HDR images. When applying the models to HDR images, high dynamic range will be scaled which inevitably causes content loss that in turn makes salient regions appear not salient or vice versa. Image contents within HDR images may not register as salient due to the significant contrast reduction, preventing the current computational models from obtaining useful results. Also, Narwaria et al. [74] have found that TMOs can modify human attention and fixation behavior significantly, making it unreliable to apply salience analysis techniques after tone mapping.

In this chapter, we introduce a novel algorithm for saliency analysis of HDR images based on virtual photographs [33]. Inspired by the process of generating HDR images from multiple LDR exposures, we decompose an HDR image into multi-exposed virtual photographs for the preservation of image content, and then incorporate the virtual photograph sequence for visual saliency analysis. Experiments with a variety of HDR images demonstrate that our algorithm can produce more consistently reliable predictions than the existing methods.

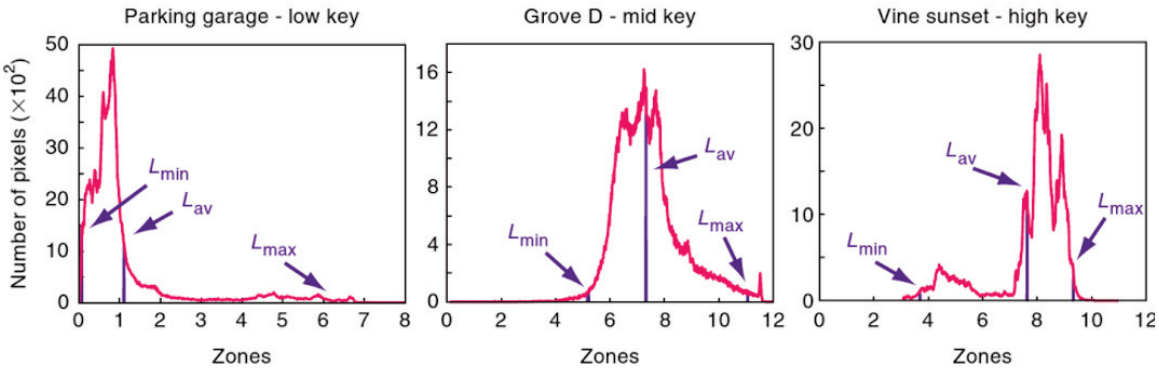



Figure 3.1: Histogram for HDR scenes that are overall dark (left), medium (middle), and light (right). Images from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2010 Elsevier Inc.

## 3.2 Virtual Photograph Based Visual Saliency Analysis

### 3.2.1 Taking Virtual Photographs

Taking virtual photographs was first introduced for flash-exposure HDR imaging [1, 87], in which a series of images with different flash intensity and exposures are captured and merged into HDR maps, and virtual photographs can then be taken for any combination of exposure and flash intensity. In these methods, the response function of the camera that is recovered when producing an HDR image is applied for taking virtual photographs from the same HDR image. Consequently, it is assumed that the response function is already available.

For HDR images that are captured directly by a digital camera, generated synthetically using ray tracing, or generated using other approaches different from those that combine multiple LDR images, a response function may not be readily available. In order to obtain well-exposed virtual photographs from an HDR image if the response function of the capturing device is not available, we adopt a simple transfer function often used in modern photography [91] in combination with the key of a scene based calibration.

For natural scenes, the lighting conditions can be revealed in the histogram of the images. As shown in Figure 3.1, a dark scene often displays a peak towards the left of the histogram, a light scene on the right of the histogram, and medium ones in the

middle of the histogram. Based on this observation, Reinhard [82] introduces the key of a scene which is a unitless number related to the overall lighting level, and proposes a calibration algorithm for HDR images. The same heuristic has been adopted to calibrate HDR images for taking virtual photographs. Based on the distance of the logarithmic average luminance to the minimum luminance in an image relative to the difference between the minimum and maximum luminance, the key of a scene $\alpha$ can be calculated as follows,

$$f = \frac{2\log L_{av} - \log L_{min} - \log L_{max}}{\log L_{max} - \log L_{min}} \ , \tag{3.1}$$

$$\alpha = 0.18 \times 4^{f} \tag{3.2}$$

in which $L_{av}$, $L_{min}$, $L_{max}$ denote the log average luminance, minimum luminance, and maximum luminance, respectively. To reduce the influence of outliers, 1% of the lightest and darkest pixels are excluded for the calculation of minimum and maximum luminance [82]. For an HDR image with luminance values $L(x, y)$, we calibrate the luminance so that the log average value $L_{av}$ can be scaled to the estimated key of the scene $\alpha$. The calibrated luminance $L'(x, y)$ is defined as

$$L'(x, y) = \frac{\alpha}{L_{av}} L(x, y). \tag{3.3}$$

With the calibrated luminance, we apply a transfer curve that mainly compresses high luminance to take virtual photographs with different exposures. The exposures can be calculated as the product of the calibrated luminance $L'(x, y)$ and exposure time $\Delta t$. Then, denoting the displayable luminance value as $L_d(x, y)$, the process of taking virtual photograph is described as

$$L_d(x, y) = \frac{L(x, y)\Delta t}{1 + L(x, y)\Delta t}. \tag{3.4}$$

A comparison between real and virtual photographs is shown in Figure 3.2, where the real photographs have been used to generate the HDR images. As we can see, the virtual photographs look very similar to the real originals. We do note some minor differences, since the exposure times of virtual photographs do not completely match those of the real photographs. Figure 3.3 below presents evidence that the differences are too small to significantly affect the subsequent saliency analysis.

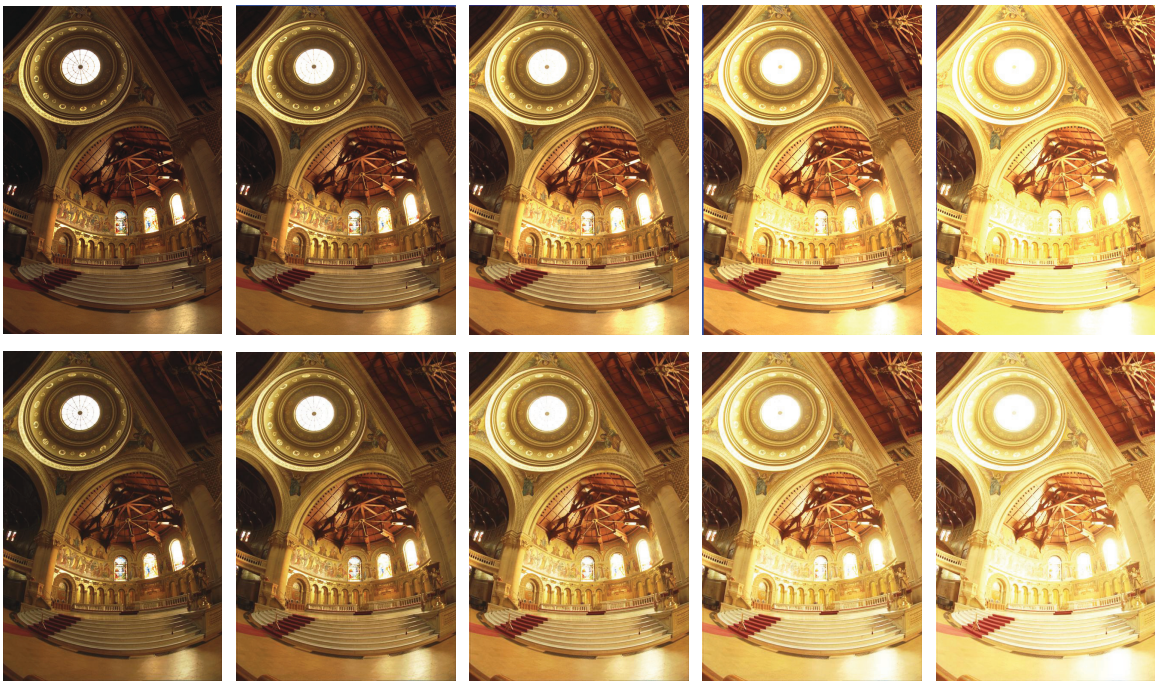


Figure 3.2: Comparison between real photographs (upper row) and virtual photographs (lower row) on HDR image "Memorial Church". Exposure times in all cases increase by a factor of two between neighboring images when moving from left to right. The HDR image that the virtual photographs are derived from as well as the real photographs are from Paul Debevec © 1997 ACM.

### 3.2.2 Analyzing Visual Saliency

In order to preserve image contents of HDR images while reducing artifacts that may be introduced by under- or over-exposed images, we calculate a sequence of nine virtual photographs with exposure times $\Delta t \in \{1/15, 1/8, 1/4, 1/2, 1, 2, 4, 8, 15\}$. We adopt the method of Itti and Koch [46] to calculate the saliency map of each image in the sequence of virtual photographs, but other salience detection operators could be used instead.

Then the saliency maps are combined into the saliency map of the HDR image. One difficulty in combining different saliency maps is the signal-to-noise ratio problem identified by Itti and Koch [46], which means that some salient objects may be weakened or entirely lost during combination. We solve this problem by adopting the spatial competition scheme used by Itti and Koch [46] for noise reduction in the feature maps, which is realized with a two-dimensional Difference-of-Gaussians (DoG) filter. The salient locations in each saliency map can be excited with counteraction

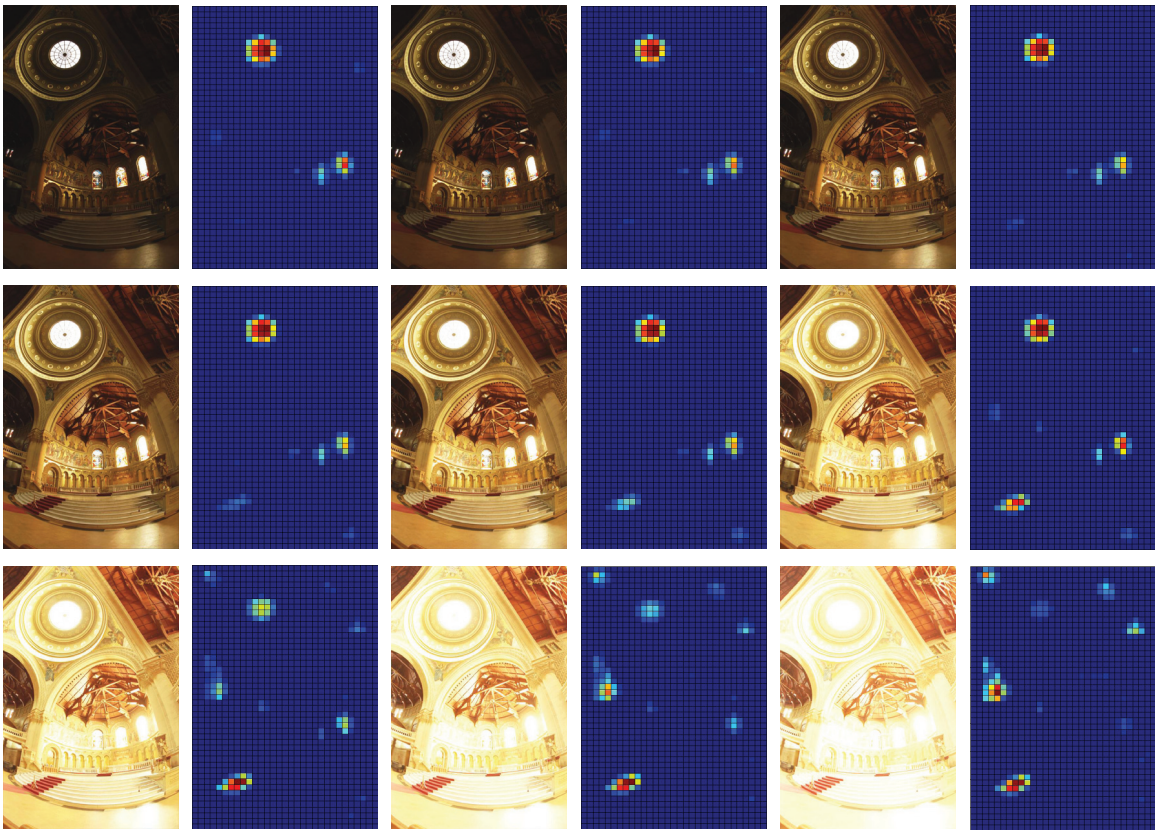


Figure 3.3: Virtual photographs (left) and their corresponding saliency maps (right). The HDR Image is from Paul Debevec © 1997 ACM.

triggered by the inhibition from the surrounding regions.

After the competition, the saliency maps are combined into the saliency map for the HDR image by computing their weighted average, where weights decrease with increasing degrees of over- and under-exposure. A similar strategy is used when producing HDR images from multiple LDR images with different exposures. Debevec and Malik [18] make use of a simple hat function for the weights while Mann and Picard [63] use the derivative of the response curve as the weighting function. We employ $\phi_\sigma(\log_2 \Delta t)$ with $\phi_\sigma(x) = \exp(-(x/\sigma)^2/2)$ and $\sigma = 1.2$ to weight the saliency map obtained from the virtual photograph with exposure time $\Delta t$.

With the saliency map, the focus of attention can be directed to the regions with the highest values. To that end, we utilize a winner-take-all (WTA) neural network [50, 46] to determine the saliency locations and a fixed-size circle (one eighth of the smaller of the input image width and height) to represent the focus of attention. WTA networks work well to simulate the selection mechanism of the human brain [31].

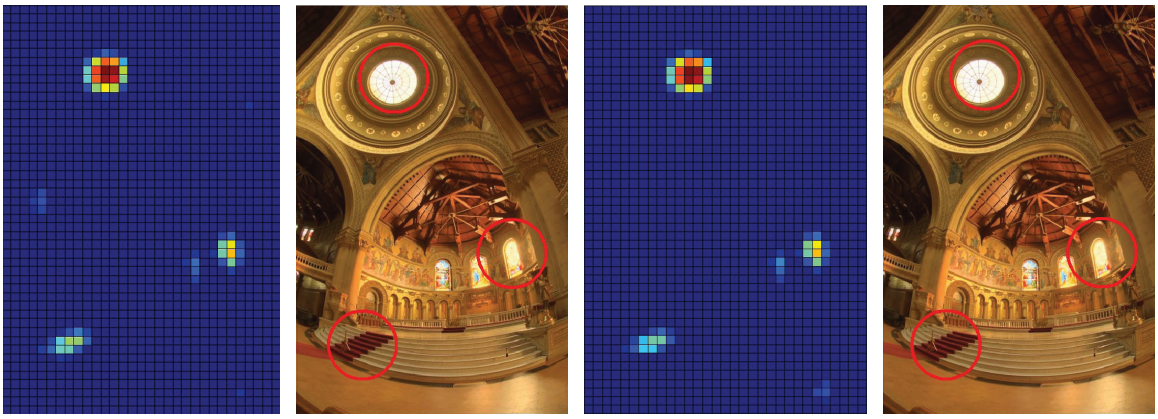


Figure 3.4: Comparison between the results generated from virtual and real photographs. From left to right: saliency map from virtual photographs, regions of visual attention (red circles) from virtual photographs, saliency map from real photographs, and regions of visual attention from real photographs. The HDR Image is from Paul Debevec © 1997 ACM.

A sequence of virtual photographs and their saliency maps are shown in Figure 3.3. It is apparent that some salient features are present only in a subset of the virtual photographs, and therefore only in some parts of the dynamic range. The final results from combining the salience maps of the virtual photographs as well as of the real photographs using the algorithm described above are shown in Figure 3.4. Here as well as in the following figures, HDR images are tone mapped using a local adaption algorithm [85] with manual manipulation of the curve and saturation for display purposes. As shown in this example, the virtual photographs contribute to similar saliency maps and the same locations of visual attention as the sequence of real photographs.

### 3.2.3 Experimental Results

A number of computational visual attention models have been proposed during the past decades. However, it is still challenging to establish an objective criteria for the performance of visual attention models. Although eye-tracking data can serve as the ground truth for human fixation, the previous studies show that predictive models of human fixation is beyond the possibilities of bottom-up visual attention models [13, 49], which limits its application for the performance judgement of visual

attention models. Similar to the prior works in the domain [47, 46], we test our algorithm on various natural HDR images to detect the targets that can be of interest for human observers, where all images are provided in color and contain a number of "targets" with discriminative features. The experimental results indicate that our method is able to produce reasonable predictions. In each case, the eye-catching targets that differ from the surrounding environment by their unique colors, intensities, or orientations, can be marked as visually salient. Three examples are shown in Figure 3.5. As we can see, the targets with discriminative visual features can be found, such as the red car (First row), the ancient buildings (Second row), and the trees (Third row).

A comparison with results generated by applying the method of Itti and Koch [46] directly to the HDR image as well as those obtained using the algorithm of Brémond et al. [13] is shown in Figures 3.6, 3.7, and 3.8. Compared with the methods of Itti and Koch [46] and Brémond et al. [13], our method addresses the problem of HDR content loss that occurs due to contrast compression, and therefore consistently leads to better performance for detecting the salient regions of HDR images. As illustrated in Figure 3.6, if the luminance of the HDR image is compressed before the saliency analysis (as is the case with the other methods), it is hard for the visual models to analyze image contents in the relatively darker areas of the image. In this example, only the sun and its reflection on the water surface are detected by the other approaches, while the targets in other regions, such as the mountains, sailboat, and house, are neglected. The same pattern occurs in Figure 3.7. The previously cited methods fail to capture the bridge in the background, which is darker than the foreground. Another example is given in Figure 3.8. Our method detects the woman that is obviously the most important target, while the other methods cannot. We hypothesize that this is because the image contents in relatively dark areas are weakened or even entirely lost as a result of significant changes in contrast in the other approaches.

## 3.3 Conclusion and Discussion

A number of computational visual attention models have been proposed for LDR images. However, relatively poor results have been observed when applying some

of those systems for saliency analysis of HDR images since lightness reduction will lead to significant content loss. To preserve content of HDR images and incorporate them into the saliency analysis, a virtual photograph based method is proposed in this chapter. The approach utilizes a sequence of virtual photographs rather than a single image for revealing HDR content. Our method reliably characterizes regions of visual attention for HDR images and has a wide variety of potential applications, such as HDR image or video coding, visual saliency based tone mapping, and objective evaluation of tone mapped images. The virtual photograph technique presented in the algorithm opens new avenues to extract image features from HDR images, which can also be used for other features, such as brightness, and local detail.

There are some future research directions which can be pursued to extend the work in this chapter. First of all, our method adopts the key of a scene based calibration when taking virtual photographs from HDR images, and the incorrect estimation of the key of a scene can somewhat effect the following feature analysis. Considering the calibration algorithm is derived from the observation on natural scenes, our method cannot guarantee reliable feature analysis on artificial HDR images. In the future, we would like to explore alternative calibration algorithms or further improvements that are applicable for different types of HDR images. Secondly, the current implementation of our method is built on the visual attention model of Itti and Koch [46], and thus it could suffer from the limitations of their particular method. Since our method is flexible enough to adopt any other visual attention models, testing variant existing models and comparing them can be considered as future work.

