VISUAL SIMILARITY FOR HDR IMAGES WITH APPLICATIONS TO TONE MAPPING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MERVE AYDINLILAR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

FEBRUARY 2021

Approval of the thesis:

VISUAL SIMILARITY FOR HDR IMAGES WITH APPLICATIONS TO TONE MAPPING

submitted by **MERVE AYDINLILAR** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Halit Oğuztüzün
Head of Department, **Computer Engineering**

Prof. Dr. Ahmet Oğuz Akyüz
Supervisor, **Computer Engineering, METU**

Prof. Dr. Sibel Tarı
Co-supervisor, **Computer Engineering, METU**

Examining Committee Members:

Prof. Dr. Tolga Kurtuluş Çapın
Computer Engineering, TED University

Prof. Dr. Ahmet Oğuz Akyüz
Computer Engineering, METU

Prof. Dr. Mine Özkar
Architecture, ITU

Assoc. Prof. Dr. Sinan Kalkan
Computer Engineering, METU

Assist. Prof. Dr. Gökberk Cinbiş
Computer Engineering, METU

Date: 15.02.2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname:	Merve Aydınlılar
Signature :	

ABSTRACT

VISUAL SIMILARITY FOR HDR IMAGES WITH APPLICATIONS TO TONE MAPPING

Aydınlılar, Merve
Ph.D., Department of Computer Engineering
Supervisor: Prof. Dr. Ahmet Oğuz Akyüz
Co-Supervisor: Prof. Dr. Sibel Tarı

February 2021, 115 pages

Assessing visual similarity between images is important for many computer vision applications. So far, investigations on visual similarity have been confined to low dynamic range images. However, recently, there is a growing interest to high dynamic range (HDR) imaging. In this thesis, the aim is to shed light on visual image similarity for HDR images by following an experimental approach. To this end, a user experiment is conducted through a novel web-based interface, in which the participants assess the pairwise similarity of HDR images. The data collected through this experiment is used to evaluate a set of image features with respect to their correlations with the participants responses. A combined feature as a linear combination of individual features is defined, and its coefficients are learned using metric learning. The learned combined feature as compared to individual features correlated better with the participants' responses. Among individual features, deep learning features are found to correlate better than others, supporting that higher level features are better than lower level ones. Using the learned similarity, *i.e. combined feature*, a style-based tone mapping algorithm is proposed that successfully imparts a user-defined style to various HDR images determined to be similar with respect to the proposed metric.

Keywords: HDR Imaging, HDR Image Similarity, HDR Tone Mapping

ÖZ

YDA İMGELERDE GÖRSEL BENZERLİK VE TON EŞLEMEYE UYGULAMALARI

Aydınlılar, Merve
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Ahmet Oğuz Akyüz
Ortak Tez Yöneticisi: Prof. Dr. Sibel Tarı

Şubat 2021 , 115 sayfa

İmgeler arası görsel benzerlik birçok bilgisayarlı görme uygulaması için yüksek öneme sahiptir. Bu konu üzerine şimdiye kadar yapılan çalışmalar düşük dinamik aralıklı imgelerle sınırlı kalmıştır. Ancak, günümüzde yüksek dinamik aralıklı (YDA) görüntülemeye ilgi her geçen gün artmaktadır. Bu tezde, YDA imgeleri arasındaki görsel benzerliğe deneysel bir yaklaşımla ışık tutulması hedeflenmiştir. Bu amaçla, kullanıcıların YDA imgelerin benzerliklerini değerlendirmelerine olanak veren web tabanlı bir deney arayüzü oluşturulmuş ve internet üzerinden ulaşılan yüksek sayıda kullanıcıdan veri toplanmıştır. Toplanan veriyle bir grup imge özniteliğinin kullanıcı verisiyle korelasyonu değerlendirilmiştir. Bu özniteliklerin lineer kombinasyonuyla birleşik bir öznitelik tanımlanmış ve bu özniteliğin katsayıları metrik öğrenme ile belirlenmiştir. Bu öğrenilmiş birleşik özniteliğin diğer özniteliklere nazaran kullanıcı verisiyle korelasyonunun daha yüksek olduğu gözlemlenmiştir. Tekil öznitelikler arasında derin öğrenmeden elde edilen özniteliklerin kullanıcı verisiyle diğerlerinden daha iyi korale etmesi, yüksek seviye özniteliklerin düşük seviye özniteliklere göre daha başarılı olmasını desteklemektedir. Elde edilen benzerlik özniteliği ile, stil bazlı ton eşlemesi algoritması geliştirilmiştir. Bu ton eşleme algoritmasıyla yaratılan stilin imgeler arası benzerlik kullanılarak başarılı bir şekilde YDA imgelere uygulandığı gösterilmiştir.