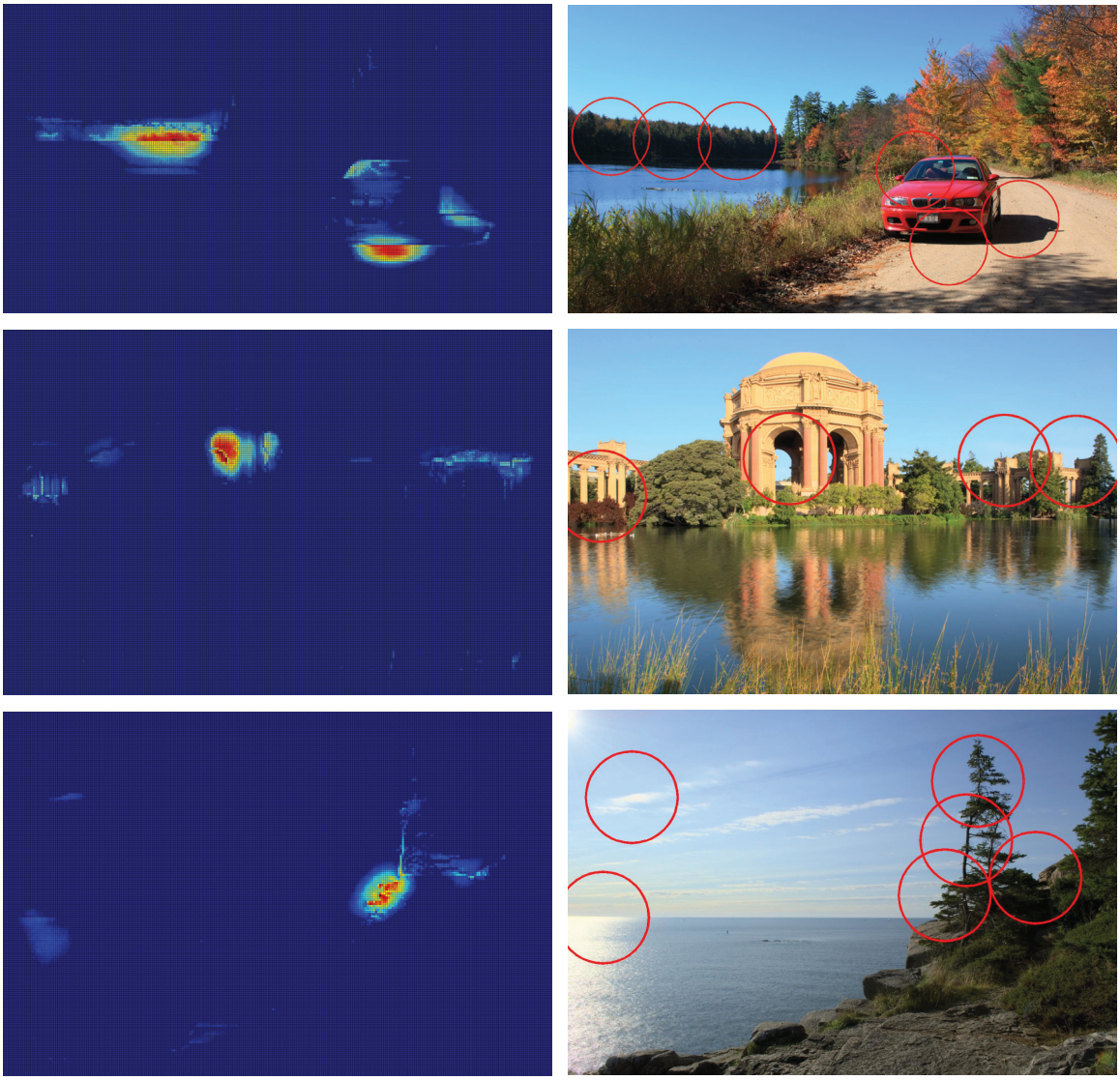


Figure 3.5: Results of saliency analysis on the HDR images “M3 Middle Pond”, “Exploratorium(1)”, and “Otter Points”. First column: saliency maps generated with our method. Second column: prediction of visual attention (red circles). The HDR images are from Mark Fairchilds HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

Figure 3.6: Comparison with other approaches on the HDR image "Bar Harbor Sunrise". From top to down, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [46], and the method of Brémond et al. [13]. The HDR image is from Mark Fairchilds HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

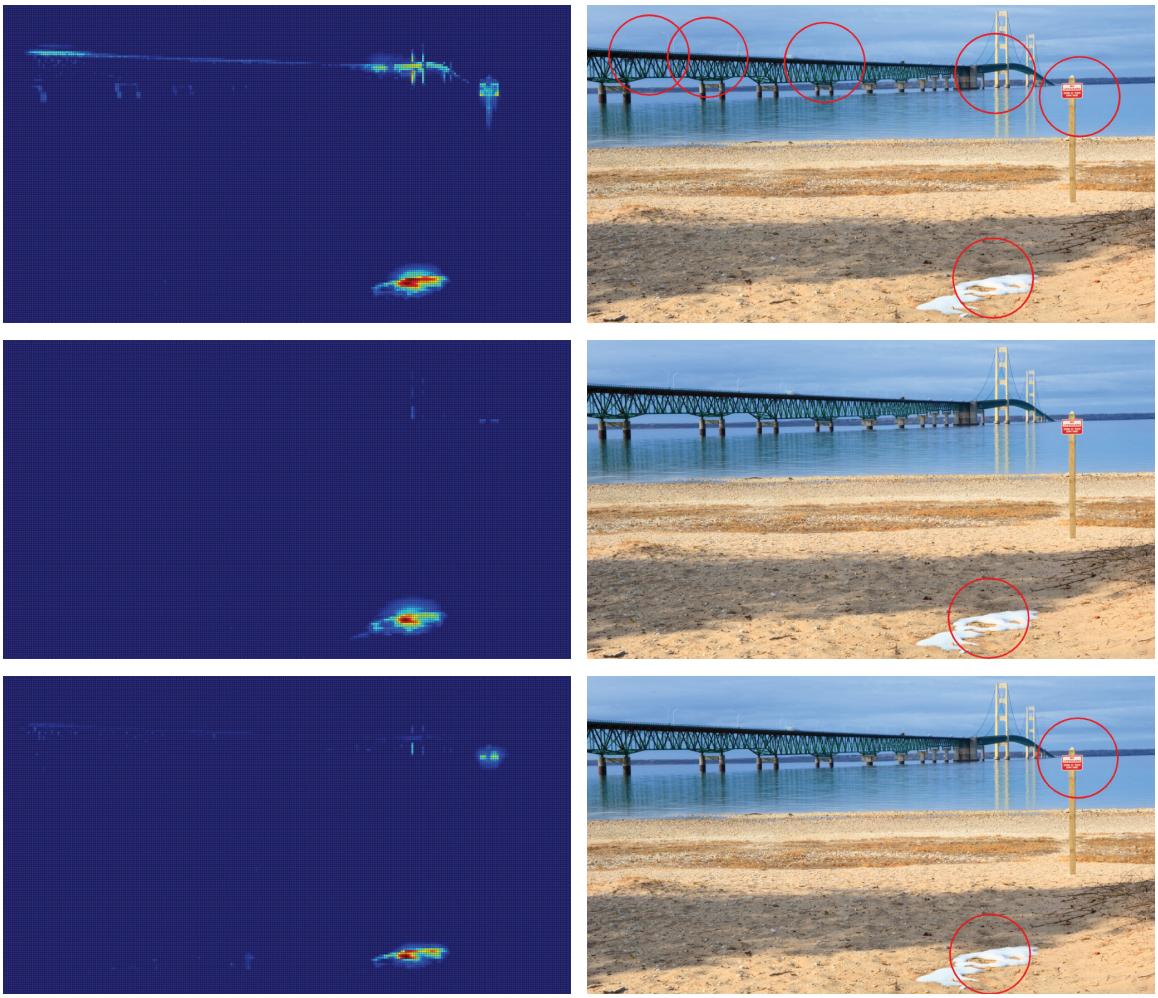


Figure 3.7: Comparison with other approaches on the HDR image "Mackinac Bridge". From top to down, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [46], and the method of Brémond et al. [13]. The HDR image is from Mark Fairchilds HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

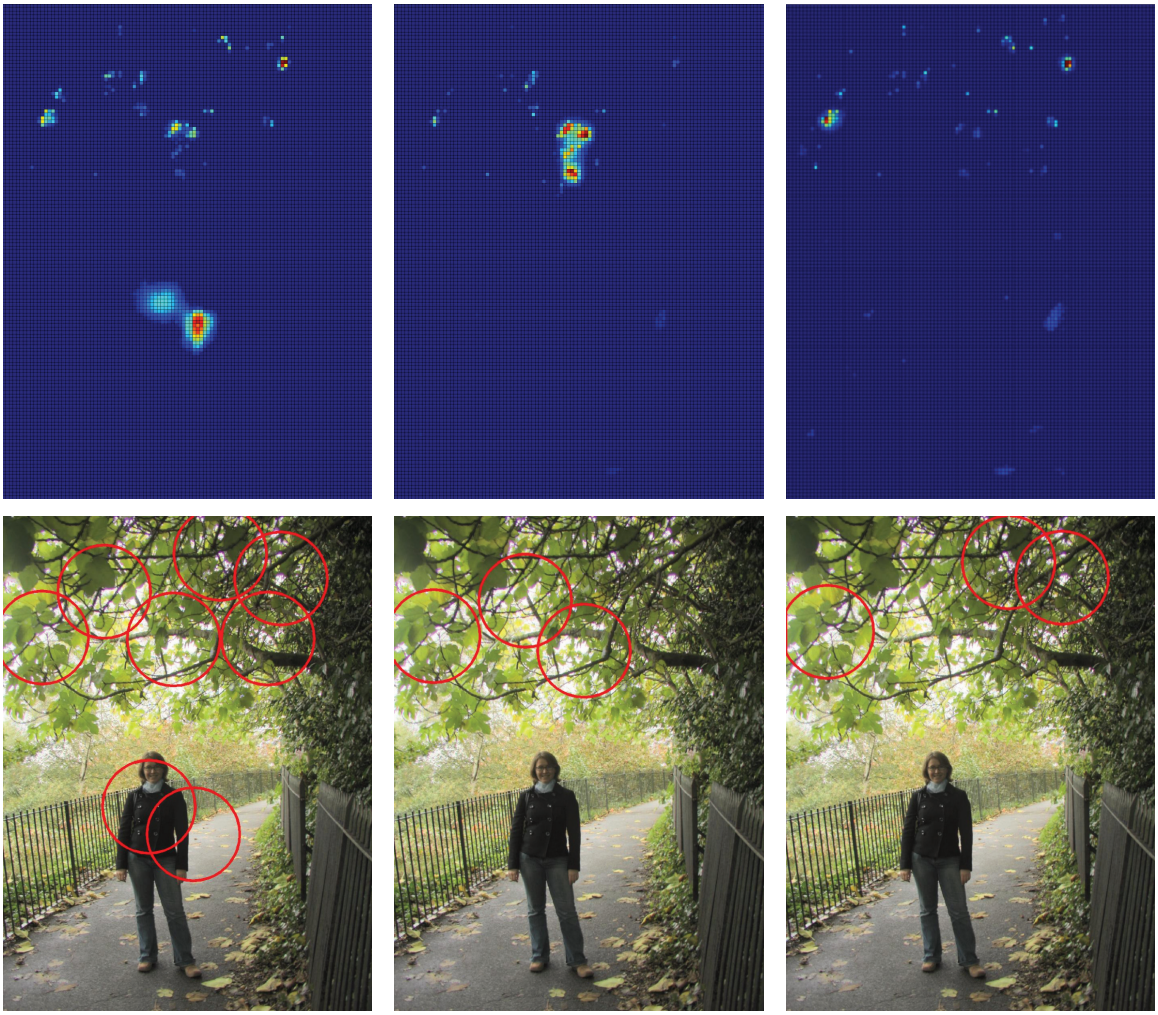


Figure 3.8: Comparison with other approaches on the HDR image "Celine1". From left to right, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [46], and the method of Brémond et al. [13]. The HDR image is from *High Dynamic Range Imaging*, published by Morgan Kaufmann Publishers, © 2006 Elsevier Inc.

# Chapter 4

# Objective Image Quality Assessment of Tone Mapped Images

## 4.1 Introduction

With the development of HDR tone mapping, it is essential to perform quality measures for tone mapped images. Without reliable quality evaluation, tone mapped images cannot be compared and further improvement of existing operators could be aimless. Although the quality assessment can be carried out by human observers in psychophysical experiments [21, 22, 55, 52], objective quality metrics are especially appealing as they are far more economic, faster, more consistent, and applicable in optimization frameworks. In this chapter, we focus on objective quality assessment of tone mapped images, which calculates the distortion between compressed LDR and original HDR images.

Conventional image quality assessment (IQA) algorithms which assume that reference images and test images have the same dynamic range is less suitable for quality evaluation of tone mapped images [102, 100]. These IQA algorithms compute distortion based on the difference of the magnitude of intensity or normalized contrast, and their predictions could be meaningless when the image pair has significantly different dynamic ranges. We assume that the quality evaluation of tone mapped images should focus on the reproduction of important image features that affect the perceived quality to human observers rather than the exact match of intensity or contrast.

The virtual photograph based algorithm bridges the gap of dynamic ranges for feature analysis, which enables the comparison of image features for quality evaluation. Using the algorithm for features analysis of HDR images, we propose two feature-based quality metrics for tone mapped images, including visual saliency distortion predictor (VSDP) [34, 35], and perceptual distortion predictor (PDP) [38, 32]. The quality metrics calculate the distortion in image features for quality prediction. A series of experiments demonstrates the effectiveness of the proposed feature-based quality metrics and their benefits over the prior works.

## 4.2 Visual Saliency Distortion Predictor

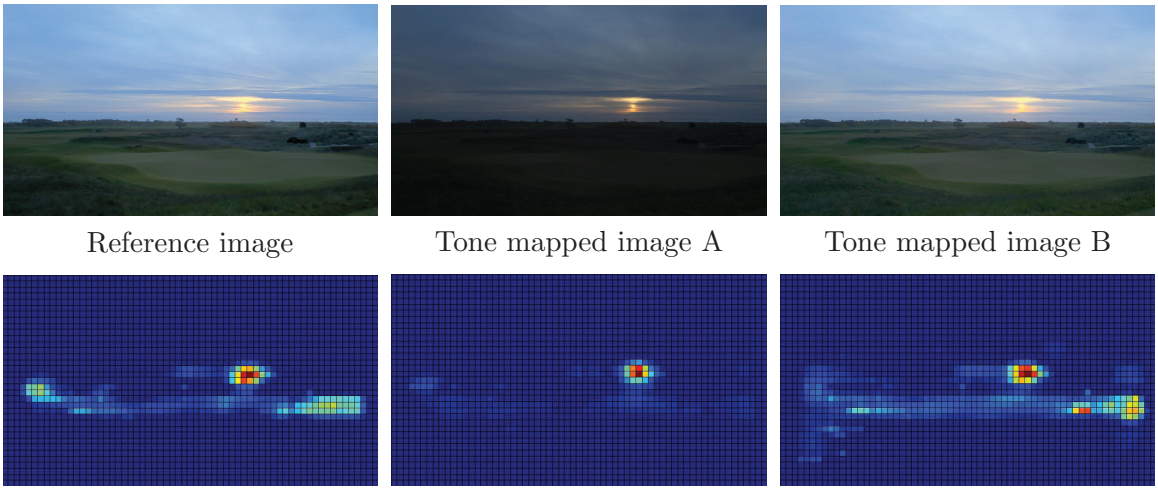
According to Reinhard et al. [86], the ultimate goal in tone mapping is to reproduce the visual appearance of HDR images. This raises the question: how we can evaluate the image quality of tone mapped images for visual realism? Based on eye-tracking data, Liu and Heynderickx [58] assert that visual attention affects how visual signals are received by the human visual system, and it would be beneficial for objective image quality assessment. Narwaria et al. [74] consider several TMOs to study the effects of tone mapping on visual attention. Their results indicate that visual attention needs to be considered for perceptual evaluation as human visual attention and fixation behaviour can be affected during tone mapping.

Inspired by these studies of visual attention [58, 74], we employ the bottom-up visual saliency calculated from computational visual attention models to measure image quality of tone mapped images, and propose an objective image quality metric called visual saliency distortion predictor (VSDP) [34, 35]. If tone mapped poorly, regions of interest (ROIs) in HDR images predicted by visual saliency may exhibit inaccurate contrast when displayed with low dynamic range and thus appear more or less salient. As shown in Figure 4.1 and Figure 4.2, different tone mapped versions of HDR images have varied saliency maps, and the tone mapped image sharing more similarity of its saliency map with that of the HDR image tends to provide a more faithful impression.

### 4.2.1 Visual Saliency Distortion

We have proposed a virtual photograph based algorithm for saliency analysis on HDR images. To do that, we employ camera curves to take multiple-exposed virtual photographs from HDR images, analyze visual saliency on the sequence of virtual photographs, and then combine the results into a unified map. The algorithm applies existing visual attention models to HDR images. In order to calculate visual saliency distortion, we use the virtual photograph based algorithm to generate saliency maps of HDR images, and the process is described as follows,

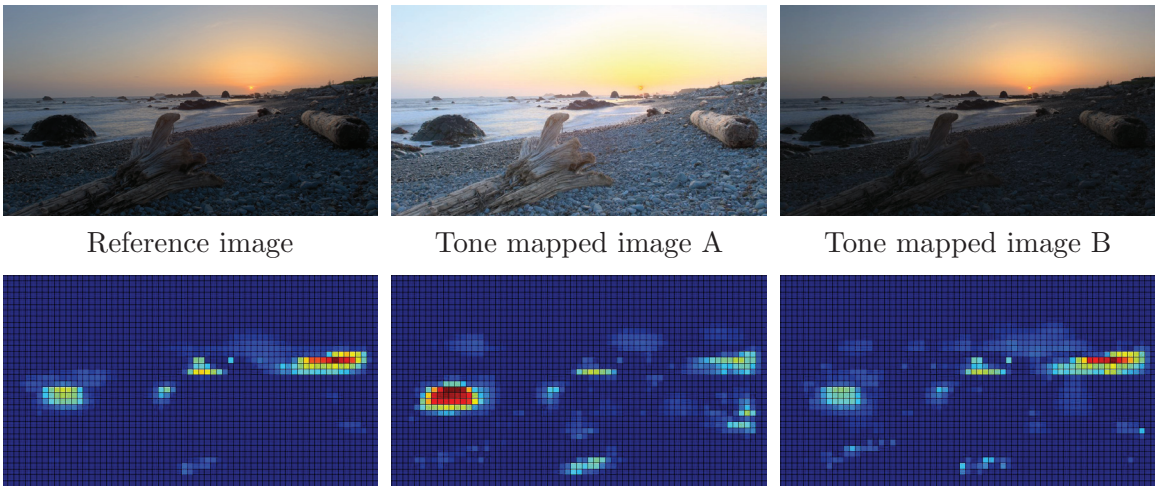
$$S_H(x,y) = \frac{1}{N}\sum_{i=1}^{N} w_i S_{Li}(x,y), \tag{4.1}$$



Reference image | Tone mapped image A | Tone mapped image B

Saliency map of HDR image | Saliency map of image A | Saliency map of image B

Figure 4.1: Application of visual saliency for quality evaluation on the HDR image "Bandon Sunset (1)". First column is the reference image tone mapped with manual manipulation and its saliency map generated by the virtual photograph based algorithm [33]. Second and third columns are tone mapped images from Schlick's operator [88] with various parameter settings and their saliency maps generated by the model of Itti and Koch [46]. The HDR image is from Mark Fairchild's HDR Photographic Survey © 2006-2007 Mark D. Fairchild.



Reference image | Tone mapped image A | Tone mapped image B

Saliency map of HDR image | Saliency map of image A | Saliency map of image B

Figure 4.2: Application of visual saliency for quality evaluation on the HDR image "Redwood Sunset". First column is the reference image tone mapped with manual manipulation and its saliency map generated by the virtual photograph based algorithm [33]. Second and third columns are tone mapped images from Schlick's operator [88] with various parameter settings and their saliency maps generated by the model of Itti and Koch [46]. The HDR image is from Mark Fairchild's HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

Here, $S_H(x, y)$ and $S_{Li}(x, y)$ denote the saliency map of HDR image and virtual photograph respectively, and the parameter $w_i$ depicts the weight of each individual map. It is worth noting that saliency maps of HDR images in this quality metric is generated by the precursor of the approach described in Chapter 3, in which we use a camera curve recovered from existing photograph sequences. Afterwards, we use the visual attention model of Itti and Koch [46] to compute the saliency map of tone mapped images $S_L(x, y)$. Adopting the same visual attention model for both HDR images and tone mapped images can better ensure an unbiased comparison.

With saliency maps of HDR and tone mapped LDR images in hand, we seek to quantify their difference which is referred to as visual saliency distortion. The normalized saliency map can be treated as a probability distribution that assigns a probability of visual attention to each pixel in an image. Therefore, visual saliency distortion can be measured by computing the information loss when the probability distribution calculated from a tone mapped image is used to approximate that of the HDR image. The Kullback-Leibler (K-L) divergence [53], or relative entropy, is used to quantify the information loss as

$$D_{KL}(H||L) = \sum_{x,y} \ln \left( \frac{S_H(x, y)}{S_L(x, y)} \right) S_H(x, y) \ . \tag{4.2}$$

Here, $D_{KL}(H||L)$ denotes the K-L divergence between the normalized saliency map of HDR image $S_H(x, y)$ and that of tone mapped image $S_L(x, y)$, where the summation extends over all pixels.

### 4.2.2 Experimental Results

The visual saliency distortion predictor (VSDP) was proposed for parameter tuning of TMOs [34, 35]. Before VSDP was published, Yeganeh and Wang [108] proposed the tone mapped image quality index (TMQI) by combining the measures of structural fidelity and statistical naturalness. Since TMQI can assign a single quality score to a entire image, it is potentially applicable for parameter optimization. Nevertheless, because the quality metric uses non-reference statistics of natural images to measure the quality of tone mapped images, it is too coarse to make accurate prediction for the breadth of real world images. More specifically, the statistical naturalness measure suggests that tone mapped images of dynamic range [0, 255] should have mean value


Figure 4.3: Comparison of image quality evaluation between VSDP and TMQI on the HDR images "Frontier" and "HDR Mark". First column: reference images tone mapped with manual manipulation. Second column: the best tone mapped images selected based on VSDP. Third column: the best tone mapped images selected based on TMQI. The HDR images are from Mark Fairchilds HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

around 116 and the standard deviation around 65 to appear natural. This measure can dominate TMQI and assign high scores to an "average" image, which will cause inaccuracy for quality evaluation of either naturally bright or dark scenes.

We have conducted an experiment to compare VSDP and TMQI for parameter optimization. In the experiment, we use 4 HDR images that cover different image contents and lighting conditions, including "Frontier", "HDR Mark", "Lab Booth", and "Blooming Gorse (1)". For each HDR image, we generate 25 visually different tone mapped images using Schlick's operator [88] with various parameter settings, and then apply both VSDP and TMQI to select the best tone mapped image. The reference images provided at [26] are used as benchmarks. Since these images are

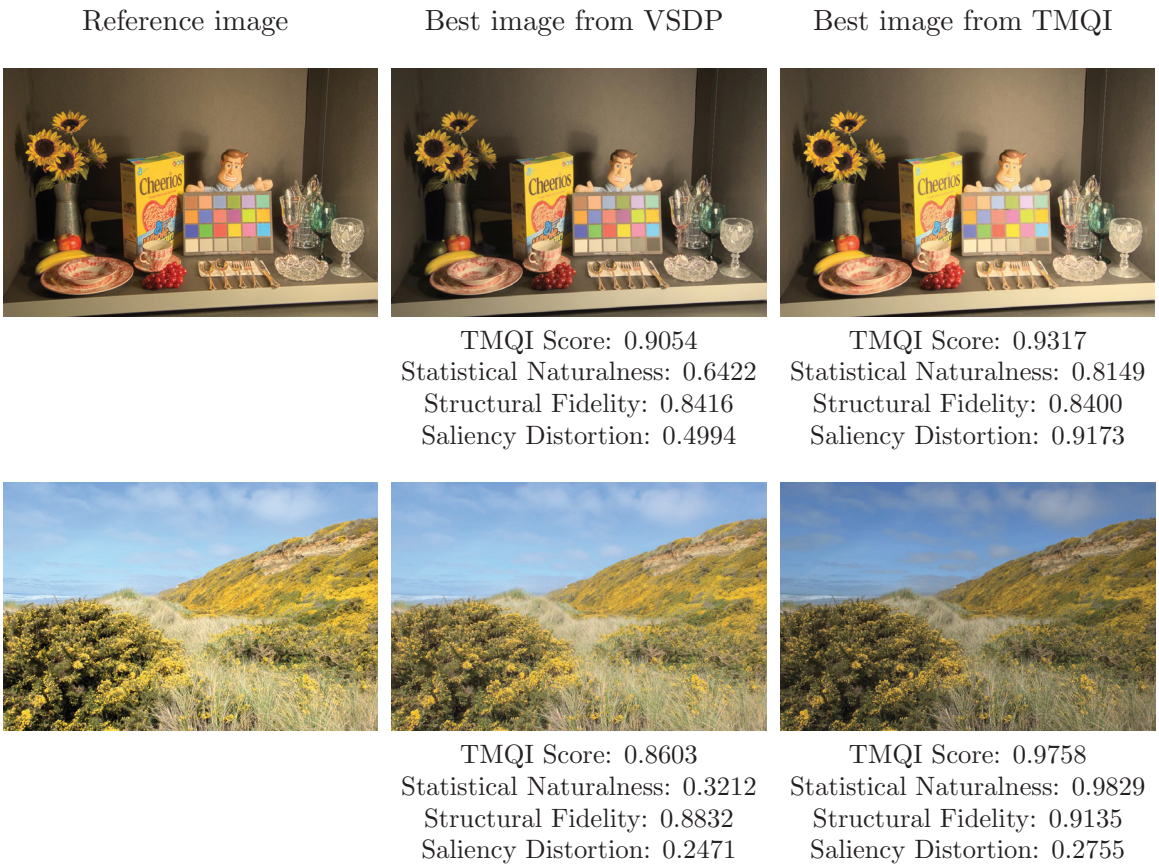



Figure 4.4: Comparison of image quality evaluation between VSDP and TMQI on the HDR images "Lab Booth" and "Blooming Gorse (1)". First column: reference images tone mapped with manual manipulation. Second column: the best tone mapped images selected based on VSDP. Third column: the best tone mapped images selected from TMQI. The HDR images are from Mark Fairchilds HDR Photographic Survey © 2006-2007 Mark D. Fairchild.

tone mapped with hand tuning of the tone-reproduction curve and saturation, they are expected to have more faithful appearance than other tone mapped versions.

Figure 4.3 illustrates the comparison on the HDR images "Frontier" and "HDR Mark". Those two examples are relatively dark scenes in both outdoor and indoor environments. As we can see, the tone mapped images selected by VSDP share similar appearance with reference images, while the images selected from TMQI are apparently too bright to faithfully reproduce these two scenes. And in Figure 4.4, we demonstrate the comparison on the HDR images "Lab Booth" and "Blooming Gorse (1)", where they depict an indoor moderate scene and an outdoor relatively bright scene respectively. VSDP yields consistently reliable predictions while TMQI does

not. Using the no-reference measure of statistical naturalness, TMQI cannot have accurate quality prediction except for the scenes with "average" brightness and contrast, making it not applicable for tone mapping optimization. A more comprehensive comparison of the quality metrics is provided in Chapter 5.

## 4.3 Perceptual Distortion Predictor

Although visual saliency has been employed in VSDP for quality evaluation, it is by no means the only image feature that can affect the perceived quality of tone mapped images, and other image features should be considered as well for more thorough evaluation. To do that, we further introduce a feature-based quality metric called perceptual distortion prediction (PDP) [38, 32], which extends the usage of virtual photograph based algorithm for multiple image features and incorporates them for the quality assessment.

In order to understand the perceived quality of tone mapped images to human observers, Drago et al. [22] conducted a series of psychophysical experiments, which found that subjective preference is most related to the apparent naturalness (the degree of resembling realistic scenes) and the apparent level of detail (the visibility of scene content). Also, in the work of Ledda et al. [55], the perceived quality of tone mapped images is measured in terms of overall similarity and detail reproduction. These perceptual studies [22, 55] suggest that the perceptual quality of tone mapped images can be explained with the global fidelity to the original HDR images and preservation of local detail.

In the quality metric, we use the feature-based measures, including brightness distortion, visual saliency distortion, and detail distortion in light and dark areas, to estimate the perceived image quality of tone mapped images. The brightness distortion and visual saliency distortion are related to the reproduction of perceived luminance and regions of interest, and they are used to measure the global fidelity to original HDR scenes. Due to the truncation of luminance during tone mapping, detail reproduction is mainly an issue in relatively light and dark areas [15]. Consequently, the detail distortion in light and dark areas are incorporated for the measurement of local detail preservation.

The framework of the perceptual distortion prediction is illustrated in Figure 4.5.

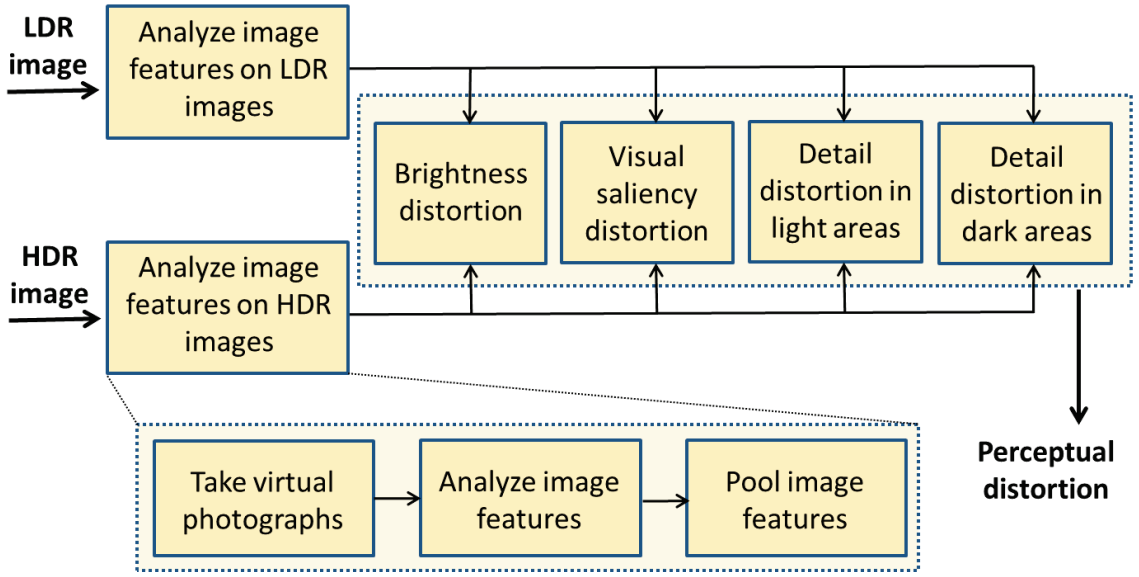



Figure 4.5: The framework of Perceptual Distortion Predictor

The quality metric takes an HDR image and a tone mapped image as the inputs and generate a distortion score for quality prediction. First of all, we take virtual photographs from the HDR scene, and analyze image features using the virtual photography sequence. After that, with the image features of both the HDR reference image and LDR test image, the quality of tone mapped images is assessed by combining the feature-based measures, including brightness distortion, visual saliency distortion, and detail distortion in light and dark areas.

### 4.3.1 Taking Virtual Photographs

We first employ the virtual photograph based algorithm for visual saliency analysis on HDR images. The virtual photographs bridge the gap of dynamic ranges for feature analysis, which enables the measure of feature distortion between HDR images and tone mapped images. In the quality metric, we use the virtual photograph sequence taken from HDR images to analyze multiple image features, including both global and local ones.

As described in Chapter 3, we calibrate HDR scenes with the key of a scene, and then take virtual photographs with the transfer curve from modern photography. We take a sequence of 11 virtual photographs from HDR images with exposure time

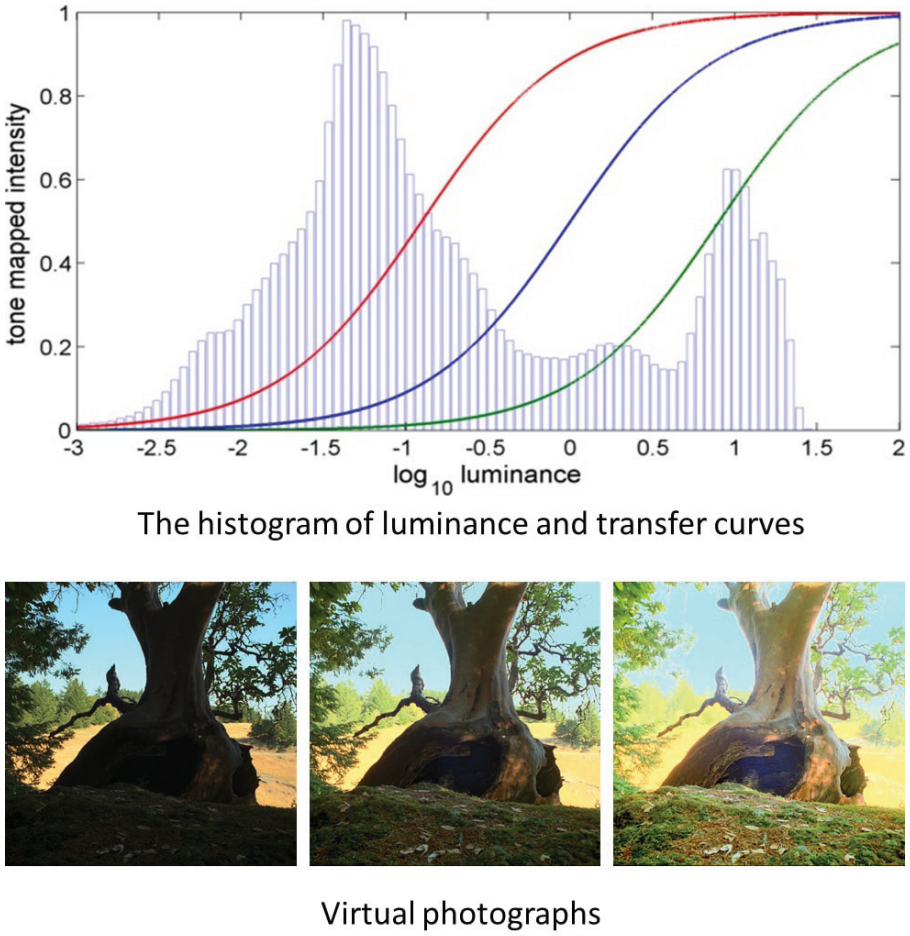



Figure 4.6: Transfer curves and virtual photographs of the HDR image "Tree". Upper row: the transfer curves with exposure time 1/8 (green), 1 (blue), and 8 (red). Lower row: virtual photographs with exposure time 1/8, 1, and 8 respectively. The HDR image is from Industrial Light & Magic.

$\Delta t \in \{1/30, 1/15, 1/8,\ 1/4, 1/2, 1, 2, 4, 8, 15, 30\}$. Examples of the curves and the virtual photographs are demonstrated in Figure 4.6. The figure on the upper shows the histogram and transfer curves with three different exposure settings. The other three images are the virtual photographs taken with the respective transfer curves. It is worth noting that the quality metric makes the quality prediction based on the distortion in image features and the similarity to the transfer curves cannot guarantee good quality of tone mapped images (see Appendix A for evidence).

### 4.3.2 Brightness Distortion

Brightness represents the visual perception of human beings for a particular luminance. In our quality metric, we aim to measure the overall brightness of HDR images as a quantity within the standard dynamic range, so a bright daylight scene and a dim night scene can be well distinguished. Similar overall brightness indicates faithful reproduction of illumination conditions, which can be crucial to the understanding of original HDR scenes.

Since the HDR images are calibrated with the key of a scene that correlates with the illumination of the original images, the virtual photograph taken without any exposure adjustment tends to produce more faithful brightness than others. Therefore, we estimate the brightness of an HDR image $B_H$ as the statistical average value of the virtual photograph that is directly generated from calibrated data ($\Delta t = 1$). Given a tone mapped image with average value $B_L$, the brightness distortion $B_{\text{dst}}$ is defined as the absolute difference

$$B_{\text{dst}} = |B_H - B_L| \ . \tag{4.3}$$

### 4.3.3 Visual Saliency Distortion

Visual saliency predicts the regions of interest that can attract human visual attention by their discriminative visual features from surrounding environment. In the problem of HDR tone mapping, consistency in visual saliency can ensure the regions of interest of HDR scenes are exhibited with accurate contrast when displayed with low dynamic ranges, which in turn produces an overall faithful impression [35].

As with the visual saliency distortion predictor (VSDP) [34, 35], we calculate the bottom-up visual saliency of virtual photographs with Itti and Koch's visual attention model [46], and combine the saliency maps into a unified map for HDR images by calculating their weighted average. The weights are scalar values based on the well-exposedness of virtual photographs, which decrease with an increasing degree of over- and under-exposure. The weight of a virtual photograph with exposure time $\Delta t$ is defined as $\exp(-(\log_2 \Delta t)^2/2\sigma^2)$ where $\sigma = 1.2$. With the saliency map of HDR image $S_H(x, y)$ and that of tone mapped image $S_L(x, y)$, the visual saliency distortion $S_{dst}$ is calculated as the Kullback-Leibler divergence between the saliency maps, which has been described in equation 4.2.

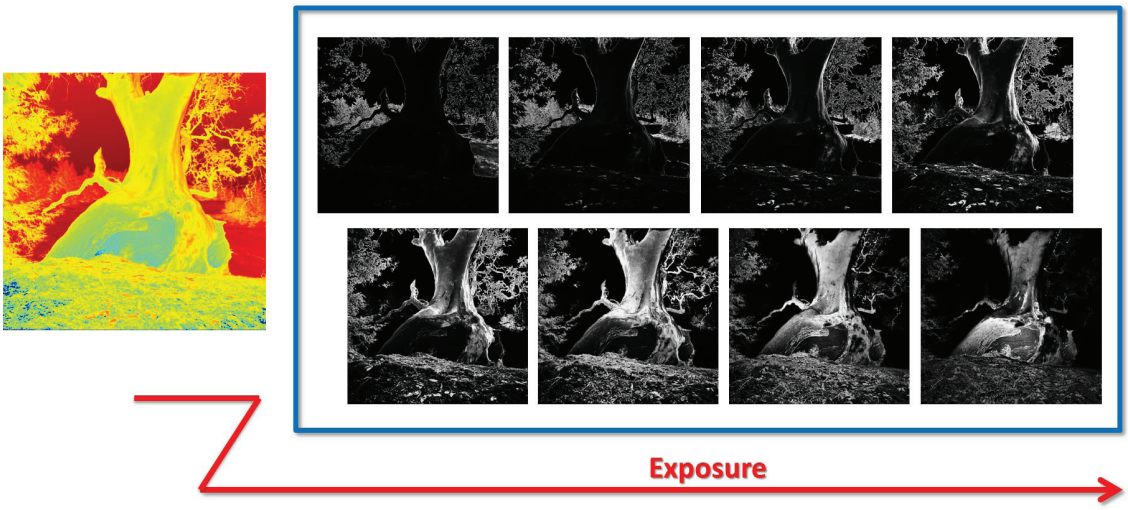



Figure 4.7: Illustration of weight maps of virtual photographs for detail analysis. The HDR image is from Industrial Light & Magic.

### 4.3.4 Detail Distortion in Light and Dark Areas

Preserving local detail during tone mapping is one of the most important concerns in quality evaluation [22, 55, 15] and algorithm design [24, 27, 5, 85]. Detail reproduction is mainly an issue in light and dark areas because the truncation of luminance in these areas can cause detail loss [15].

We calculate detail maps of virtual photographs with the Laplacian responses to high frequency detail, and then combine the detail maps into a unified map for HDR images. Since detail belongs to local features, a weight map will be generated for each virtual photograph based on the well-exposedness of pixel values. In our implementation, the weight function $\exp(-(p-M)^2/0.08)$ defined in [69] for exposure fusion is adopted, where $p$ denotes the normalized values and $M$ denotes middle grey in sRGB colour space. The function is applied for the three colour channels separately, and the results are multiplied to yield the weight of each pixel. Figure 4.7 illustrates the weight maps for the virtual photograph sequence with increasing exposures. We calculate the weighted average along each pixel to fuse the detail maps. The detail distortion in light areas $D^l_{\text{dst}}$ and dark areas $D^d_{\text{dst}}$ are computed as

Figure 4.8: Weight maps that represent light (left) and dark (right) areas. The HDR image is from Industrial Light & Magic.

$$D^l_{\text{dst}} = \sqrt{\sum_{x,y} \left(D_H(x,y) - D_L(x,y)\right)^2 W_l(x,y)}$$
$$D^d_{\text{dst}} = \sqrt{\sum_{x,y} \left(D_H(x,y) - D_L(x,y)\right)^2 W_d(x,y)} \ ,$$

where $D_H(x,y)$ and $D_L(x,y)$ are the detail maps of the HDR and tone mapped images, and $W_l$ and $W_d$ are weight maps that represent light and dark areas in the HDR scene (illustrated in Figure 4.8). We first estimate the weight maps according to the cumulative probability density of luminance. Then, we calculate the log average $L_{av}$ and the weighted log average using the estimated weight maps of bright and dark areas $L^l_{av}$ and $L^d_{av}$, and further adjust the weight maps by multiplying strength factors $\theta$ and $\phi$, where $\theta = \left((L^l_{av} - L_{av})/(L^d_{av} - L_{av})\right)^2$, and $\phi = 1/\theta$, which means that higher weights will be assigned to the bright or dark areas with more conspicuous luminance (higher degree of being bright or dark).

### 4.3.5 Overall Perceptual Distortion

For the measure of overall distortion, we combine the normalized distortion measures together and assign a single quality score for each tone mapped image. Since $B_{\text{dst}}$ is already within the range of $[0,1]$, we only need to address $S_{\text{dst}}$, $D^l_{\text{dst}}$, and $D^d_{\text{dst}}$. Because it is difficult to estimated outside boundary of distortion values, we adopt the scaling function $f(x) = x/(x+c)$ to compress the large margin values, where

$c$ is a parameter that can be estimated from the distribution of quality predictions. By applying the feature-based measures on 160 image pairs generated from sixteen HDR images and ten TMOs, we determine the parameters to ensure similar standard deviation for all the distortion measures. In our implementation, $c$ is set to 10 for visual saliency distortion and 20 for detail distortion. With the normalized distortion measures, the overall perceptual distortion is calculated as

$$O_{\text{dst}} = w_1 B_{\text{dst}} + w_2 S_{\text{dst}} + w_3 D^l_{\text{dst}} + w_4 D^d_{\text{dst}} \ , \tag{4.4}$$

where $w_i, i \in \{1, 2, 3, 4\}$, denote the weights of brightness distortion, visual saliency distortion, and detail distortion in light and dark areas, respectively. Without further evidence that describes the interaction of feature-based measures for quality assessment, we assign the same weights 0.25 to different quality measures by emphasizing their equal importance. A similar pooling strategy have been adopted in TMQI-II [61].

### 4.3.6 Experimental Results

We test the proposed quality metric on a broad range of HDR scenes and tone mapped images from different TMOs and parameter settings. In the test cases, we confirmed that the common artifacts of tone mapped images that lead to quality reduction can be well predicted by the feature-based measures.

Figure 4.9 illustrates the brightness artifacts which happen frequently during tone mapping. In this example, the tone mapped images (middle and right) are apparently either too dark or bright to depict the daylight scene, and the artifacts can be detected by the measure of brightness distortion. Figure 4.10 shows example of contrast artifacts of tone mapped images. These artifacts are usually caused by an inappropriate operation of local adaption, and they may cause structural changes, making the tone mapped images appear "unreal" or "unnatural". As shown in the figure, the tone mapped image with relatively lower visual saliency distortion (left) tends to have more faithful and natural appearance than others. An example of detail artifacts is provided in Figure 4.11. Compared with the left tone mapped images, the other two images (middle and right) suffer from detail loss in either light or dark regions (the color charts inside and outside of the box), and the perceptual difference

Brightness Distortion: 0.0409 Brightness Distortion: 0.2907 Brightness Distortion: 0.1787

Figure 4.9: Image quality assessment using brightness distortion on the HDR image "C08". The images are generated from various TMOs with some manipulation of parameter settings. The HDR image is from H. Nemeto et al. [76].

can be predicted by corresponding detail distortion measures. For instance, because of over and under-exposure in light and dark areas respectively, the middle image has higher distortion values in these areas.

These examples give a sense of how the distortion measure operates. For a systematic validation, we compare the objective quality assessment results with subjective evaluation data. The experiment is carried out with the subject-rated image database introduced by Yeganeh and Wang [108], which is the largest database for the evaluation of objective quality metrics of tone mapped images. The database consists of 15 image sets and 120 tone mapped images with subjective rankings from 20 human observers. The tone mapped images are generated using five TMOs developed by Reinhard et al [82], Drago et al. [23], Durand and Dorsey [24], Mantiuk et al. [64], and Pattanaik et al. [79], and three built-in operators in Adobe Photoshop, namely "Exposure and Gamma", "Equalize Histogram", and "Local Adaption". In each test, the subjects are asked to rank the 8 tone mapped images in each image set from the best to the worst, and the subjective ranking for each image is calculated as the mean ranking score of subjects. To quantify the correlation between the objective and subjective ranks, two statistical metrics: Spearman's rank correlation coefficient (SRCC) [71] and Kendall's rank correlation coefficient (KRCC) [48], have been used. The metrics are non-parametric measures of ranking correlation whose value increases with the increasing similarity of ranks (the value is 1 for identical ranks).

Spearman's rank correlation coefficient (SRCC) is defined as follows,

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{4.5}$$

where $N$ represents the numbers of observations in the ranks, and $d_i$ denotes the

Saliency Distortion: 0.0154 | Saliency Distortion: 0.2327 | Saliency Distortion: 0.1700

Figure 4.10: Image quality assessment using visual saliency distortion on the HDR image "C10". The images are generated from various TMOs with some manipulation of parameter settings. The HDR image is from H. Nemeto et al. [76].

Detail Distortion in Light Areas: 0.4479 Detail Distortion in Dark Areas: 0.2596 | Detail Distortion in Light Areas: 0.6195 Detail Distortion in Dark Areas: 0.3415 | Detail Distortion in Light Areas: 0.4842 Detail Distortion in Dark Areas: 0.2840

Figure 4.11: Image quality assessment using detail distortion in light and dark areas on the HDR image "Test Chart". The images are generated from various TMOs with some manipulation of parameter settings. The HDR image is from Empa Media Technology.

difference between two ranks of each observation.