Anahtar Kelimeler: YDA Görüntüleme, YDA İmge Benzerliği, YDA Ton Eşlemesi

To my grandmother

ACKNOWLEDGMENTS

First of all, I would like to thank my advisors Prof. Dr. Ahmet Oğuz Akyüz and Prof. Dr. Sibel Tarı for their guidance and encouragement during this thesis and also before that as my lecturers. Needless to say, without their supervision this thesis would not be possible. I am deeply grateful for the opportunity to work with them. These both challenging and exciting years made me a better person both academically and personally thanks to them.

I can not thank enough to my thesis progress committee members Prof. Dr. Mine Özkar and Assoc. Prof. Dr. Sinan Kalkan for their guidance and support over the years in the course of this thesis. Also I would like to thank my thesis defense committee members Prof. Dr. Tolga Kurtuluş Çapın and Assist. Prof. Dr. Gökberk Cinbiş for their valuable comments. Also, I would like to thank Assist. Prof. Dr. Serdar Çiftçi for his participation to my thesis defense.

I am grateful for the support and encouragement of my friends that I worked with during my system administration years, Özgür Kaya, Hüsnü Yıldız, Gökdeniz Karadağ, İlkcan Keleş, Can Eroğul and especially Adnan Kılıç who also helped to materialize this thesis. Working with you all has always been a pleasure.

I also would like to thank my friends Utku Erdoğdu, Elvan Gülen, Okan Akalın, Can Hoşgör, Ömer Ekmekçi, Kazım Işık, Yunus Olgun, Tuba Demirtaş and Aslıhan Örüm. Their friendship means a lot to me. And I would particularly like to thank Kerem Hadımlı with whom I started working on my doctoral studies.

Lastly, I would like to thank my parents for their support during this thesis. My grandmother, whom I dedicated this thesis, had always believed in me and wanted me to pursue higher education. I am thankful for the encouragement I received from Christian and his parents. Special thanks to my dearest sister Melike who I am so lucky to have. Also, to Zahter, who brings joy to our lives everyday.

TABLE OF CONTENTS

CHAPTERS

APPENDICES

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	
1D	1 Dimensional
2D	2 Dimensional
2AFC	2 Alternative Forced Choice
BF	Bileteral Filter
CNN	Convolutional Neural Network
DCNN	Deep Convolutional Neural Network
EMD	Earth Mover’s Distance
HDR	High Dynamic Range
HSV	Hue, Saturation, Value
HVS	Human Visual System
LDR	Low Dynamic Range
MOS	Mean Opinion Scores
SIFT	Scale-Invariant Feature Transform
SURF	Speeded Up Robust Features
TMO	Tone Mapping Operator
UM	Unsharp Masking

CHAPTER 1

INTRODUCTION

With the developments in digital imaging, the means of producing digital images became widely available. This led to a vast amount of images and a diverse set of applications. For several image related applications, including image retrieval and indexing [5], classification and clustering [6], image editing and style transfer [7], assessing the visual similarity of a pair of images is important. Due to its importance, significant amount of research is dedicated to measuring image similarity. Depending on the context of the application, image similarity might be defined differently hence yielding different measures and techniques.

Typically, it is hard to estimate a certain degree of similarity between two images without a given context. Unlike some computer vision tasks like depth estimation or object detection, image similarity does not have a ground truth. The lack of ground truth makes the research of image similarity harder in two aspects, i.e., learning models from the ground truth data directly and measuring the performance of the proposed image similarity method reliably. This makes subjective experiments highly valuable for the image similarity research.

The need for High Dynamic Range (HDR) imaging was realized in computer graphics in order to deal with the requirements of the physically accurate lighting simulation systems [8]. Such systems produced numerically unbounded pixel values, necessitating their storage in HDR formats [9]. HDR images are typically been termed as “scene-referred” as opposed to “display-referred” – a term used for Low Dynamic Range (LDR) images [10]. However, as display devices have traditionally been low dynamic range, displaying these images on LDR devices requires tone mapping [11, 12], i.e., the compression and quantization of luminance from high dynamic range to low dynamic range. Numerous tone mapping operators (TMOs) have been developed in literature ranging from simple contrast adjustments to complex algorithms modeling the human visual system [13] and the properties of display devices [14]. These operators are developed with different motivations hence yield perceptually and statistically different images. Many operators have also been produced to create photographic HDR images of real-world scenes [15], including dynamic scenes [16, 17]. Besides computer graphics applications, HDR imaging has other applications including studying of fossils [18], cultural heritage and archaeology [19], structural engineering [20], architecture [21], medical imaging [22, 23], forensics [24], and automotive industry [25].

While the majority of the produced and stored images are LDR images, HDR imaging offers to capture much wider luminance range, closer to the human eye, compared to standard LDR imaging. Currently, HDR video formats HDR10+ [26] and Dolby Vision [27] are becoming the main HDR video standards and the number of HDR compatible commercial TVs, Blu-ray devices, gaming consoles and mobile devices supporting these video format standards is increasing. Besides, lately main streaming platforms like Amazon Prime Video and Netflix started serving HDR video content. These progress on the standardization and the developed devices will increase the demand in HDR content production very soon. Additionally, with the emerging technologies that require to exceed the human visual system limitations such as autonomous driving, HDR imaging gains more importance in safety critical computer vision applications.

Many research problems and applications are common to both HDR and LDR imaging. However, LDR images consist of 8 bit integers, whereas HDR images are represented with floating point data and if they have not been calibrated a single pixel intensity may refer to completely different illumination values on different scenes. Therefore, using the solutions and methods that are originally devised for standard LDR images may not be feasible for HDR images.

1.1 Problem Definition

Since image similarity is a perceptual phenomenon without ground truth, there are several studies that conducted human experiments. All these works, however, assume that the image is given in standard low dynamic range format where the brightness information is suitably quantized to match the dynamic range of traditional image display devices. Growing number of applications utilize High Dynamic Range (HDR) images with unbounded brightness values. HDR images contain potentially uncalibrated floating point data and two images that have vastly different pixel values may actually be very similar to each other. The richness of information in an HDR image, despite causing difficulties, may aid in similarity assessment. For example, pixel values corresponding to a bright light source can be much higher than that of a white reflecting surface in an HDR image, while the two objects are likely to map to similar intensities in an LDR image. Hence, there is a need for investigating visual similarity for HDR images. It may be argued that HDR images pose no extra challenges and approaches designed for LDR images may directly be used to assess visual similarity of HDR images. However, using a standard similarity measure for an HDR image requires tone mapping, a problem for which a multitude of algorithms, each with a number of parameters, exist [28].

Investigating the assessment of visual similarity between two HDR images is the motivation of this thesis. To this end, subjective human judgments using crowdsourcing is collected and features are evaluated by comparing them to human judgments. To our knowledge this thesis serves as the first rigorous attempt to evaluate how visual similarity can be assessed between HDR images. Using the findings from the perceptual study, a tone mapping methodology is proposed where tone mapping parameters are automatically computed to impart a certain user-defined style to a given HDR image using the similarity between this image and several calibration images that are used to create this style.

1.2 Contributions and Outline of the Thesis

The motivation of this thesis is to investigate the visual similarity for HDR images and demonstrate an application that benefits from these investigations. The main contributions of this thesis are listed below:

- A novel web-based interface to conduct a perceptual similarity experiment for HDR images, and the assessment data collected through crowdsourcing,
- Analysis on the experiment data, in the sense of the agreement of various image features and the corresponding distance metrics to the user responses, besides examining the effect of tone mapping operators,
- A similarity model that consists of combination of image features with contribution of each feature estimated by user experiment data,
- Two different methodologies to improve the presented tone mapping operator by employing the findings obtained by user experiment,
- A tone mapping methodology, namely style-based tone mapping, that uses image similarity to automatically estimate the tone mapping parameters from the tone mapping parameters of a set of given images to follow a certain style.

The organization of the rest of this thesis is as follows, in Chapter 2, a review of HDR imaging and image similarity is given. In Chapter 3, the image features and the corresponding distance metrics used in the rest of the thesis are given. Due to the subjective nature of the problem, an experimental study that assesses visual similarity between HDR images is conducted. In Chapter 4, the experimental setup and the details about data gathering is given. The analyses on the user experiment such as the correlation of the image features and the corresponding distance metrics to the user responses, and the effect of the tone mapping operators to these analyses are given in this chapter. In addition, two combined models that are estimated using experiment data are also given. Style-based tone mapping operator is presented in detail in Chapter 5 with some example styles replicated with the developed tool and the tone mapping results following these styles. After that, two methods that improves the style-based tone mapping operator by using the results obtained through user experiment are presented.