Kendall's rank correlation coefficient (KRCC) is defined as follows,

$$KRCC = \frac{N_c - N_d}{N(N-1)/2}. \tag{4.6}$$

where $N_c$ and $N_d$ are the number of concordant pairs (the pairs with consistent rank order) and discordant pairs (the pairs with inconsistent rank order) respectively.

In the experiment, we generate the rankings from the objective quality metrics for all fifteen image sets, and then compare the generated objective predictions with the subjective results provided in the database, where the similarity between objective and subjective ranks is measured in terms of SRCC and KRCC. The alternative full-reference methods are used as the benchmark for comparison, including the visual saliency distortion predictor (VSDP) [35], and the improved variant of Tone Mapped Image Quality Index (TMQI-II) [61], whose implementation is provided by the original

Table 4.1: Comparison with the alternative objective quality metrics (VSDP and TMQI-II) and the mean behavior of individual subject.

| **Test data** | **SRCC** | | | **KRCC** | | |
|---|---|---|---|---|---|---|
| | **VSDP** | **TMQI-II** | **Subject** | **VSDP** | **TMQI-II** | **Subject** |
| Positive Ranks | 10 | 12 | 10 | 10 | 9 | 9 |
| Negative Ranks | 2 | 1 | 5 | 2 | 1 | 6 |
| Ties | 3 | 2 | 0 | 3 | 5 | 0 |
| **p Value** | **0.020** | **0.002** | **0.363** | **0.016** | **0.010** | **0.650** |

authors. We do not use Aydın et al.'s method [7] because it is not obvious how the quality maps could be condensed into a single score for rank prediction. In addition to the alternative quality metrics, the mean behaviour of subjects is also incorporated for comparison (the benchmark is provided in the original database). To generate that, the subjective ranks from individual subjects are compared with the subjective ranks, and the mean performance of all individual subjects is computed accordingly [108]. Table 4.2 lists the performance of VSDP, the performance of TMQI-II, the mean performance of individual subjects, and the performance of our method in terms of both SRCC and KRCC. Based on the average values, our method produces better results than the alternative metrics and the mean behavior of individual subjects.

To further analyze the difference, we carry out a paired test between the performance of our method and those of the other approaches. A Wilcoxon signed-rank test is used to compare the matched samples and assess whether they are significantly different from each other. Table 4.1 shows the outcomes of the paired test, including the positive ranks (the number of image sets in which our method performs better), the negative ranks (the number of image sets in which the alternatives perform better), the ties (the number of image sets with the same performance), and the asymptotic significance (two-tailed), which is also known as $p$ value. If the $p$ value is less than 0.05, the test groups are deemed to have a statistically significant difference. It can be seen that, according to both SRCC and KRCC, our method can perform significantly better than both VSDP and TMQI-II, and it performs similarly with mean behavior of individual subjects.

Finally, Figure 4.12 illustrates the performance in terms of SRCC of the individual feature-based measures and the overall perceptual measure. As we can see,

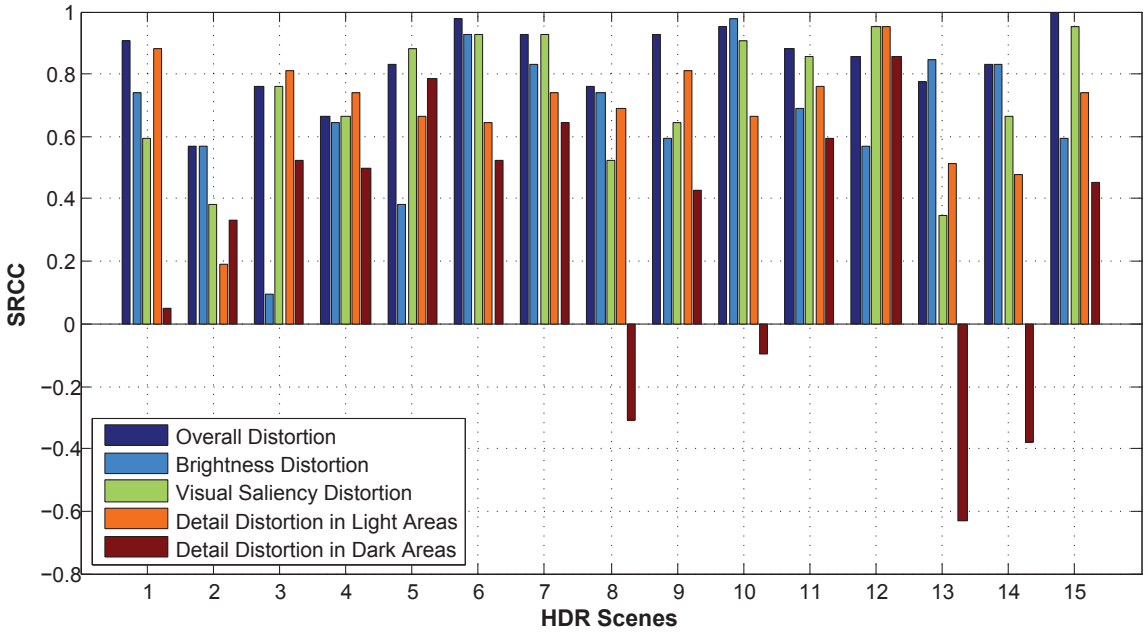



Figure 4.12: SRCC performance of the distortions in various image features and the overall perceptual distortion.

an individual feature-based measure cannot guarantee good quality predictions for different HDR images, and the overall perceptual measure leverages the strengths of each considered measure for better performance.

Table 4.2: Performance evaluation using subjective database [108]. For the 15 image sets in the database, the performance of VSDP, the performance of TMQI-II, the mean performance of individual subject, and the performance of our method are listed in terms of both SRCC and KRCC.

| **Image Set** | **SRCC** | | | | **KRCC** | | | |
|---|---|---|---|---|---|---|---|---|
| | **VSDP** | **TMQI-II** | **Subject** | **Our Method** | **VSDP** | **TMQI-II** | **Subject** | **Our Method** |
| 1 | 0.5952 | 0.9048 | 0.9071 | 0.9048 | 0.4286 | 0.7857 | 0.8071 | 0.7857 |
| 2 | 0.3810 | 0.5000 | 0.8251 | 0.5714 | 0.2857 | 0.2857 | 0.7269 | 0.4286 |
| 3 | 0.7619 | 0.6905 | 0.8797 | 0.7619 | 0.5714 | 0.5714 | 0.7642 | 0.5714 |
| 4 | 0.6667 | 0.6905 | 0.9130 | 0.6667 | 0.4286 | 0.5000 | 0.8107 | 0.4286 |
| 5 | 0.8810 | 0.6667 | 0.6000 | 0.8333 | 0.7857 | 0.5000 | 0.4714 | 0.7143 |
| 6 | 0.9286 | 0.9762 | 0.7630 | 0.9762 | 0.8571 | 0.9286 | 0.6464 | 0.9286 |
| 7 | 0.9286 | 0.8333 | 0.8285 | 0.9286 | 0.8571 | 0.7143 | 0.7250 | 0.8571 |
| 8 | 0.5238 | 0.6667 | 0.8023 | 0.7619 | 0.2857 | 0.5000 | 0.7000 | 0.5714 |
| 9 | 0.6429 | 0.8095 | 0.7857 | 0.9286 | 0.5000 | 0.7143 | 0.6607 | 0.8571 |
| 10 | 0.9048 | 0.9048 | 0.9276 | 0.9524 | 0.7857 | 0.7857 | 0.8418 | 0.8571 |
| 11 | 0.8571 | 0.8333 | 0.8523 | 0.8810 | 0.6429 | 0.7143 | 0.7428 | 0.7143 |
| 12 | 0.9524 | 0.5952 | 0.7595 | 0.8571 | 0.8571 | 0.4286 | 0.6250 | 0.7143 |
| 13 | 0.3452 | 0.7500 | 0.6970 | 0.7738 | 0.1786 | 0.6071 | 0.5637 | 0.6071 |
| 14 | 0.6667 | 0.7619 | 0.7702 | 0.8333 | 0.4286 | 0.5714 | 0.6214 | 0.6429 |
| 15 | 0.9524 | 0.9048 | 0.9035 | 1.0000 | 0.8571 | 0.7857 | 0.8142 | 1.0000 |
| **average** | **0.7325** | **0.7659** | **0.8143** | **0.8421** | **0.5833** | **0.6262** | **0.7014** | **0.7119** |

## 4.4 Conclusion and Discussion

Using virtual photograph based analysis for feature extraction and comparison of image pairs with different dynamic ranges, we have proposed two full-reference quality metrics for tone mapped images. The fundamental principle of our metric is that the quality evaluation of tone mapped images should focus on the reproduction of important image features that are related to the quality as perceived by human observers. Subjective and numerical experiments have been used to verify the effectiveness of our method and its benefits over prior works. The feature-based quality metrics can be applied for comparison of TMOs and automated tone mapping optimization.

The proposed feature-based quality metrics suffer from some limitations which could be addressed or improved in the future. First of all, the quality metrics mainly focus on the luminance channel, making it less effective to measure color reproduction. The chromatic information has been used for visual saliency analysis, but the operation is limited to the central-surround difference across several color pairs. Considering most HDR scenes are captured in color, we would like to further investigate the quality assessment of color reproduction in the future, and incorporate the measures in the current quality metrics for more comprehensive evaluation. Secondly, in the absence of further evidence, the perceptual quality predictor adopts a rather simple pooling strategy by emphasizing the same importance of different quality measures. It would be useful to perform a deeper study involving the relationship of the feature-based measures and how they determine the image quality of any particular HDR image in finer detail.

# Chapter 5

# Tone Mapping Optimization

## 5.1 Introduction

A large number of TMOs have been proposed and many of them depend on user parameters that impact the quality of the tone mapped images. Without human guidance, it is difficult to select an appropriate operator and parameter settings for a particular HDR image. The development of objective image quality measures for tone mapped images opens up the possibility of generating optimized tone mapped images without the need for user input. In this chapter, we employ the proposed feature-based quality metrics to solve the tone mapping optimization problems including parameter tuning of TMOs [34, 35] and blended tone mapping [36]. The tone mapping optimization algorithms are implemented with an evolutionary strategy which iteratively searches for the optimal solution in the parameter space of operators.

## 5.2 Parameter Tuning of Tone Mapping

Many TMOs provide several parameters for further adjustment [86]. Since the optimal parameters settings could be image-dependent, the default settings cannot work well for any arbitrary HDR image, and manual adjustments are required to achieve satisfying results. However, parameter tweaking is a trial-and-error process that can be tedious and time-consuming even for knowledgeable users with complete understanding of the parameters.

A number of methods have been proposed with the attempts to fine-tune parameters. Based on subjective experiments, Yoshida et al. [109] determine two generic parameters for a global operator that can be partly estimated from image characteristics. Also, by minimizing the visible contrast distortion between HDR images and tone-mapped images displayed on a particular display device, Mantiuk et al. [65]

address the tone mapping problem as an optimization problem in which a set of parameters that define a linear tone-curve is automatically adjusted. Nevertheless, these methods are designed for particular parameters, and they cannot easily be adapted for arbitrary parameters of other operators.

We seek to provide a universal solution for parameter tuning of TMOs without the requirement for user interaction. Based on the proposed visual saliency distortion predictor (VSDP), we employ the bottom-up visual saliency to measure image quality for parameter tuning [34, 35]. If tone mapped poorly, regions of interest in HDR images predicted by visual saliency may exhibit inaccurate contrast when displayed with low dynamic range and thus appear more or less salient. Visual saliency based parameter tuning helps to ensure that regions of interest will be reproduced with more accurate contrast. With the visual saliency distortion calculated by VSDP, parameter tuning can be treated as an optimization problem. We use an evolutionary algorithm to iteratively find parameter settings to minimize distortion.

### 5.2.1 Search Space

The search space would be the parameter space of the selected operator. We choose three TMOs are selected to validate the parameter tuning algorithm: Schlick's operator [88], Ashikhmin's operator [5], and Reinhard and Devlin's operator [83]. Schlick's operator is a simple, yet effective tone mapping algorithm that can be used to generate plausible results by tuning user parameters. Ashikhmin's operator tends to preserve local details during tone mapping with user parameters provided for further adjustment. Similarly, Reinhard and Devlin's operator is derived from the photoreceptor mechanisms of the human visual system, and it provides user parameters for the control of overall intensity, as well as light and chromatic adaptation. Ashikhmin's operator is local, while the other two operators are global. The three TMOs are commonly used for subjective evaluation [15, 10], interactive tone mapping [16], and in end user tools, such as *Luminance HDR*[1].

We use the implementation of the operators provided by Reinhard et al. [86]. Table 5.1 lists the default settings, ranges, and short descriptions of the parameters. The default value for gamma correction and the pre-scale factor are from the

[1] `qtpfsgui.sourceforge.net`

Table 5.1: Default parameter values and ranges of TMOs.

| **Parameter** | **Default** | **Range** | **Description** |
|---|---|---|---|
| Schlick | | | |
| $n$ | 0.18 | $[0.0, 1.0]$ | just noticeable difference |
| $\gamma$ | 2.0 | $[1.5, 2.5]$ | gamma correction |
| Ashikhmin | | | |
| $\log_{10}(p)$ | -1.0 | $[-2.0, -4.0]$ | pre-scale factor |
| $t$ | 0.5 | $[0.0, 1.0]$ | threshold for scale |
| $\gamma$ | 2.0 | $[1.5, 2.5]$ | gamma correction |
| Reinhard and Devlin | | | |
| $f$ | 0.0 | $[-8.0, 8.0]$ | scale factor |
| $m$ | *[1] | $[0.3, 1.0]$ | exponent for semi-saturation |
| $a$ | 1.0 | $[0.0, 1.0]$ | light adaption |
| $c$ | 0.0 | $[0.0, 1.0]$ | chromatic adaption |
| $\gamma$ | 2.0 | $[1.5, 2.5]$ | gamma correction |

implementation, and the default values for other parameters are derived from the recommendations of the original papers. It is worth mentioning that we have the exponent for semi-saturation and chromatic adaption parameter remain unchanged for Reinhard and Devlin's operator. The exponent for semi-saturation is not included in the optimization as its value can be well estimated by an algorithm recommended in the original paper. We found out adding the exponent into parameter tuning cannot contribute to extra gain in visual saliency distortion. As for the chromatic adaption parameter, we use the default value because the measure of visual saliency mainly focuses on luminance channel, making it less effective for the color related parameters.

### 5.2.2 Minimization of Visual Saliency Distortion

The optimization method used for minimizing visual saliency distortion is a $(1 + \lambda)$ evolution strategy (see Chapter 2 for the related terminology). Evolution strategies (ESs) are a sub-class of nature-inspired search methods which have mutation, recombination, and selection applied for the population of individuals that contain candidate solutions. They are mostly useful for optimization problems where derivative information is not available and relatively robust with regard to the ruggedness

[1]The default value of $m$ is estimated by an algorithm presented by Reinhard et al. [86].

**Algorithm 2** Parameter Tuning of TMOs

Set the generation counter $t = 0$;
Initialize the strategy parameters $\lambda$, $\sigma$;
Create the population;
Conduct tone mapping with the operator;
Calculate visual saliency distortion with VSDP;
**while** iteration number is smaller than the threshold **do**
  **for** $i \in \{1, ..., \lambda\}$ **do**
    Generate offspring with mutation;
    Conduct tone mapping with the operator;
    Calculate visual saliency distortion with VSDP;
  Update the population with generated offspring;
  Select the individual with the least saliency distortion;
  Update the step size $\sigma$;
  $t = t + 1$;
Return the optimal result;

of the objective function. We adopt the ESs here because the distortion measure of tone mapped images can only be calculated through application of TMOs and no convenient mathematical assumptions regarding the objective can be made. Moreover, ESs are relatively easy to implement and they are often capable of generating substantial improvements in a small number of iterations.

ESs use real-value vectors to encode an individual that represents candidate solution. For the problem of parameter tuning, the candidate solutions $\mathbf{x} \in \mathbb{R}^n$ are comprised of parameters of a TMO with their ranges normalized to a unit width. In iteration $t$ of the algorithm, $\lambda > 1$ offspring are generated from the parent $\mathbf{x}^{(t)}$ as

$$\mathbf{y}_i^{(t)} = \mathbf{x}^{(t)} + \sigma^{(t)}\mathbf{z}_i^{(t)} \qquad i = 1, \ldots, \lambda \tag{5.1}$$

where $\sigma^{(t)}$ denotes the step size and $\mathbf{z}_i^{(t)} \in \mathbb{R}^n$ is a standard normally distributed mutation vector. The step size $\sigma^{(t)}$ controls the expected distance of the offspring from their parent and is initialized to $1/3$. Out-of-range values of variables are clamped to $[0, 1]$. Then, the candidate solution that leads to minimal visual saliency distortion is selected and adopted as the parent $\mathbf{x}^{(t+1)}$ for the next iteration. If the parent candidate solution is superior to all of its offspring, then it is not replaced in iteration $t$. As with

Chisholm et al. [16], we update the step size in subsequent iterations where the parent is superior to all of its offspring by multiplying it with an adaption factor $s$ which is set to 0.8. The evolutionary algorithm typically requires relatively few iterations to obtain a result with significantly reduced visual saliency distortion. The optimization process is described in Algorithm 2.

### 5.2.3 Experimental Results

In our implementation, the number $\lambda$ of offspring in each iteration is set to 8, and the number of iterations to 7. Based on preliminary tests, these settings can yield good results for the tested HDR images. Unless indicated otherwise, the initial values of user parameters are randomly selected from within their ranges. The HDR images in the database provided by Fairchild [26] are used as test images. The database contains 105 HDR images with a broad range of luminance values, as well as the corresponding reference images manually tone mapped for appearance.

Figures 5.2, 5.3, 5.4, 5.5, and 5.6 demonstrate the results of our parameter tuning algorithm on five HDR images. Compared with the images tone mapped with default parameter settings, our results generally have closer visual appearance to the reference images. An example of the evolution of saliency distortion is shown in Figure 5.1, demonstrating that our approach can reduce the saliency distortion significantly within a small number of iterations. However, we do notice differences between the reference images and the tone mapped images with optimized parameter settings. For manual tone mapping, there is flexibility for manipulation, such as changing the tone mapping curve directly to show more details, and adjusting the saturation for a more vivid look. In parameter tuning, we can only tune the user parameters provided by the original authors, and the results suffer from the limitations of the particular TMO used. Therefore, parameter tuning can generate tone mapped images that have similar visual appearance with reference images, but it cannot yield identical results in most cases.

The computational cost of our algorithm is dominated by the cost of tone mapping. For the computation of saliency distortion, our MATLAB implementation takes around 0.5 seconds for a 1028×577 pixel image running on an Intel Quad-Core 2.66GHz CPU with 4 GB of RAM. As each iteration requires several independent

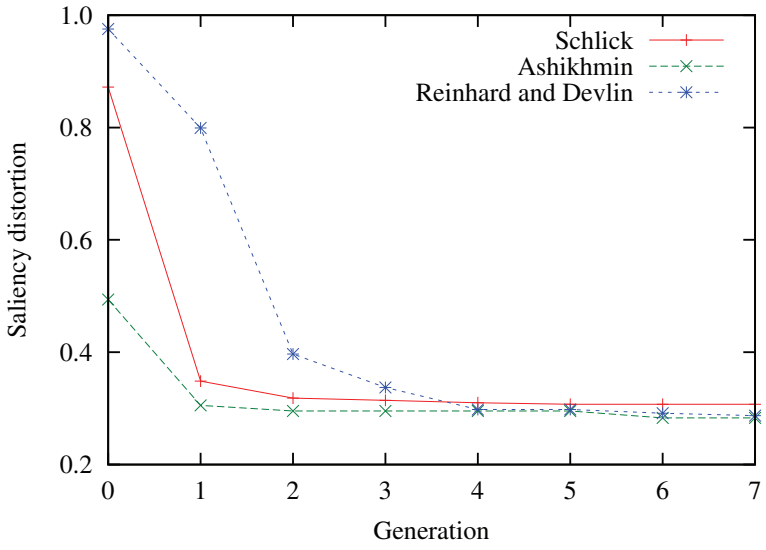



Figure 5.1: Evolution of visual saliency distortion in sample parameter tuning runs for "Bandon Sunset (1)".

applications of the TMO with different parameter settings, multiple cores can easily be exploited.

### 5.2.4 Systematic Analysis

For a thorough evaluation, we apply the parameter tuning algorithm to all images in the database [26]. Since reference images are available, the performance of optimized and default parameter settings can be quantified with full-reference image quality assessment metrics. Two widely used metrics are chosen to evaluate the performance: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [103, 101]. PSNR is a quantity related to mean squared error (MSE), which measures the average squared luminance difference between the reference and test images. The SSIM index by Wang et al. [103] has been proposed as an alternative approach for quality assessment of image similarity. The SSIM index is able to capture the distortion of image structure, which is asserted to be crucial for human visual perception [103]. In our test, we adopt both saliency distortion and TMQI in the optimization framework to search for optimal results. Then, by comparing with the reference images, the performance of the optimized and default parameter settings are calculated in terms of PSNR and SSIM index. Considering the stochastic nature of the search, we

Table 5.2: Comparison between visual saliency based optimized parameter settings and default parameter settings.

| TMO | PSNR | | | SSIM | | |
|---|---|---|---|---|---|---|
| | Optimized | Default | $p$ Value | Optimized | Default | $p$ Value |
| Schlick | 56 | 49 | 0.315 | 61 | 44 | 0.001 |
| Ashikhmin | 105 | 0 | 0.000 | 86 | 19 | 0.000 |
| Reinhard & Devlin | 105 | 0 | 0.000 | 82 | 23 | 0.000 |

cannot guarantee the exact same output from each parameter tuning process. We thus conduct five independent runs for each image in the database.

Three experiments[2] are carried out to analyze the performance of the proposed approach and compare VSDP with TMQI for parameter tuning. In the first one, we conduct a paired difference test between the average performance of optimized parameter settings and the performance of default settings to measure the improvement of saliency based parameter tuning. A Wilcoxon signed-rank test is adopted to compare the matched samples and assess whether they differ from each other. Table 5.2 shows the numbers of images for which optimized parameter settings perform better than default settings and vice versa, and the asymptotic significance (two-tailed), which is also known as $p$ value. If the $p$ value is less than 0.05, the test groups are deemed to have statistically significant difference. According to PSNR, optimized parameter settings perform significantly better than default parameter settings for Ashikhmin's and Reinhard and Devlin's operators, but not for Schlick's operator. Based on the SSIM index, optimized parameter settings perform better than default settings for all three operators.

In the second experiment, we perform a comparison for various TMOs using a Friedman test. The Friedman test ranks the performance of operators for each image, and calculates the mean ranks to detect differences. The mean ranks of each operator and the $p$ value are given in Table 5.3, which indicates that the ranks of operators can be greatly changed with parameter tuning. Since higher values of mean ranks imply better performance, Schlick's operator is preferable to the other two operators for

[2]Some other experiments are provided in Appendix B. They are used to prove that the virtual photograph based method used in parameter optimization cannot be simply replaced by computing visual saliency on log-HDR values.

Table 5.3: Comparison among TMOs.

| TMO | PSNR | | SSIM | |
|---|---|---|---|---|
| | Optimized | Default | Optimized | Default |
| Schlick | 1.79 | 3.00 | 2.14 | 2.36 |
| Ashikhmin | 2.21 | 1.30 | 1.94 | 1.58 |
| Reinhard & Devlin | 2.00 | 1.70 | 1.91 | 2.06 |
| $p$ value | 0.010 | 0.000 | 0.196 | 0.000 |

Table 5.4: Comparison between VSDP and TMQI for parameter tuning.

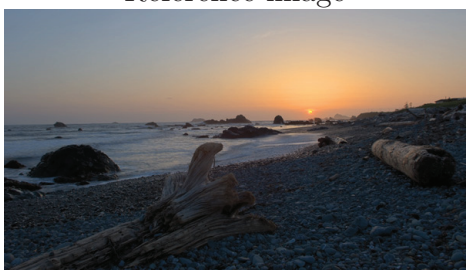
| TMO | PSNR | | | | SSIM | | | |
|---|---|---|---|---|---|---|---|---|
| | VSDP | TMQI | Ties | $p$ Value | VSDP | TMQI | Ties | $p$ Value |
| Schlick | 49 | 24 | 32 | 0.003 | 43 | 26 | 36 | 0.020 |
| Ashikhmin | 42 | 17 | 46 | 0.001 | 41 | 15 | 49 | 0.001 |
| Reinhard & Devlin | 45 | 32 | 28 | 0.138 | 37 | 36 | 32 | 0.907 |

default parameter settings in terms of both PSNR and SSIM index. However, after parameter tuning, those two operators can exhibit performance similar to Schlick's operator in terms of SSIM index, and even better results in terms of PSNR. Some current assessments of TMOs only consider default parameter settings [51, 55], and different rankings might be arrived at by considering the potential of operators for parameter tuning.

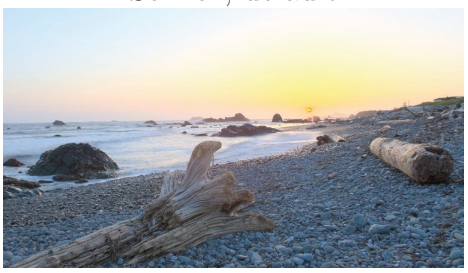
The third experiment is a comparison between parameter tuning based on VSDP and parameter tuning based on TMQI. Rather than simply averaging the performance, we use ten samples for each image in a Mann-Whitney U test to determine if the two sets of data are significantly different from each other. Then, the output of the Mann-Whitney U test, including both wins and ties, is further analyzed with a Wilcoxon signed-rank test for measuring the overall difference. Table 5.4 lists the numbers of images, for which VSDP performs better than TMQI, TMQI performs better than VSDP, and they have similar performance, as well as the $p$ values. For all three TMOs, VSDP results in visual appearance closer to the reference images than TMQI for a greater number of images. The difference is statistically significant for Schlick's and Ashikhmin's operators.


Figure 5.2: Results of the parameter tuning algorithm on HDR image "Bandon Sunset (1)". Shown are the reference image tone mapped with manual manipulation (provided at [26]), images tone mapped with Schlick's, Ashikhmin's, and Reinhard and Devlin's operators and default parameter settings, and images tone mapped with optimized parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

Reference image

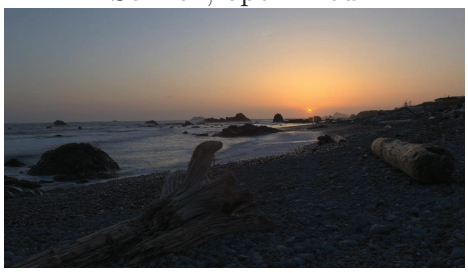


Schlick, default



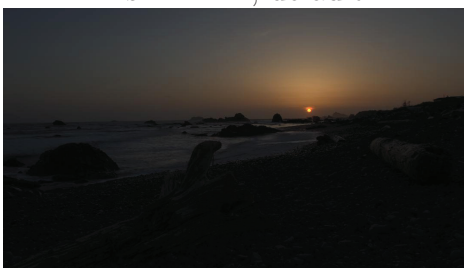
Saliency Distortion: 1.5237

Schlick, optimized



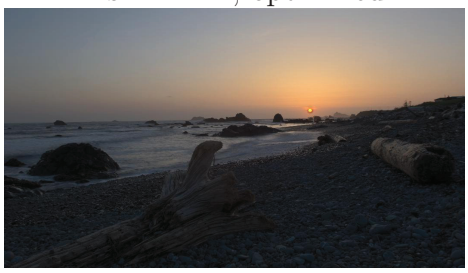
Saliency Distortion: 0.3948

Ashikhmin, default



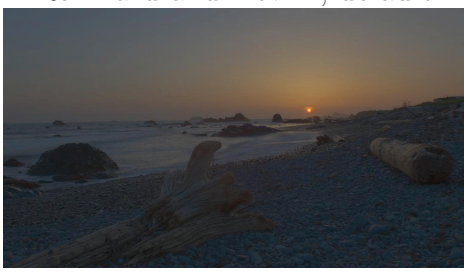
Saliency Distortion: 1.2039

Ashikhmin, optimized



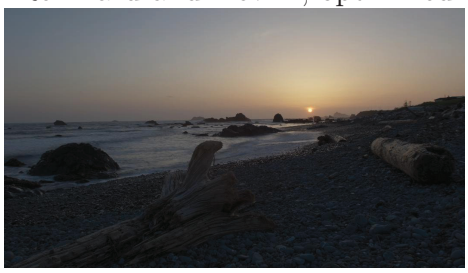
Saliency Distortion: 0.3905

Reinhard and Devlin, default



Saliency Distortion: 0.6871

Reinhard and Devlin, optimized



Saliency Distortion: 0.3303

Figure 5.3: Results of the parameter tuning algorithm on HDR image "Redwood Sunset". Shown are the reference image tone mapped with manual manipulation (provided at [26]), images tone mapped with Schlick's, Ashikhmin's, and Reinhard and Devlin's operators and default parameter settings, and images tone mapped with optimized parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

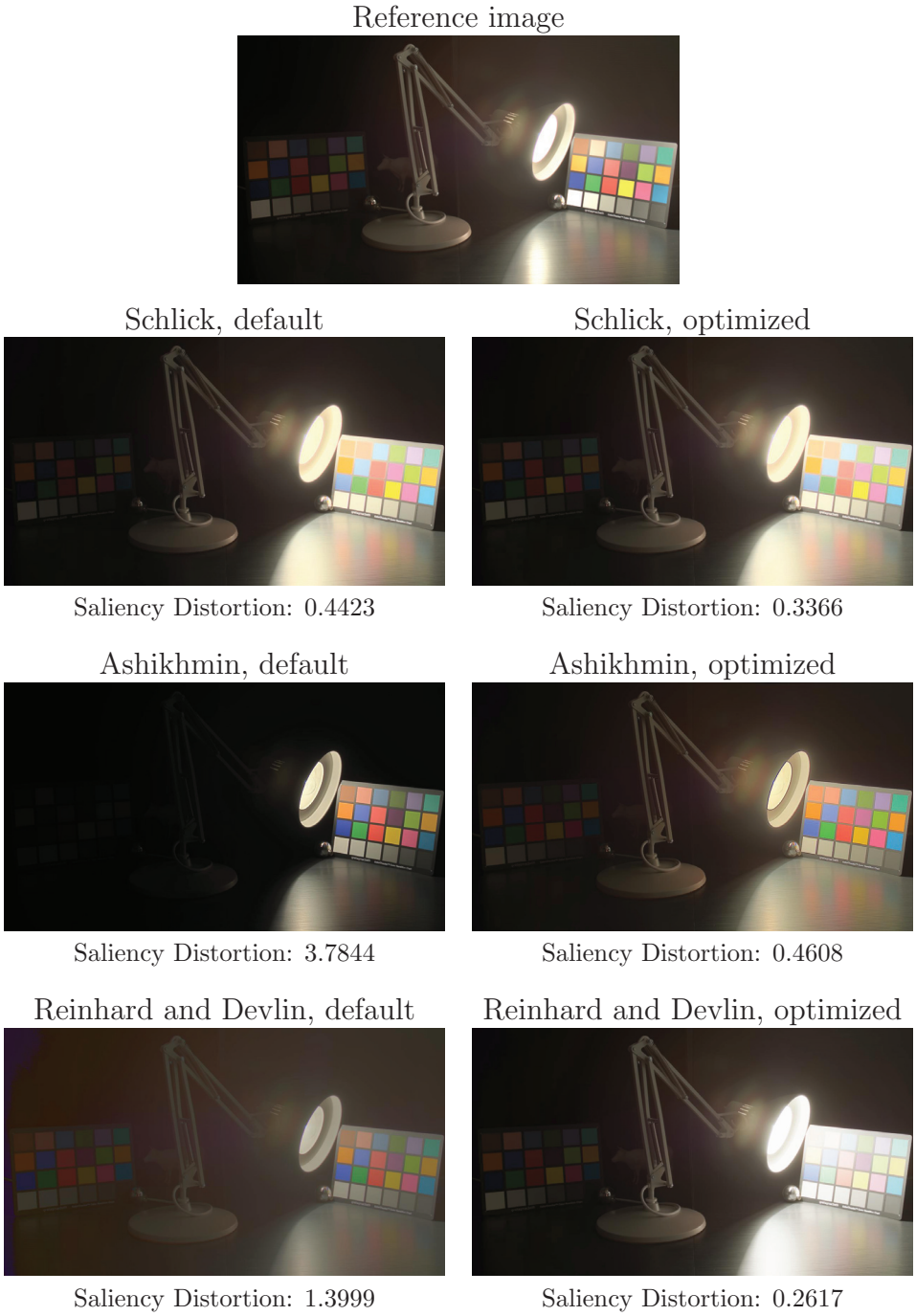
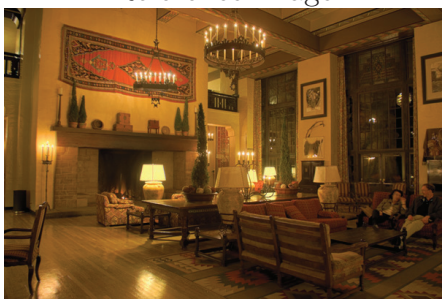
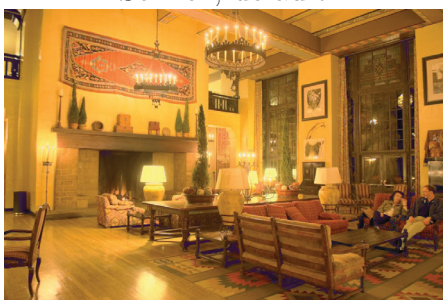



Figure 5.4: Results of the parameter tuning algorithm on HDR image "Luxo Double Checker". Shown are the reference image tone mapped with manual manipulation (provided at [26]), images tone mapped with Schlick's, Ashikhmin's, and Reinhard and Devlin's operators and default parameter settings, and images tone mapped with optimized parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

Reference image

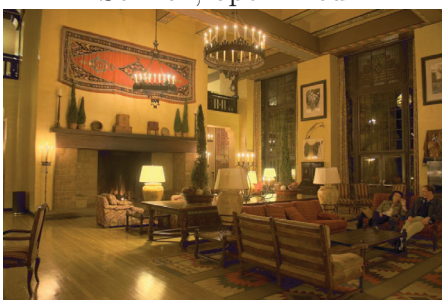


Schlick, default



Saliency Distortion: 0.8270

Schlick, optimized

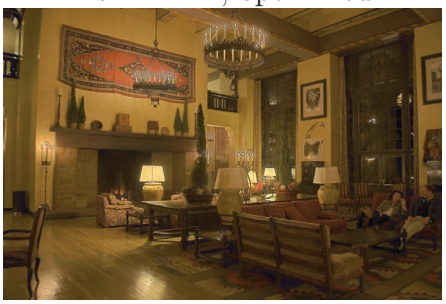


Saliency Distortion: 0.4416

Ashikhmin, default

Saliency Distortion: 1.3593

Ashikhmin, optimized

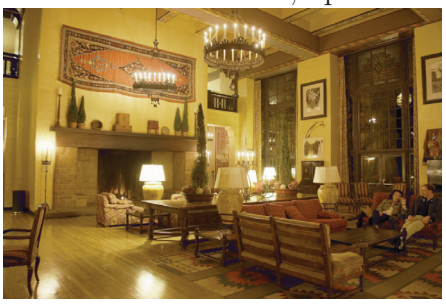


Saliency Distortion: 0.5996

Reinhard and Devlin, default

Saliency Distortion: 1.1225

Reinhard and Devlin, optimized



Saliency Distortion: 0.2636

Figure 5.5: Results of the parameter tuning algorithm on HDR image "Ahwahnee Great Lounge". Shown are the reference image tone mapped with manual manipulation (provided at [26]), images tone mapped with Schlick's, Ashikhmin's, and Reinhard and Devlin's operators and default parameter settings, and images tone mapped with optimized parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

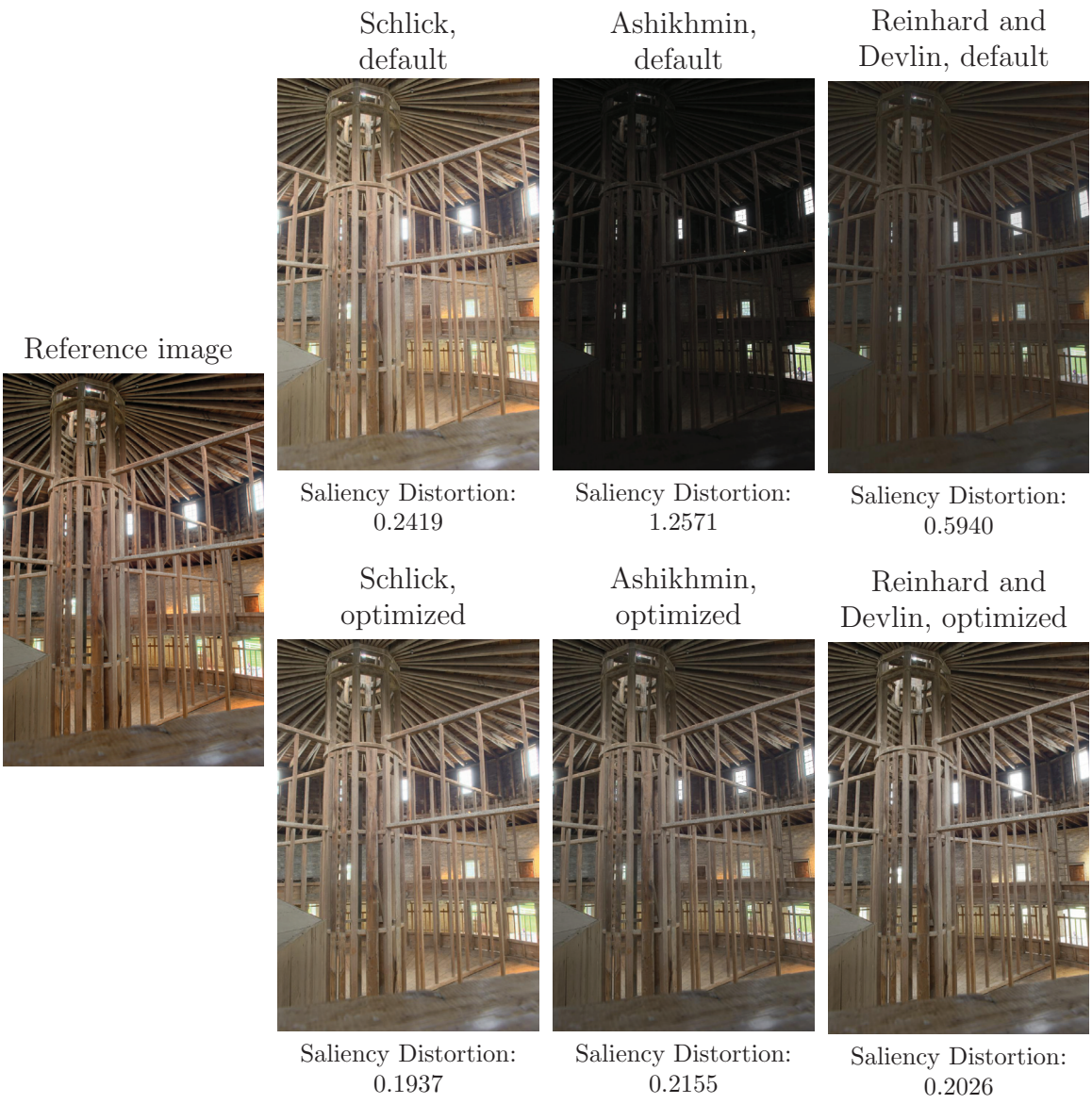



Figure 5.6: Results of the parameter tuning algorithm on HDR image "Round Barn Inside". Shown are the reference image tone mapped with manual manipulation (provided at [26]), images tone mapped with Schlick's, Ashikhmin's, and Reinhard and Devlin's operators and default parameter settings, and images tone mapped with optimized parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

## 5.3 Blended Tone Mapping

During the last two decades, a large number of TMOs have been proposed. However, the performance of TMOs are image-dependent [55, 52], and there is no single "best" operator that yields optimal results under all conditions. With a limited parameter space for tone mapping, a single operator could be insufficient to generate satisfying results for all HDR images. The limitation of an individual operator is demonstrated in Figures 5.7 and 5.8 using Schlick's operator [88] and Kuang et al.'s operator (iCAM06) [51]. As shown in the first example, Schlick's operator [88] can only provide a limited solution space and it does not offer good detail reproduction in both bright and dark areas at the same time. Even the best image that is subjectively selected fails to maintain the details in the relatively bright areas such as the distant woods. As for the second example, iCAM06 [51] does not grant sufficient flexibility to reproduce the luminance of the original HDR scene, and the subjectively-selected best image still appears too dim especially in the interior of the room, which can cause inaccurate perception for human observers.

By searching for the optimal results in a larger solution space, blending images from a number of commonly-used operators can overcome the limitations of each single operator. The process is referred to as blended tone mapping in the current context. Blended tone mapping has the potential to achieve results that cannot be produced by any individual operator. Chisholm et al. [16] proposed an interactive blended tone mapping and they demonstrate that good tone mapped images can often be obtained within a small number of iterations. With an user interface, people can select the best image from a set of blended tone mapped images, and an evolution strategy iteratively improves the appearance of the images until a satisfying result is found. Their work enables rapid manual manipulation but user interaction is still needed in the optimization process.

Objective image quality measures for tone mapped images makes it feasible to replace manual user interaction with an automatic assessment. In this section, we propose an automated blended tone mapping algorithm guided by the perceptual distortion predictor (PDP). The quality metric assesses tone mapped images in terms of brightness distortion, visual saliency distortion, and detail distortion in light and dark areas. With the quality metric, blended tone mapping can be automatically solved in

Figure 5.7: Tone mapped images across parameter space and the subjectively-selected best one of Schlick's operator [88] on HDR image "Tree". The images are generated with parameter setting for just noticeable difference which is provided by the original author of the operator. The HDR image is from Industrial Light & Magic.

Figure 5.8: Tone mapped images across parameter space and the subjectively-selected best one of Kuang's operator [51] on HDR images "Lab Windows". The images are generated with parameter settings of overall contrast and gamma value respectively where the ranges are provided by the original author of the operator. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

an optimization framework, where the parameters of TMOs and the weights that determine the relative influence of each operator are tuned to generate images with the least perceptual distortion. We validate the performance of the proposed algorithm with three commonly used TMOs and a large number of test images from various sources. There results show that the blending tone mapping algorithm can generate visual appealing tone mapped images without the requirement of user interaction.

### 5.3.1 Search Space

We select three TMOs for blended tone mapping to establish the viability of the approach: the operators proposed by Schlick [88], Kuang et al. [51], and Drago et al. [23]. Schlick's operator [88] provides a simple but effective scheme for tone mapping with a small number of user parameters. Kuang et al.'s operator [51], which is referred to as iCAM06, incorporates an edge-preserving filter with light adaptation functions modelled on the human visual system, and uses the modified color appearance model for HDR images rendering. Based on the logarithmic compression of luminance that imitates the human response to light, Drago et al.'s operator [23] introduces a bias power function to adaptively adjust logarithmic bases for good preservation of detail and contrast. In order to set up an efficient solution space that can produce good images with a relatively small number of parameters, several aspects, such as the diversity of operators that can enlarge the solution space, existence of parameters, and computational speed, were considered for the selection of operators.