Lastly, in Chapter 6, discussion about the findings and the results are given and the summary of the thesis, several limitations and future directions are given in Chapter 7.

CHAPTER 2

RELATED WORK

2.1 HDR Imaging

While HDR imaging has long been an active field of research, recent developments in HDR imaging [10, 29, 30], in particular those pertaining to HDR image and video capture [31, 32] and display systems [33], and HDR video streaming standards [34] to allow direct rendition of HDR images will likely make HDR images more common in the near future. However, despite the practical improvements in the field, there is also a need for fundamental and experimental research that explores various aspects related to HDR imaging and dynamic range. Hanhart et al. [35] investigated the performance of various objective metrics in quantifying visual distortions of HDR images commensurate with subjective opinions. Hanhart et al. found HDR-VDP-2 [36] and HDR-VQM [37] to be the best predictors of visual quality. In another study, Grimaldi et al. [38] investigated how image statistics change as a function of dynamic range and found that there are indeed differences between HDR and LDR images. Grimaldi et al., also found, however, that the majority of these differences are accounted for by the early visual processing that takes place in the human visual system. Additionally, Rana et al. [39] compared HDR and LDR images in terms of feature detection potential and showed that it is possible to extract more effective features from HDR images. However, these works do not consider the HDR image similarity problem.

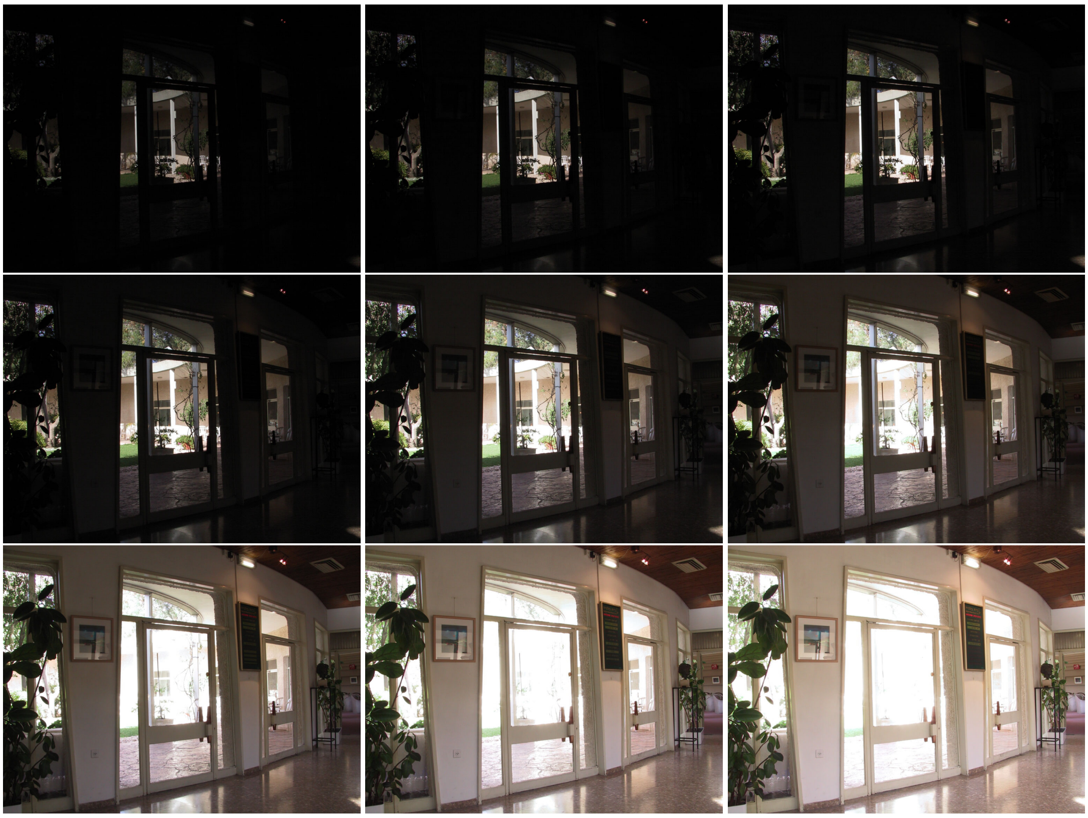


Figure 2.1: "Belgium House" scene captured with different exposures (image courtesy of Dani Lischinski).