For the selected operators, iCAM06 belongs to local operators, while the others are global operators. Any finite set of operators could be employed instead. The implementation of iCAM06 was provided by the original authors, and the implementation of Schlick's operator and of the operator by Drago et al. are due to Banterle [9]. The search space of blended tone mapping is composed of the parameters of TMOs and the weights that determine the relative contributions in the blending process and the user parameters provided by the individual TMOs. The ranges and initial values of parameters are from the recommendations of the original papers. In order to avoid the multiple applications of gamma correction, we do not use gamma as an extra parameter unless it is used by the authors of the TMO. As for the weights, they are real-valued variables within the range of 0 to 1, and the initial value is set to 1/3. The user parameters and weights of TMOs are summarized in Table 5.2. Altogether, the search space of blended tone mapping is thus seven-dimensional, due to four parameters and three weights.

### 5.3.2 Minimization of Perceptual Distortion

Blended tone mapping is optimized with the strategy $(1 + \lambda) - ES$. Candidate solutions $\mathbf{x} \in \mathbb{R}^n$ are real-valued vectors that consist of the user parameters and

Table 5.5: Search space of blended tone mapping.

| **Parameter** | **Default** | **Range** | **Description** |
|---|---|---|---|
| User Parameters | | | |
| $n$ | 0.18 | $[0.0, 1.0]$ | just noticeable difference |
| $p$ | 0.7 | $[0.6, 0.85]$ | overall contrast parameter |
| $\gamma$ | 1.0 | $[1.0, 1.2]$ | gamma adjustment parameter |
| $b$ | 0.85 | $(0.0, 1.0]$ | bias parameter |
| Weights | | | |
| $w1$ | 1/3 | $[0.0, 1.0]$ | weight of Schlick's operator |
| $w2$ | 1/3 | $[0.0, 1.0]$ | weight of iCAM06 |
| $w3$ | 1/3 | $[0.0, 1.0]$ | weight of Drago et al.'s operator |

weights of the selected operators. In iteration $t$, we generate $\lambda$ offspring from the parent $\mathbf{x}^{(t)}$, where $\lambda$ and $\sigma$ are set to 10 and 0.3 respectively in our implementation. Out-of-range values of variables are clamped to the ranges of the parameters, and weights are normalized to sum to one after mutation. With the parameters encoded in the candidate solutions, we apply the three operators for tone mapping and blend their results together. The tone mapped image are blended pixel-wise with contributions from individual operator multiplied with the normalized weight. In order to avoid undesired artifacts, the blending process is conducted in CIELab color space that is more perceptually uniform. Then, the perceptual distortion predictor (PDP) is employed to calculate the quality of blended images. The candidate solution that yields the least perceptual distortion is selected and adopted as the parent $\mathbf{x}^{(t+1)}$ for the next iteration. If the parent candidate solution is superior to all of its offspring, then it is not replaced in iteration $t$. As with Chisholm et al., we update the step size in those iterations where the parent is superior to all of its offspring by multiplying it with 0.8. If the parent has not changed for three consecutive iterations, the process will end and the current parent will be returned. The algorithm is demonstrated in Algorithm 3.

**Algorithm 3** Blended Tone Mapping

Set the generation counter $t = 0$;
Initialize the strategy parameters $\lambda$, $\sigma$;
Create the population;
Generate tone mapped images with the operators;
Blend the images in CIELab color space;
Calculate perceptual distortion with PDP;
**while** Stopping condition is not fulfilled **do**
    **for** $i \in \{1, ..., \lambda\}$ **do**
        Generate offspring with mutation;
        Generate tone mapped images with the operators;
        Blend the images in CIELab color space;
        Calculate perceptual distortion with PDP;
    Update the population with generated offspring;
    Select the individual with the least perceptual distortion;
    Update the step size $\sigma$;
    $t = t + 1$;
Return the optimal result;

### 5.3.3 Experimental Results

We have tested our blended tone mapping algorithm on a variety of HDR scenes from difference sources, including some challenging scenes with large contrast ratios. In all cases, our algorithm can produce good results without the requirement of user interaction. Once again, the computational cost of our algorithm is dominated by the cost of tone mapping. For the computation of perceptual distortion, our MATLAB implementation takes around $0.8s$ for a $1024 \times 768$ pixel image running on an Intel Quad-Core 2.66 GHz CPU with 4 GB of RAM.

Several examples are demonstrated in Figure 5.10. Next to each image there are thumbnail images showing some of the exposures used to construct HDR scenes. As we can see, our blended tone mapping algorithm is capable of combining the local details that can only be captured in various exposures in a realistic manner, such as the foreground grasses and background woods (First row; Tree), the texture of the stone walls and plants (Second row; Knossos Ruins), the stained glass and tables (Second

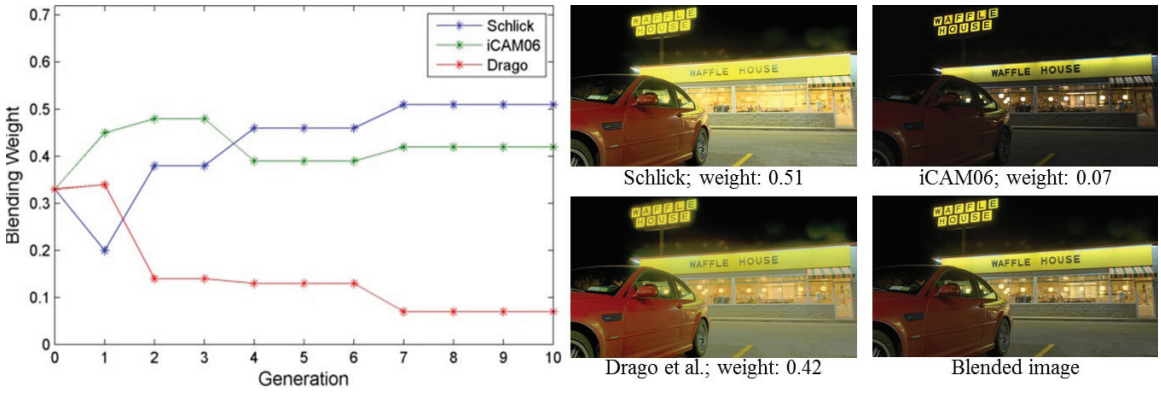



Figure 5.9: Evolution of blending weights in a sample run of automatic blended tone mapping on the "Waffle House" HDR image. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

row; Stained Glass), the scenery outside window and items on the table (Third row; Lab Window), and the color charts (Third row; Luoxo Double Checker). Guided by the perceptual distortion predictor, the algorithm produces visually pleasing tone mapped images that can appear faithful in overall impression while at same time preserving fine details in relatively bright and dark areas.

Figure 5.9 illustrates the evolution of blending weights. As we can see, the blending operation can overcome the limitations of each single TMO, and the optimization converges within a relatively small number of iterations. In this example, Schlick's operator [88] and iCAM06 [51] have much higher weights than Drago et al.'s operator [23]. Schlick's operator [88] can provide sufficient flexibility to tune the overall light levels, while iCAM06 [51] can well preserve the details in bright areas such as the interior of restaurant. By combining the advantages of both operators, blended tone mapping can contribute to an overall desirable result. The weight of Drago et al.'s operator [23] reduces markedly in the evolutionary process. We assume it is because the operator may cause halo artifacts for this HDR image.

Figure 5.11, 5.12, 5.13 compare our algorithm with the frequently-used TMOs using parameters recommended by the original authors. Eight TMOs are adopted in our experiment, including the ones published by Ward [104], Schlick [88], Larson et al. [54], Reinhard et al. [85], Durand and Dorsey [24], Ashikhmin [85], Fattal et al.[28], and Tumblin and Turk [95], where the first three are global operators and the others

are local operators. Generally speaking, global operators cannot perform well for visual details reproduction especially in light and dark areas, while local operators may produce artifacts making the image appear unnatural. The shortcomings of global and local operators have been well addressed by the blended tone mapping algorithm.

As shown in Figure 5.11, the tone mapped image generated by Schlick's operator (First row; left) provides an overall realistic appearance but the details in bright regions such as the clouds in sunset are washed out. Although the results from Reinhard et al. and Durand and Dorsey's operators have better preservation of the details, they suffer from overemphasized edges (First row; right) and a non-photorealistic appearance (Second row; left), which could impact the overall similarity to real world scenes. In comparison, the clouds in sunset are rather smoothly and naturally preserved in our result without causing any undesired artifacts (Second row; right). Figure 5.12 provides an additional example with Ward's, Ashikhmin's, and Fattal et al.'s operators. Our algorithm can generate realistic results with rich details in light and dark areas while the operators cannot. The same pattern occurs in Figure 5.13. Larson et al.'s operator loses details in the areas around street light (left) and the Tumblin and Turk's operator introduces "embossed" and unnatural appearances (middle). In contrast, blended tone mapping can strike a balance between overall similarity and detail reproduction, and yield images with appealing appearance even for challenging scenes.

"Tree", and "Synagogue"

"Knossos Ruins", and "Stained Glass"

"Lab Window", and "Luxo Double Checker"

Figure 5.10: Tone mapped images generated by our blended tone mapping algorithm with HDR images from various sources. The HDR image "Tree" and "Stained Glass" are from Industrial Light & Magic, the HDR image "Synagogue" is from Dani Lischinski, the HDR image "Knossos Ruins" is from EMPA Media Technology, and the HDR image "Lab Window" and "Luxo Double Checker" are from Mark D. Fairchild.

Schlick's operator

Perceptual Distortion: 0.1007

Reinhard et al.'s operator

Perceptual Distortion: 0.1216

Durand and Dorsey's operator

Perceptual Distortion: 0.1853

Blended tone mapping

Perceptual Distortion: 0.0902

Figure 5.11: Comparison of blended tone mapping with TMOs on HDR image "Swiss Sunset". Shown are the tone mapped images generated by Schlick's, Reinhard et al.'s, Durand and Dorsey's operators, and blended tone mapping. The HDR image is from EMPA Media Technology.

Ward's operator

Perceptual Distortion: 0.1626

Ashikhmin's operator

Perceptual Distortion: 0.1452

Fattal et al.'s operator

Perceptual Distortion: 0.2004

Blended tone mapping

Perceptual Distortion: 0.1064

Figure 5.12: Comparison of blended tone mapping with TMOs on HDR image "Desk Lamp". Shown are the tone mapped images generated by Ward's, Ashikhmin's, Fattal et al.'s operators, and blended tone mapping. The HDR image is from Martin Čadík.


Figure 5.13: Comparison of blended tone mapping with TMOs on HDR image "Foggy Night". Shown are the tone mapped images generated by Ward et al.'s, and Tumblin and Turk's operators, and blended tone mapping. The HDR image is from Jack Tumblin.

## 5.4 Conclusion and Discussion

In this chapter, we use the proposed feature-based quality metrics to address the automatic optimization problems for tone mapping. The visual saliency based parameter tuning algorithm enables the automatic parameter selection within any arbitrary TMOs. Meanwhile, the perceptual quality guided blended tone mapping algorithm searches for the optimal solution across various TMOs for a particular HDR image. We shows that the optimization can be accomplished by an evolution strategy. As one of the first attempts for automated tone mapping optimization, the proposed algorithms advance the application of existing TMOs and inspire further improvements on this research topic.

The performance of the tone mapping optimization algorithms can be improved in a number of aspects. The time complexity of the algorithms are dominated by the running time of TMOs. Due to its requirement of independent applications of tone mapping, we are planing to distribute the computation of the quality scores predicted by objective quality metrics of the offspring across multiple cores to reduce the running time. Also, considering a simple evolution strategy is used in the algorithms, it would be interesting to further analyze the influence of strategy selections, and explore more standard strategies such as CSA-ES for tone mapping optimization. Additionally, the parameter space of TMOs is adopted as the solution space for tone mapping optimization. Although it is an important component that can affect the final results, little work has been done to exploit and compare different solution spaces. Further studies and comparison towards solution space can be conducted in the future.

# Chapter 6

# Comparison of Optimization Methods

## 6.1 Introduction

The development of reliable image quality measures for the assessment of tone mapped images constitutes a significant advancement in high dynamic range imaging. The ability to objectively assess the quality of tone mapped images allows treating tone mapping as an optimization problem that can be solved by automated algorithms, without the need for human input. By applying evolutionary algorithms (EA) to parameter spaces of TMOs, we introduce a general framework to realize objective quality assessment based tone mapping optimization. The framework has successfully been used for various optimization problems, including parameter tuning of TMOs [34, 35], and blended tone mapping [36].

More recently, Ma et al. [61] present an optimization method in connection with the improved variant of tone mapped image quality index (TMQI-II). The proposed method interleaves the optimization of structural fidelity using gradient ascent with that of statistical naturalness. Experiments show that it can significantly improve the TMQI-II scores, albeit usually at a high computational cost. Although the algorithm is especially designed for their quality metric, it represents an alternate solution for tone mapping optimization which makes use of gradient information of quality metrics to search for the optimal results in image space.

Compared with Ma et al.'s method [61], EA optimizes parameters of TMOs, rather than tone mapped images. These two approaches have been applied for different quality metrics, and no direct comparison has been performed. In this chapter, we compare them under a common platform [32]. To establish that, we apply the EA for the optimization of TMQI-II addressed by Ma et al. [61]. The objective of EA is set to TMQI-II, and the solution space is defined as the parameter space of a generic TMO [67]. A comparison is conducted with an HDR image benchmark set and various starting points. According to the experimental results, EA can outperform Ma et

al.'s method with significantly reduced computational effort, which means it can be more efficient and practical to perform objective quality metric based optimization on parameter space.

## 6.2 Gradient-based Optimization

With TMQI-II as the objective, Ma et al. [61] solve the tone mapping optimization in image space, whose dimensionality is equal to the number of pixels. To address that, they conduct an iterative numerical optimization based on gradient information. They first improve the the structural fidelity with a gradient ascent method, and then update the statistical naturalness with a point-wise intensity transformation, where the parameters are estimated with a gradient projection algorithm. These two steps continue until convergence.

### 6.2.1 Structural Fidelity Update

Given an HDR image $X$ and a tone mapped image $Y$, Ma et al. [61] calculate the gradient of the structural fidelity $S(X, Y)$ with respect to $Y$. The updating of the $k$-th iteration input image $Y_k$ is described as

$$\hat{Y}_k = Y_k + \lambda \nabla_Y S(X, Y)\Big|_{Y=Y_k}, \tag{6.1}$$

in which $\hat{Y}_k$ denotes the output image and $\lambda$ is the step size. Similar with the calculation of structural fidelity that uses a sliding window across images to predict local detail preservation, they calculate the gradient of overall structural fidelity $\nabla_Y S(X, Y)$ by combining the gradient of local image patches $x$ and $y$ as follows:

$$\nabla_Y S(X, Y) = \frac{1}{M} \sum_{i=1}^{M} R_i^T \nabla_y S_{local}(x, y)\Big|_{x=x_i, y=y_i}. \tag{6.2}$$

In the equation, $M$ is the number of local patches; $x_i = R_i(X)$ and $y_i = R_i(Y)$ are the $i$-th patches in $X$ and $Y$; $R_i$ and $R_i^T$ are the operators that take $i$-th local patch from the image and place it back respectively.

### 6.2.2 Statistical Naturalness Update

With the image $\hat{Y}_k$ from the optimization procedure of structural fidelity, the statistical naturalness is improved through a three-segment equipartition monotonic piecewise linear function:

$$y_{k+1}^i = \begin{cases} (3/L)a\hat{y}_k^i & 0 \le \hat{y}_k^i \le L/3 \\ (3/L)(b-a)\hat{y}_k^i + (2a-b) & L/3 \le \hat{y}_k^i \le 2L/3 \\ (3/L)(L-b)\hat{y}_k^i + (2b-2L) & 2L/3 \le \hat{y}_k^i \le L, \end{cases} \tag{6.3}$$

where $\hat{y}_k^i$ and $y_{k+1}^i$ denote the pixels of the input and output images at $i$-th patch, $L$ is the dynamic range of a tone mapped image, and $a$ and $b$ are parameters to establish the desired brightness (mean) and contrast (standard derivation).

The parameters $a$ and $b$ (where $0 \le a \le b \le L$) are chosen so that the brightness $\mu$ and contrast $\sigma$ of $Y_{k+1} = \{y_{k+1}^i \text{ for all } i\}$ better approximate the brightness $\mu_e$ and contrast $\sigma_e$ of the desired image. To generate the parameters $a$ and $b$, they first estimate the brightness and contrast of $Y_{k+1}$:

$$\begin{aligned} \mu_{k+1}^e &= \hat{\mu}_k + \lambda_m(\mu_e - \hat{\mu}_k) \\ \sigma_{k+1}^e &= \hat{\sigma}_k + \lambda_d(\lambda_e - \hat{\lambda}_k). \end{aligned} \tag{6.4}$$

In the equation, $\hat{\mu}_k$ and $\hat{\sigma}_k$ denote the brightness and contrast of $\hat{Y}_k$; $\lambda_m$ and $\lambda_d$ denote the parameters that control the updating speed. Afterwards, they calculate parameters $a$ and $b$ by solving the constrained optimization problem:

$$\{a,b\}_{opt} = \arg\max_{\{a,b\}} \|\mu_{k+1} - \mu_{k+1}^e\|^2 + \eta\|\sigma_{k+1} - \sigma_{k+1}^e\|^2, \tag{6.5}$$

in which $\eta$ is the weight that adjusts the two terms. They use a gradient projection algorithm [77] to solve the problem, and then plug the optimal values of $a$ and $b$ into the equation 6.3.

The above-mentioned operations result in an image $Y_{k+1}$ which is the output of $(k+1)$-th iteration. The improvement of structural fidelity and statistical naturalness alternate until the Euclidean norm of tone mapped images $\|Y_{k+1} - Y_k\|$ is smaller than a predefined threshold.

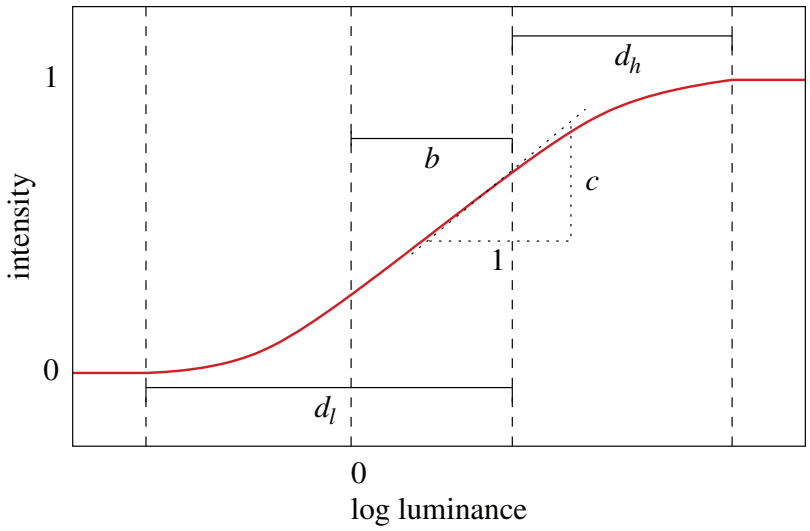



Figure 6.1: Tone curve used in the generic TMO and its parameters (adapted from [67]). Parameter $b$ allows for brightness adjustment, $d_l$ and $d_h$ determine the lower and higher midtone ranges,and $c$ governs contrast.

### 6.3 Evolutionary Optimization

In our algorithm, the solution space is defined by the generic TMO, and the optimization is accomplished with an evolution strategy. The generic TMO by Mantiuk and Seidel [67] aims to provide the ability to emulate different TMOs using computationally inexpensive image processing operations. We choose it for its low computational cost as well as due to its capacity to generate tone mapped images with plausible TMQI-II scores using a relatively small set of parameters. The operator maps intensity values as

$$C_{\mathrm{LDR}} = f_{\mathrm{MT}}(f_{\mathrm{TC}}(L_{\mathrm{HDR}})) \cdot \left(\frac{C_{\mathrm{HDR}}}{L_{\mathrm{HDR}}}\right)^{s}, \tag{6.6}$$

where $C_{\mathrm{HDR}}$ and $C_{\mathrm{LDR}}$ are the color channels of the HDR image and the tone mapped image, respectively, $L_{\mathrm{HDR}}$ is the luminance of the HDR image, $f_{\mathrm{TC}}$ denotes the tone curve, $f_{\mathrm{MT}}$ represents the modulation transfer function, and $s$ is a parameter used for saturation adjustment. The tone curve is a sigmoidal function, with parameters $b$, $d_l$, $d_h$, and $c$ provided for tuning the curve shape as illustrated in Fig. 6.1. The modulation transfer function allows specifying several parameters that determine a 1D function of spatial frequency and allows the tuning of blurring and sharpening operations applied to an image. The function involves band-pass filtering implemented with

difference of Gaussian operators, with parameters $m_1$, $m_2$, and $m_3$ for the adjustment of different frequency components [67].

Since TMQI-II scores are computed on the basis of luminance values, without taking color information into account, we choose not to modify the saturation adjustment parameter $s$. This leaves a total of seven parameters available for tuning. We use a $(1+\lambda)$-ES to solve the optimization of TMQI-II scores in the seven-dimensional parameter space. The application of the evolutionary approach to optimization is motivated by the lack of availability of analytical gradients and the potential for ruggedness resulting from the choice of quality criterion.

There are no restrictions regarding the setting of parameters of the generic TMO. However, we find that allowing parameters to grow without bounds may result in very marginal changes to TMQI-II scores and thus in ill-conditioning that negatively impacts the ability of the simple evolution strategy to optimize image quality. We thus impose boundary constraints that prevent the parameters from moving past their useful ranges. In order to be able to define image-independent ranges, we calibrate the logarithmic luminance values of the pixels in an HDR image by subtracting the mean logarithmic luminance. Ranges and short descriptions of the parameters are listed in Table 6.1.

Candidate solutions are seven-dimensional real vectors that consist of the parameters of the generic TMO. In each iteration of the algorithm, $\lambda > 1$ offspring are generated from the parental candidate solution $\mathbf{x} \in \mathbb{R}^7$ as

$$\mathbf{y}_i = \mathbf{x} + \sigma \mathbf{z}_i \qquad i = 1, \ldots, \lambda \tag{6.7}$$

where $\sigma \in \mathbb{R}$ denotes the step size parameter and the $\mathbf{z}_i \in \mathbb{R}^7$ are independent, standard normally distributed mutation vectors. Out-of-range values of variables are clamped to the boundaries. In light of potential issues with decreasing step size as a result of constraint handling such as discussed by Arnold [3], we have also experimented with an exterior penalty approach, but not observed a significant difference in performance. In each iteration, the candidate solution that leads to the highest TMQI-II value among the union of the parent and the set of all offspring is selected and adopted as the parent for the next iteration. The offspring number $\lambda$ is set to 10 throughout, and the step size parameter is initialized to 0.5 at the start of a run. That parameter is decreased by multiplication with 0.8 in each iteration where the parental

Table 6.1: Parameters of the generic TMO.

| Parameter | Range | Description |
|---|---|---|
| Parameters of tone curve | | |
| $b$ | $[-2.0, 2.0]$ | brightness factor |
| $d_l$ | $[0.0, 2.5]$ | lower midtone range factor |
| $d_h$ | $[0.0, 2.5]$ | higher midtone range factor |
| $c$ | $[0.2, 1.5]$ | contrast factor |
| Parameters of modulation transfer function | | |
| $m_1$ | $[-2.0, 2.0]$ | high frequency factor |
| $m_2$ | $[-2.0, 2.0]$ | medium frequency factor |
| $m_3$ | $[-2.0, 2.0]$ | low frequency factor |

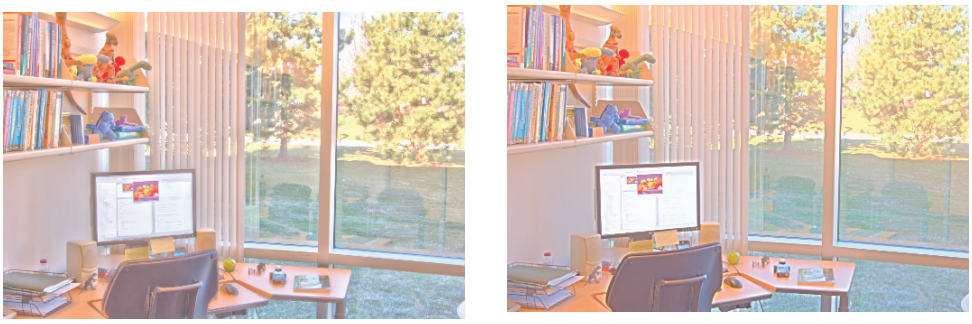
candidate solution is superior to all of its offspring; it is unchanged in those iterations where an offspring candidate solution is successful. We terminate a run when the change in the best TMQI-II value has been less than $10^{-4}$ for six consecutive iterations and return the best candidate solution found as the final result.

## 6.4 Methods Comparison

### 6.4.1 Experiment Settings

To evaluate the performance of the evolutionary approach, we conduct a comparison with the algorithm by Ma et al. [61] for the optimization of TMQI-II scores. We carry out the comparison on a set of sixteen HDR images. This set is identical with that used by Ma et al. in the evaluation of their approach, except of that we omitted one image to which we do not have access.

For each HDR image, we establish three starting points in the search for optimal tone mapped images. Ma et al. use starting points generated by various TMOs with default parameters settings, including the the logarithmic operator, Durand and Dorsey's operator [24], and Mantiuk's operator [65]. In order to ensure the same starting points for both algorithms, we determine parameter settings for the generic TMO such that the resulting tone mapped images closely match those generated by the various TMOs. Figure 6.2 shows a typical example of the generic TMO's ability to



Durand's TMO | Generic TMO (SSIM = 0.9419)

Mantiuk's TMO | Generic TMO (SSIM = 0.9918)

Figure 6.2: Comparison between images tone mapped using the TMOs by Durand (top left) and Mantiuk (bottom left) using default parameter settings, and corresponding images generated using the generic TMO (right) with parameters chosen to maximize SSIM scores.

emulate other TMOs and find matches that are visually nearly indistinguishable. The parameter settings and simulated tone mapped images are adopted as the starting points for EA and Ma et al.'s method respectively.

We use the implementation by Ma et al. [61] for TMQI-II and their optimization algorithm. That implementation is in *Matlab* and, according to the authors, not optimized for speed. Since the same implementation of TMQI-II is used for image quality assessment in both of the optimization algorithms, the comparison here is meaningful for establishing relative performance. The EA as well as the generic TMO are implemented in *Matlab* and not optimized for speed either. Running times reported are for a PC with Intel Quad-Core 2.66GHz CPU with 4GB of RAM. As the optimization approach by Ma et al. does not exploit parallelism, we have chosen not to make use of more than one CPU core in the implementation of the EA. Making use of parallel computational resources by evaluating offspring simultaneously on multiple cores is straightforward. On our hardware, the computation of a single TMQI-II score

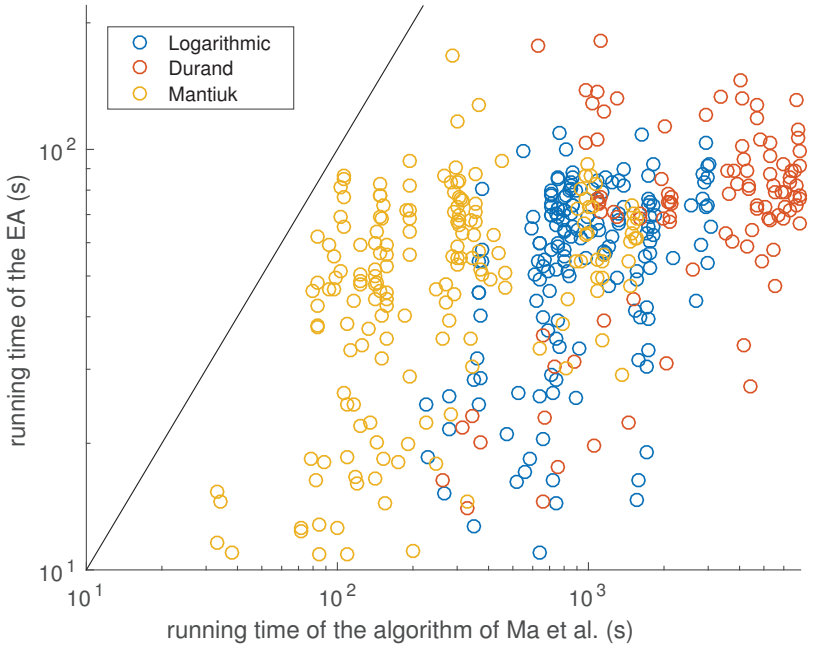



Figure 6.3: Running times of the EA plotted against running times required by the algorithm by Ma et al. [61] to reach equivalent TMQI-II scores. Data are shown for starting points matching those generated by the logarithmic operator, the Durand's operator, and Mantiuk's TMO. The black line indicates the identity.

for images of the size considered here (approximately $360 \times 500$ pixels) takes about 0.3 seconds.

For each of the sixteen HDR images and three starting points, we conduct eleven independent runs of the EA for the optimization of TMQI-II, for a total of 528 runs. All of the TMQI-II scores are calculated from tone mapped images stored in PNG format (i.e., with lossless compression, but with only eight bits per color channel). Table 6.2 shows TMQI-II scores of the starting points as well as median and standard deviation of the scores obtained after evolutionary optimization. Also shown are median and standard deviation of the computational times. We then ran the algorithm of Ma et al. [61] and recorded the time it requires to reach the TMQI-II scores obtained by the EA. Median and standard deviation of those times are shown in the last column of the table. Running time data from all 528 runs are represented graphically in Figure 6.3.

Table 6.2: Comparison between the EA and the algorithm of Ma et al. [61]. The table lists TMQI-II scores of starting points, scores after evolutionary optimization, the computational time (in seconds) required to generate those results, and the computational time (in seconds) for the algorithm of Ma et al. to reach equivalent TMQI-II scores. The computation time is omitted (—) when an equivalent score could not be reached within 7200 seconds. Shown are median values of eleven independent runs, with standard deviations given in parentheses.

| HDR Image | Starting Point | Initial TMQI-II Score | Optimized TMQI-II Score | Running Time (EA) | Running Time (Ma et al.) |
|---|---|---|---|---|---|
| Foggy Night | Logarithm | 0.3807 | 0.9780 (0.0047) | 58.5 (23.6) | 756.0 (39.0) |
| 340 × 512 | Durand | 0.3844 | 0.9766 (0.0020) | 71.0 (40.4) | 1263.7 (304.9) |
| | Mantiuk | 0.9091 | 0.9779 (0.0011) | 74.6 (26.3) | 357.6 (39.2) |
| Clock Building | Logarithm | 0.4409 | 0.9757 (0.0029) | 80.6 (22.6) | 1035.7 (288.3) |
| 384 × 512 | Durand | 0.4415 | 0.9752 (0.0040) | 68.4 (24.2) | 5748.9 (2126.1) |
| | Mantiuk | 0.9692 | 0.9769 (0.0031) | 38.0 (16.2) | 83.9 (22.8) |
| Dani Cathedral | Logarithm | 0.3999 | 0.9704 (0.0099) | 76.3 (27.8) | 2835.3 (1079.6) |
| 384 × 512 | Durand | 0.4186 | 0.9703 (0.0011) | 62.9 (17.3) | — (—) |
| | Mantiuk | 0.4616 | 0.9700 (0.0100) | 79.0 (24.4) | 293.0 (46.9) |
| Kitchen | Logarithm | 0.3651 | 0.9714 (0.0002) | 52.1 (12.0) | 1756.7 (51.6) |
| 342 × 512 | Durand | 0.3804 | 0.9715 (0.0160) | 73.3 (22.8) | 6760.6 (2120.2) |
| | Mantiuk | 0.8896 | 0.9712 (0.0033) | 56.9 (20.9) | 147.4 (19.1) |
| Memorial Church | Logarithm | 0.4442 | 0.9795 (0.0041) | 63.0 (20.5) | 1716.9 (65.7) |
| 340 × 512 | Durand | 0.4520 | 0.9804 (0.0046) | 63.7 (24.6) | — (—) |
| | Mantiuk | 0.9086 | 0.9798 (0.0015) | 43.5 (16.3) | 116.0 (11.3) |
| Women | Logarithm | 0.4151 | 0.9806 (0.0074) | 64.0 (26.1) | 1015.0 (336.7) |
| 342 × 512 | Durand | 0.4135 | 0.9809 (0.0077) | 96.2 (27.1) | 5283.8 (1771.6) |
| | Mantiuk | 0.5189 | 0.9808 (0.0120) | 67.4 (41.6) | 299.6 (61.0) |
| Desk 1 | Logarithm | 0.3881 | 0.9801 (0.0002) | 83.9 (8.4) | 857.7 (7.9) |
| 512 × 384 | Durand | 0.4256 | 0.9800 (0.0028) | 75.1 (18.0) | 2079.3 (431.1) |
| | Mantiuk | 0.7600 | 0.9800 (0.0067) | 56.2 (25.0) | 192.9 (31.4) |
| Desk 2 | Logarithm | 0.3765 | 0.9654 (0.0045) | 54.0 (17.8) | 800.8 (29.9) |
| 512 × 384 | Durand | 0.4031 | 0.9655 (0.0083) | 76.4 (47.9) | 1088.4 (236.4) |
| | Mantiuk | 0.8178 | 0.9652 (0.0069) | 51.5 (27.0) | 102.6 (14.1) |
| Display1000 | Logarithm | 0.4004 | 0.9649 (0.0038) | 59.6 (31.0) | 2810.3 (1033.1) |
| 512 × 384 | Durand | 0.4220 | 0.9648 (0.0070) | 99.3 (36.1) | 6983.4 (—) |
| | Mantiuk | 0.7236 | 0.9649 (0.0059) | 63.2 (25.3) | 985.1 (351.5) |
| Belgium House | Logarithm | 0.4096 | 0.9778 (0.0005) | 73.0 (14.3) | 707.7 (19.4) |
| 512 × 384 | Durand | 0.4186 | 0.9777 (0.0030) | 91.8 (20.2) | 5986.0 (1400.7) |
| | Mantiuk | 0.8552 | 0.9774 (0.0052) | 47.8 (18.8) | 141.3 (25.9) |
| Woods | Logarithm | 0.0708 | 0.9844 (0.0056) | 75.3 (24.3) | — (—) |
| 512 × 340 | Durand | 0.3541 | 0.9845 (0.0069) | 77.3 (30.3) | — (—) |
| | Mantiuk | 0.4957 | 0.9843 (0.0024) | 54.7 (14.7) | 1078.2 (356.3) |
| Lawn | Logarithm | 0.4434 | 0.9864 (0.0069) | 68.1 (17.8) | 1318.3 (317.5) |
| 512 × 381 | Durand | 0.4585 | 0.9861 (0.0043) | 89.2 (35.9) | 4276.2 (1558.4) |
| | Mantiuk | 0.9689 | 0.9861 (0.0019) | 39.3 (19.1) | 278.4 (69.2) |
| Bristol Bridge | Logarithm | 0.3968 | 0.9843 (0.0128) | 52.0 (27.0) | 746.5 (191.7) |
| 512 × 384 | Durand | 0.3891 | 0.9845 (0.0001) | 83.8 (23.0) | — (—) |
| | Mantiuk | 0.9520 | 0.9841 (0.0071) | 54.3 (25.7) | 887.5 (364.3) |
| Office | Logarithm | 0.4477 | 0.9729 (0.0012) | 54.3 (18.7) | 368.2 (4.2) |
| 512 × 340 | Durand | 0.4245 | 0.9738 (0.0020) | 71.2 (24.5) | — (—) |
| | Mantiuk | 0.9503 | 0.9742 (0.0048) | 44.1 (16.7) | 158.2 (36.1) |
| Vine Sunset | Logarithm | 0.4246 | 0.9636 (0.0051) | 50.6 (21.3) | 613.3 (123.6) |
| 512 × 345 | Durand | 0.4440 | 0.9639 (0.0007) | 89.7 (27.8) | — (—) |
| | Mantiuk | 0.4617 | 0.9641 (0.0003) | 73.9 (12.6) | 1498.9 (56.8) |
| Wreathbu | Logarithm | 0.491 | 0.9652 (0.0010) | 58.8 (18.0) | 1204.1 (354.4) |
| 512 × 384 | Durand | 0.4365 | 0.9655 (0.0010) | 88.6 (34.0) | 3814.1 (716.4) |
| | Mantiuk | 0.7982 | 0.9660 (0.0010) | 49.0 (21.0) | 406.1 (82.1) |

### 6.4.2 Experimental Results

As shown in the Table 6.2, EA can significantly improve TMQI-II scores. Final scores are within a narrow range, both across test images and across starting points. Standard deviations are such that the empirical coefficient of variation of TMQI-II scores rarely exceeds 0.01. Running times for the EA range from well under a minute to no more than three minutes in the longest of the 528 runs. In comparison, the algorithm of Ma et al. [61] requires significantly more time to generate tone mapped images with equivalent TMQI-II scores, and in a number of instances remains unsuccessful even after two hours (where runs are terminated).

It is worth noting that Ma et al. [61] report TMQI-II scores in excess of most of the values attained by the EA and reported in Table 6.2 (running times are not reported with the TMQI-II scores). The EA operates in a low-dimensional search space that may implicitly limit the quality of the tone mapped images that can be achieved by parameter optimization. However, we notice that their TMQI-II scores are reported on floating-point image data. When storing the data into image file formats with eight bits per colour channel and pixel, any further improvement can disappear due to quantization noise.

Clearly, the running times of both the EA and the algorithm by Ma et al. are impacted by the size of the images being processed. However, the approach relying on the generic TMO and the EA for optimization admits a simple technique for reducing running time: Rather than performing the optimization on potentially sizable images, shrink the images before applying the EA to obtain parameter settings for the generic TMO, and then use the parameter settings obtained on the small images to tone-map the full-sized images. Two examples of results from this approach can be found in Figure 6.4. The images on the left have been obtained through optimization using the full-sized images of size $1024 \times 768$. The optimization took 344 seconds for the "Bristol Bridge" image and 200 seconds for the "Belgium House" image. The images on the right are the result of computing parameter settings on images of size $256 \times 192$ and applying those settings to the full-sized images. The TMQI-II scores decrease in both cases, but the results are visually nearly indistinguishable, and optimization in the latter case is accomplished in 11 and 9 seconds respectively. Another technique for reducing running time while not having to contend with reduced TMQI-II scores

TMQI-II: 0.9794 TMQI-II: 0.9759

TMQI-II: 0.9716 TMQI-II: 0.9660

Figure 6.4: Comparison between images with parameters of the generic TMO obtained through evolutionary optimization on different sized versions of the images. The images on the left have been obtained from optimization on the full-sized images of size $1024 \times 768$. The images on the right are the result of shrinking the images to size $256 \times 192$, solving the optimization problem using the EA, and then using the TMO parameter settings obtained on the full-sized images.

is to start with small images and to increase the size of the images being processed during the run of the EA.

Finally, in addition to the reduced running time, we have found our method to often be preferable to the algorithm by Ma et al. [61] in that it generates images with more consistent appearance across various starting points. Figure 6.5 shows examples where the appearance of the images generated using the algorithm by Ma et al. differs from starting point to starting point and suffers from artifacts such as over- and under-saturation, while the results generated using the EA look rather uniform. Further examples can be found in the complete set of experimental data which is available at `www.cs.dal.ca/~xgao/EAdata.rar`. Figure 6.6 shows an example for a deliberately poorly chosen starting point for the search. The starting point is encoded with eight


Figure 6.5: Comparison of results for HDR image "Vine Sunset". First row: starting points for the search; second row: images optimized with the EA; third row: images optimized with the algorithm of Ma et al. [61] (the algorithm of Ma et al. cannot reach our TMQI-II score within 7200 seconds for the second one).

bits per colour channel and pixel, and the dark regions in it are solidly black. The algorithm by Ma et al. is not able to restore the image content in those regions and thus converges to a suboptimal solution while the EA generates a satisfactory solution.

## 6.5 Conclusion and Discussion

To conclude, we have used an EA to solve the tone mapping problem based on maximization of TMQI-II scores. Compared to TMQI-II optimization by interleaving gradient based maximization of structural fidelity with optimization of statistical naturalness, we observe significantly reduced running times. Clearly, this is not to say that the EA is the most efficient tool for the task. However, it is remarkable


Figure 6.6: Comparison of results for HDR image "Woods" and a poorly chosen starting point. First image: starting point for the search; second image: images optimized with the EA; third image: images optimized with the algorithm of Ma et al. [61].

that an algorithm as simple as the EA can obtain results much faster than a much more sophisticated approach. By distributing the computation of the TMQI-II values of the offspring across multiple cores or obtaining parameter settings for the generic tone mapping operator by (initially) optimizing using reduced-size images, obtaining optimized tone mapped images with the simple EA requires but a few seconds.

In future work, we will consider the suitability of other image quality assessment techniques for tone mapped images as well as other image processing tasks that are commonly performed using gradient based techniques in high-dimensional spaces, but that may conceivably be solved using much lower-dimensional parametric models. Also, we would like to further analyze the influence of strategy selections, and more standard strategies such as CSA-ES will be explored.

# Chapter 7

# Conclusion

In this thesis, we introduced computational models of feature analysis on HDR scenes and feature-based objective quality assessment for tone mapped images. Furthermore, we employed the proposed quality metrics to solve tone mapping optimization problems including parameter tuning and blended tone mapping. This chapter summarizes our contributions and discusses the potential directions for future research.

In Chapter 3, we proposed an algorithm for visual saliency analysis of HDR images. Taking virtual photographs is the inverse process of generating HDR images from multiple LDR exposures, and the virtual photograph sequence has the capacity to more comprehensively reveal salient content in HDR images. Our algorithm takes variously-exposed virtual photographs and incorporates them for saliency analysis. Experimental results show our method can more reliably characterize the salient regions than existing methods. By bridging the gap between dynamic ranges, virtual photograph based algorithm can make it possible to apply conventional feature analysis methods to HDR scenes.

In Chapter 4, we focused on objective quality assessment of tone mapped images and proposed two feature-based quality metrics. Compared with typical IQA methods built on the magnitude of intensity and normalized contrast, the quality metrics measure the distortion of important image features that affect the perceived quality of human observers. Based on the assumption that regions of interest predicted by bottom-up visual attention models should be preserved during tone mapping, we proposed the visual saliency distortion predictor (VSDP) that measures the distortion in visual saliency for quality assessment. Moreover, by combining brightness distortion, visual saliency distortion, and detail distortion in light and dark areas, we present the perceptual distortion predictor (PDP) for more comprehensive evaluation. The subjective and numerical experiments have verified the effectiveness of the feature-based quality metrics and their superiority to alternative works.

In Chapter 5, we employed the proposed quality metrics to solve the problems of tone mapping optimization. By minimizing visual saliency distortion, we developed an automatic parameter tuning algorithm that can optimize the parameters of arbitrary TMOs. We showed that the minimization can be accomplished with an evolutionary algorithm (EA) with individuals represented as parameters and the fitness value as saliency distortion. Experiments show that our method can significantly improve the quality of tone mapped images by generating results with more faithful appearance, and the improvement is validated with systematic statistical analysis. Also, using the perceptual distortion predictor, we introduced an automated blended tone mapping algorithm to leverage the strengths of different operators for particular HDR scenes. We demonstrate that the algorithm can produce compelling results in comparison to individual global and local operators.

In Chapter 6, we compared the EA with the gradient-based optimization algorithm using a common platform of tone mapping optimization. With a generic TMO, we have used the EA for TMQI-II based optimization. Compared to the gradient-based algorithm carried out with the introduction of TMQI-II which interleaves the optimization of structural fidelity and statistical naturalness, we observe significantly reduced running times and more consistent results.

## 7.1 Future Research Directions

The research described in the thesis represents several attempts for objective quality metrics based tone mapping optimization, and they can be further improved and extended in many directions. In addition to the future works discussed in each chapter, we hereby provide some more general ideas for future research.

Currently, most objective quality assessment approaches [108, 35, 61] are designed for the general purpose of tone mapping, which could be described as reproducing the visual appearance of HDR images during contrast compression [86]. Nevertheless, the quality criteria should be adaptive for specialized applications. For instance, in photograph, naturalness or overall fidelity may be more important, while for medical images, detail preservation might be more critical. It would be useful to analyze the requirements of different applications and incorporate them into the design of quality metrics.

The feature-based quality metrics are proposed for HDR images, and their extension can be the quality evaluation of tone mapped videos. In recent years, with the development of capture techniques, HDR-video tone mapping is becoming more and more feasible. Several several subjective studies have been conducted to evaluate existing TMOs for HDR-video [25, 68], but there are few work that focus on objective quality assessment and the corresponding parameter optimization [107]. Compared with static HDR images, HDR video sequences pose new challenges for quality evaluation. It is necessary to carry out not only spatial measures but also temporal measures that can capture artifacts, such as flickering, inconsistent brightness and contrast, and noise [25].

Another interesting direction is tone mapping optimization for HDR videos. With both spatial and temporal quality assessment available, a possible method is to collect the key frames with unsupervised classification, apply the optimization on the key frames with spatial quality measures, and then spread the optimized results to all the frames by optimizing the temporal quality measures. Nevertheless, the quality evaluation and the algorithm details of video optimization need to be carefully addressed in the future.

## Bibliography


[1] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics*, 24(3):828–835, 2005.

[2] A.O. Akyüz, K. Hadimli, M. Aydinlilar, and C. Bloch. Style-based tone mapping for hdr images. In *Proceedings of the SIGGRAPH Asia 2013 Technical Briefs*, SA '13, pages 23:1–23:4. ACM, 2013.

[3] D. V. Arnold. Resampling versus repair in evolution strategies applied to a constrained linear problem. *Evolutionary Computation*, 21(3):389–411, 2013.

[4] D. V. Arnold and H.-G Beyer. Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49(4):617–622, 2004.

[5] M. Ashikhmin. A tone mapping algorithm for high contrast images. In *Proceedings of the 13th Eurographics Workshop on Rendering*, pages 145–156. Eurographics Association, 2002.

[6] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, 26(3):10:1–10:10, 2007.

[7] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. Seidel. Dynamic range independent image quality assessment. *ACM Transaction on Graphics*, 27(3):69:1–69:10, 2008.

[8] S. Bae, S. Paris, and F. Durand. Two-scale tone management for photographic look. *ACM Transaction on Graphics*, 25(3):637–645, July 2006.

[9] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), 2011.

[10] M. Barkowsky and P. Le Callet. On the perceptual similarity of realistic looking tone mapped high dynamic range images. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, 2010.

[11] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 35(1):185–207, 2013.

[12] A.C. Bovik. *Handbook of image and video processing*. Academic Press, 2010.

[13] R. Brémond, J. Petit, and J. P. Tarel. Saliency maps of high dynamic range images. In *Proceedings of the Trends and Topics in Computer Vision — ECCV 2010 Workshops*, pages 118–130. Springer Verlag, 2012.

[14] N. D. B. Bruce and Tsotsos. J.K. An attentional framework for stereo vision. In *Proceedings of the Canadian Conference on Computer and Robot Vision*, 2005.

[15] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi. Image attributes and quality for evaluation of tone mapping operators. In *Proceedings of the 14th Pacific Conference on Computer Graphics and Applications*, pages 35– 44, Taipei, Taiwan, 2006.

[16] S. B. Chisholm, D. V. Arnold, and S. Brooks. Tone mapping by interactive evolution. In *Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation – GECCO 2009*, pages 515–522. ACM Press, 2009.

[17] S. Daly. Digital images and human vision. chapter The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity, pages 179–206. MIT Press, 1993.

[18] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques — SIGGRAPH '97*, pages 369–378. ACM Press, 1997.

[19] R. Desmone and J. Duncan. Neural mechanisms of selective visual attention. *Annual Reviews of Neuroscience*, 18:193–222, 1995.

[20] J.E. Dowling. *The retina: an approach part of the brain.* Cambridge, MA: Belknap Press, 1987.

[21] F. Drago, W. Martens, K. Myszkowski, and H-P. Seidel. Perceptual evaluation of tone mapping operators with regard to similarity and preference. Research Report MPI-I-2002-4-002, Max-Planck-Institut für Informatik, 2002.

[22] F. Drago, W. Martens, K. Myszkowski, and H-P. Seidel. Perceptual evaluation of tone mapping operators. In *Proceedings of the ACM SIGGRAPH 2003 Sketches &Amp; Applications*, SIGGRAPH '03, pages 1–1. ACM, 2003.

[23] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22:419–426, 2003.

[24] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 257–266, New York, NY, USA, 2002. ACM.

[25] G. Eilertsen, R. Wanat, R. Mantiuk, and J. Unger. Evaluation of tone mapping operators for hdr-video. *Computer Graphics Forum*, 32(7):275–284, 2013.

[26] M. Fairchild. High dynamic range data @ONLINE. http://www.cis.rit.edu/fairchild/HDR.html, 2008.

[27] M. D. Fairchild and G. M. Johnson. Meet iCAM: A next-generation color appearance model. In *Proceedings of the IS&T/SID 10th Color Imaging Conference*, pages 33–38, 2002.

[28] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, July 2002.

[29] F. Fraundorfer and H. Bischof. Utilizing saliency operators for image matching. In *International Workshop on Attention and Performance in Computer Vision*, pages 17–24, 2003.

[30] S. Frintrop. *VOCUS: A visual attention system for object detection and goal-directed search*. Springer Verlag, 2006.

[31] S. Frintrop, E. Rome, and H. I. Christensen. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 7(1):6:1–6:39, 2010.

[32] X. Gao, S. Brooks, and D. V. Arnold. A feature-based quality metric for tone mapped images. *ACM Transaction on Applied Perception*.

[33] X. Gao, S. Brooks, and D. V. Arnold. Virtual photograph based saliency analysis of high dynamic range images. In *Proceedings of the Symposium on Computational Aesthetics*, CAE '13, pages 87–92. ACM, 2013.

[34] X. Gao, S. Brooks, and D. V. Arnold. Saliency-based parameter tuning for tone mapping. In *Proceedings of the 11th European Conference on Visual Media Production*, CVMP '14. ACM, 2014.

[35] X. Gao, S. Brooks, and D. V. Arnold. Automated parameter tuning for tone mapping using visual saliency. *Computer & Graphics*, 52:171–180, 2015.

[36] X. Gao, S. Brooks, and D. V. Arnold. Automatic blended tone mapping through evolutionary optimization. In *Proceedings of IEEE Congress on Evolutionary Computation)*, CEC' 16. IEEE, 2016.

[37] X. Gao, S. Brooks, and D. V. Arnold. Evolutionary optimization of tone mapped image quality index. In *Proceedings of Evolution Artificielle*, EA 17, 2017.

[38] X. Gao, S. Brooks, and D. V. Arnold. A feature-based quality metric for tone mapped images. In *Proceedings of ACM Symposium on Applied Perception*, SAP' 17. ACM, 2017.

[39] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang. Details preservation inspired blind quality metric of tone mapping methods. In *Proceedings of IEEE International Symposium on Circuits and Systems*, pages 518–521, 2014.

[40] N. Hansen, D. V. Arnold, and A. Auger. *Evolution strategies.* Handbook of Computational Intelligence. Springer, 2015.

[41] G. Heidemann, R. Rae, H. Bekel, I. Bax, and H. Ritter. Integrating context-free and context-dependent attentional mechanisms for gestural object reference. *Machine Vision and Applications*, 16(1):64–73, 2004.

[42] L. Itti. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of Technology, 2000.

[43] L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, 13(10):1304–1318, 2004.

[44] L. Itti. Visual salience. *Scholarpedia*, 2(9):3327, 2007.

[45] L. Itti, N. Dhavale, and F. Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention. In *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology*, 2003.

[46] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40:1489–1506, 2000.

[47] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[48] M. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–89, 1938.

[49] E. I. Knudsen. Fundamental components of attention. *Annual Review Neuroscience*, 30:57–78, 2007.

[50] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Neural Computation*, 17:219–227, 1985.

[51] J. Kuang, Johnson G. M., and M. D. Fairchild. icam06: A refined image appearance model for hdr image rendering. *Journal of Visual Communication and Image Representation*, 18:406–414, 2007.

[52] J. Kuang, H. Yamaguchi, C. Liu, G. M. Johnson, and M. D. Fairchild. Evaluating hdr rendering algorithms. *ACM Transactions on Applied Perception*, 4(2), July 2007.

[53] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[54] G. W. Larson, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3(4):291–306, 1997.

[55] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics*, 24(3):640–648, July 2005.

[56] K. Lee, H. Buxton, and J. Feng. Selective attention for cue-guided search using a spiking neural network. In *International Workshop on Attention and Performance in Computer Vision (WAPCV)*, pages 55–62, 2003.

[57] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski. Interactive local adjustment of tonal values. *ACM Transactions on Graphics*, 25(3):646–653, 2006.

[58] H. Liu and I. Heynderickx. Visual attention in objective image quality assessment: Based on eye-tracking data. *IEEE Transactions on Circuits and Systems for Video Technology*, 21:971–982, 2011.

[59] X. Liu, L. Zhang, H. Li, and J. Lu. Integrating visual saliency information into objective quality assessment of tone-mapped images. In *Proceedings of International Conference on Intelligent Computing*, pages 376–386, 2014.

[60] M. Livingstone. *Vision and Art: The Biology of Seeing*. Harry N. Abrams, 2002.

[61] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang. High dynamic range image compression by optimizing tone mapped image quality index. *IEEE Transactions on Image Processing*, 24(10):3086–3097, 2015.

[62] A. Maki, P. Nordlund, and J. O. Eklundh. Attentional scene segmentation: Integrating depth and motion. *Computer Vision and Image Understanding*, 78(3):351–373, 2000.

[63] S. Mann and R. W. Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of the 48th Annual IS&T Conference*, pages 422–428. Society for Imaging Science and Technology, 1995.

[64] Myszkowski K. Mantiuk, R. and H. P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception*, 3(3):286–308, 2006.

[65] R. Mantiuk, S. Daly, and L. Kerofsky. Display adaptive tone mapping. *ACM Transaction on Graphics*, 27(3):68:1–68:10, 2008.

[66] R. Mantiuk, S. Daly, K. Myszkowski, and H. P. Seidel. Predicting visible differences in high dynamic range images - model and its calibration. In *Proceedings of Human Vision and Electronic Imaging X, IS&T/SPIEs 17th Annual Symposium on Electronic Imaging*, pages 204–214, 2005.

[67] R. Mantiuk and H. P. Seidel. Modeling a generic tone-mapping operator. *Computer Graphics Forum*, 27(2):699–708, 2008.

[68] M. Melo, M. Bessa, K. Debattista, and A. Chalmers. Evaluation of hdr video tone mapping for mobile devices. *Signal Processing Image Communication*, 29(2):247–256, 2013.

[69] T. Mertens, J. Kautz, and F. V. Reeth. Exposure fusion. In *Proceedings of the15th Pacific Conference on Computer Graphics and Application*, pages 382–390, 2007.

[70] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):349–361, 2001.

[71] J. L Myers and Well A. D. *Research Design and Statistical Analysis*. Lawrence Erlbaum, 2003.

[72] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, and M. Cheriet. Fsitm: A feature similarity index for tone-mapped images. *IEEE Signal Processing Letters*, 22(8):1026–1029, 2015.

[73] Y. Nagai. Learning to predict where humans look. In *Proceedings of the IEEE 8th International Conference on Development and Learning*, 2009.

[74] M. Narwaria, M. Perreira Da Silva, P. Le Callet, and R. Pépion. Effect of tone mapping operators on visual attention deployment. In *SPIE Proceedings Vol. 8499 — Applications of Digital Image Processing XXXV*. International Society for Optics and Photonics, 2012.

[75] H. R. Nasrinpour and N. D. Bruce. Saliency weighted quality assessment of tone-mapped images. In *Proceedings of IEEE International Conference on Image Processing*. 2015.

[76] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi. Visual attention in ldr and hdr images. In *Proceedings of the 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2015.

[77] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer-Verlag, 2006.

[78] B. Olshausen and D. J. Field. How close are we to understanding v1? *Human Neurobiology*, 4(8):1665–1699, 2005.

[79] Tumblin J. Yee H. Pattanaik, S. N. and D. P. Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pages 47–54, 2000.

[80] S. E. Ralmer. *Vision Science, Photons to Phenomenology*. MIT Press, 1999.

[81] I. Rechenberg. *Evolutionsstrategie — Optimierung technischer steme nach Prinzipien der biologischen Evolution.* Friedrich Frommann Verlag, 1973.

[82] E. Reinhard. Parameter estimation for photographic tone reproduction. *Journal of Graphics Tools*, 7(1):45–52, 2002.

[83] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics*, 11(1):13–24, January 2005.

[84] E. Reinhard, T. Pouli, T. Kunkel, B. Long, A. Ballestad, and G. Damberg. Calibrated image appearance reproduction. *ACM Transactions on Graphics*, 31(6):1–11, November 2012.

[85] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Transactions on Graphics*, 21(3):267–276, 2002.

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

[87] C. Richardt. Flash-exposure high dynamic range imaging: Virtual photograph and depth-compensating flash. Technical Report UCAM-CL-TR-712, University of Cambridge, 2008.

[88] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. In *Proceedings of Photorealistic Rendering Techniques*, pages 7–20. Springer Verlag, 1994.

[89] H. P. Schwefel. *Evolution and Optimum Seeking.* Wiley, 1995.

[90] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang. Probabilistic exposure fusion. *IEEE Transaction on Imgage Processing*, 21(1):341–357, January 2012.

[91] L. Stroebel, J. Compton, Current. I., and R. Zakia. *Basic Photographic Material and Process, second edition.* Focal Press, 2000.

[92] A. Treisman and S. Gormican. A feature integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.

[93] A. Treisman and S. Gormican. Feature analysis in early vision: evidence from search asymmetries. *Psychology Review*, 95:15–48, 1988.

[94] J.K. Tsotsos, Y. Liu, J. C. Martinez-Trujillo, M. Pomplun, E. Simine, and K. Zhou. Attending to visual motion. *Journal of Computer Vision and Image Understanding*, 100:3–40, 2005.

[95] J. Tumblin and H. Rushmeier. Tone reproduction for computer generated image. *IEEE Computer Graphics and Application*, 13(6):42–48, 1993.

[96] S. Vijayakumar, J. Conradt, T. Shibata, and S. Schaal. Overt visual attention for a humanoid robot. In *Proceedings of International Conference on Intelligence in Robotics and Autonomous Systems*, pages 2332–2337, 2001.

[97] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[98] D. Walther. *Interactions of visual attention and object recognition: computational modeling, algorithms, and psychophysics.* PhD thesis, California Institute of Technology, 2006.

[99] Y. S. Wang, C. L. Tai, O. Sorkine, and T. Y. Lee. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics*, 27(5):1–8, 2008.

[100] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool Publishers, 2006.

[101] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, Jan 2009.

[102] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[103] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *Conference Record of the37th Asilomar Conference on Signals, Systems and Computers*, 2004.

[104] G. J. Ward. A contrast-based scalefactor for luminance display. In *Proceedings of Graphics Gems IV*, pages 415–421. Academic Press Professiona, 1994.

[105] Kandler K. Fitzpatrick D. Weliky, M. and L. C. Katz. Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron*, 15:541–552, Oct 1995.

[106] H Yeganeh. *Cross Dynamic Range And Cross Resolution Objective Image Quality Assessment With Applications*. PhD thesis, University of Waterloo, 2014.

[107] H. Yeganeh, S. Wang, K Zeng, and Z. Wang. Objective quality assessment of tone-mapped videos. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2016.

[108] H. Yeganeh and Z. Wang. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing*, 22(2):657–667, Feb 2013.

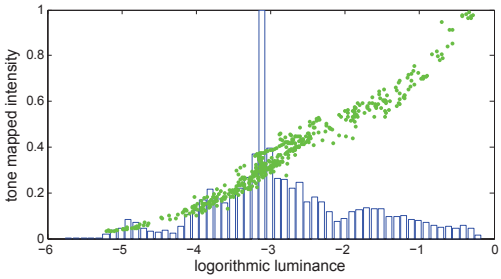
[109] A. Yoshida, R. Mantiuk, K. Myszkowski, and H.-P. Seidel. Analysis of reproducing real-world appearance on displays of varying dynamic range. *Computer Graphics Forum*, 25(3):415–426, 2006.

[110] B. Zenger and D. Sagi. Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection. *Vision Research*, 36:2497–2513, 1996.

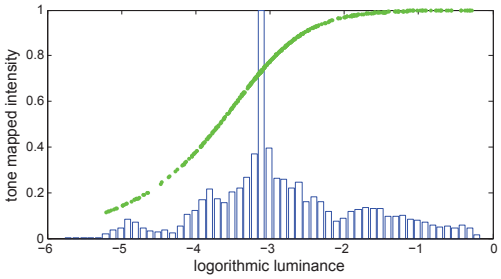
# Appendix A

## Comparison of Tone Mapped Images and Their Tone Mapping Curves

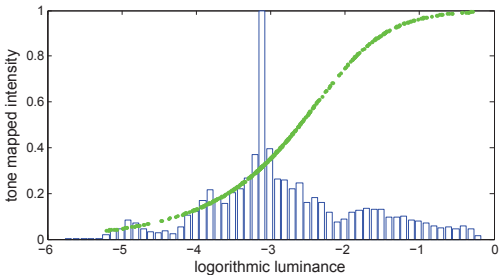
The proposed quality metric makes the quality prediction based on the distortions in image features, and the similarity to the transfer curves used for taking virtual photographs cannot guarantee good quality of the tone mapped images. Figure A.1 illustrates the comparison of several tone mapped images and their tone mapping curves. The first row is the reference image generated with iterative tone mapping optimization, which could be used as the benchmark. The second and third rows are tone mapped images generated by the sigmoidal curves that are very similar to what we use for taking virtual photographs. As shown in the example, these similar curves don't necessarily produce satisfying results. Compared with the benchmark, the tone mapped image generated by Schlick's operator [88] (Second row) suffers from more distortion in all the feature-based measures, while the image generated by Reinhard et al.'s operator [85] (Third row) does not perform well for detail reproduction in light areas.




Reference Image
Brightness Distortion: 0.0007
Visual Saliency Distortion: 0.0193
Detail Distortion in Light Areas: 0.5237
Detail Distortion in Dark Areas: 0.3202




Tone Mapped Image (Schlick's Operator)
Brightness Distortion: 0.3231
Visual Saliency Distortion: 0.0994
Detail Distortion in Light Areas: 0.6951
Detail Distortion in Dark Areas: 0.3341




Tone Mapped Image (Reinhard et al.'s Operator)
Brightness Distortion: 0.0332
Visual Saliency Distortion: 0.0139
Detail Distortion in Light Areas: 0.6621
Detail Distortion in Dark Areas: 0.3271

Figure A.1: Comparison of tone mapped images and their tone mapping curves. First row: reference image generated with tone mapping optimization. Second row: tone mapped image generated by Schlick's operator with default parameter settings. Third row: tone mapped image generated by Reinhard et al.'s operator with default parameter settings. The HDR image is from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

# Appendix B

## Comparison with Visual Saliency Analysis on Log-HDR Values for Parameter Optimization

In the visual saliency based parameter optimization, we employ the virtual photograph based method [35] to analyze visual saliency on HDR images, which cannot replaced by simple computing visual saliency on log-HDR values. Figure B.1 presents a comparison between the virtual photograph based method and applying the visual attention model proposed by Itti and Koch [46] on log-HDR values (log-HDR method) for parameter optimization. For each HDR image, we generate 25 visually different tone mapped images with Schlick's operator [88], and then use the saliency distortion calculated from these two methods to select the best tone mapped image. As illustrated in the examples, saliency distortion calculated from the virtual photograph based method yields more reliable quality predictions. We hypothesize that this is because virtual photograph sequence has the capacity to more comprehensively and accurately preserve image content of HDR images for visual saliency analysis than log-HDR values.

For a thorough evaluation, we apply the parameter tuning of Schlick's operator [88] for the image database provided by Mark D. Fairchild [17], in which 105 HDR images and the corresponding reference images are provided. The saliency distortion calculated from virtual photograph based method and log-HDR method are adopted for automatic parameter tuning. With the reference images as the benchmark, the performance of the optimized parameter settings are measured in terms of PSNR and SSIM. We conduct a Mann-Whitney U test for each image, and then apply the outputs for a Wilcoxon signed-rank test to measure the overall difference. As shown in Table B.1, virtual photograph based method performs significantly better than log-HDR method.


Reference image | Saliency Distortion: 0.1831 Log-HDR Saliency Distortion: 0.9126 | Saliency Distortion: 0.3823 Log-HDR Saliency Distortion: 0.7086

Reference image | Saliency Distortion: 0.5136 Log-HDR Saliency Distortion: 1.0090 | Saliency Distortion: 0.5478 Log-HDR Saliency Distortion: 0.8561


Reference image | Saliency Distortion: 0.4994 Log-HDR Saliency Distortion: 3.2805 | Saliency Distortion: 3.3241 Log-HDR Saliency Distortion: 1.7334

Reference image | Saliency Distortion: 0.2471 Log-HDR Saliency Distortion: 1.4305 | Saliency Distortion: 0.2895 Log-HDR Saliency Distortion: 0.5394

Figure B.1: Comparison between virtual photograph based method and log-HDR method for parameter optimization. First column: reference images tone mapped with manual manipulation (provided at [17]). Second column: the best tone mapped images selected based on saliency distortion calculated from virtual photograph based method (denoted as saliency distortion). Third column: the best tone mapped images selected based on saliency distortion calculated from log-HDR method (denoted as log-HDR saliency distortion). The HDR images are from Mark D. Fairchild's HDR Photographic Survey and © 2006-2007 Mark D. Fairchild.

Table B.1: Comparison between virtual photograph based method and log-HDR method for parameter tuning.

| Quality metrics | Our method | Log-HDR method | Ties | p Value |
|---|---|---|---|---|
| PSNR | 61 | 19 | 25 | 0.000 |
| SSIM | 61 | 16 | 28 | 0.000 |