2.1.1 HDR Image Creation

Standard camera sensors are not capable of capturing HDR images natively. Currently, there are several approaches to generate HDR images [40]. The first approach is capturing HDR images directly with specialized HDR sensors [41]. Although there has been advancements on HDR sensors, these sensors are still under development and expensive. The second and the most commonly used method is combining multiple exposure images taken by a camera with a standard sensor. This method relies on using different exposure times to capture the details of different regions of the scene, longer exposure times for low luminance areas and shorter exposure times for brighter areas. Figure 2.1 shows the images captured with different exposures of the same scene. This scene has a high dynamic range that is not possible to capture both the lower luminance (indoor) and higher luminance (garden) regions with a single exposure setting.

After the series of images are captured with the same camera, it is possible to recover irradiance, E, for each pixel, x with [40]:

$$E(x) = \frac{\sum_{i=1}^{N_e} w(I_i(x)) \frac{I_i(x)}{\Delta t_i}}{\sum_{i=1}^{N_e} w(I_i(x))} \tag{2.1}$$

where I_i is the image captured with exposure i, Δt_i is the exposure time, N_e is the number of used exposures and w is a weighting function to remove outliers. Note that camera response is assumed linear. For the cases that this assumption does not hold corrections need to be made. Dynamic scenes poses challenges for this approach, known as ghosting, and there are many deghosting algorithms to overcome this problem [42]. The third approach is using a single camera with multiple sensors to capture different exposures of the same scene simultaneously [31, 43]. However, these cameras are costly and hard to calibrate. The last approach to create HDR images is using a single LDR image and expanding the dynamic range approximately, known as inverse tone mapping. There are many studies on inverse tone mapping [44, 45, 46, 47] and this topic will gain more importance with the necessity of displaying LDR images on HDR displays arises as the usage of HDR displays increasing. Recently, there are also studies [48, 49] that employs deep neural network based approaches for inverse tone mapping.

2.1.2 Tone Mapping

HDR images can capture the true dynamic range of a scene by using wider representations compared to standard 8 bit images. However, in order to display or print HDR images on conventional devices, the dynamic range of the image should be compressed to match the dynamic range of the device. This operation, namely tone mapping, aims to compress the dynamic range while keeping the visual appearance intact.

Tone mapping is an extensively studied research area and there are many tone mapping operators [10]. Comparison and subjective evaluation of the tone mapping operators is also an active research area [50, 51]. In this thesis, to investigate the effect of tone mapping operators to visual similarity, eleven commonly used tone mapping operators are selected. Basically, these tone mapping operators can be divided into three categories depending on their operation domain: global and local operators in spatial domain, and gradient domain operators [10].

Global tone mapping operators apply the same non-linear function to all luminance values without considering pixel neighbourhood. This makes these operators fast compared to other tone mapping operators however, they may result with contrast loss in dark and bright regions. Reinhard et al. [52] introduces a transfer function that compresses low luminance values less and high luminance values more. Drago et al. [53] uses logarithmic compression for tone mapping while adaptively changing the logarithm base depending on the luminance value. A global algorithm is proposed by Pattanaik et al. [54], which takes in to account the illumination adaptation time of HSV. Other global methods that model photoreceptors of HVS is the work from Reinhard & Devlin [55] and Ferradans et al. [56]. Mantiuk et al. [57] models the display as well as HSV. On the other hand, Mai et al. [58] proposed a TMO that minimizes the mean square error between the input HDR and the result of inverse tone mapping.

Local TMOs are computationally more expensive than global TMOs but they may preserve the details better [10]. A local TMO that is inspired dodging and burning technique of traditional photography is also given in [52]. This operator darkens or brightens the pixel using Gaussian filters on different scales to increase the contrast depending on the luminance of the pixels in the neighborhood. Another local TMO, proposed by Durand & Dorsey [59], separated the image to base and detail using bilateral filter and compress only the base level to preserve the details.

There are also TMOs that operate on gradient domain rather than the spatial domain. Fattal et al. [60] suggest to reduce the magnitudes of the high gradients more than the gradients with low magnitude. Then, a low dynamic range image is obtained by solving a Poisson equation on this magnitude-wise altered gradient map. Mantiuk et al. [61] follows a similar approach with some improvements on local contrast by using a bigger neighborhood and multi level Gaussian pyramid to enhance global contrast.

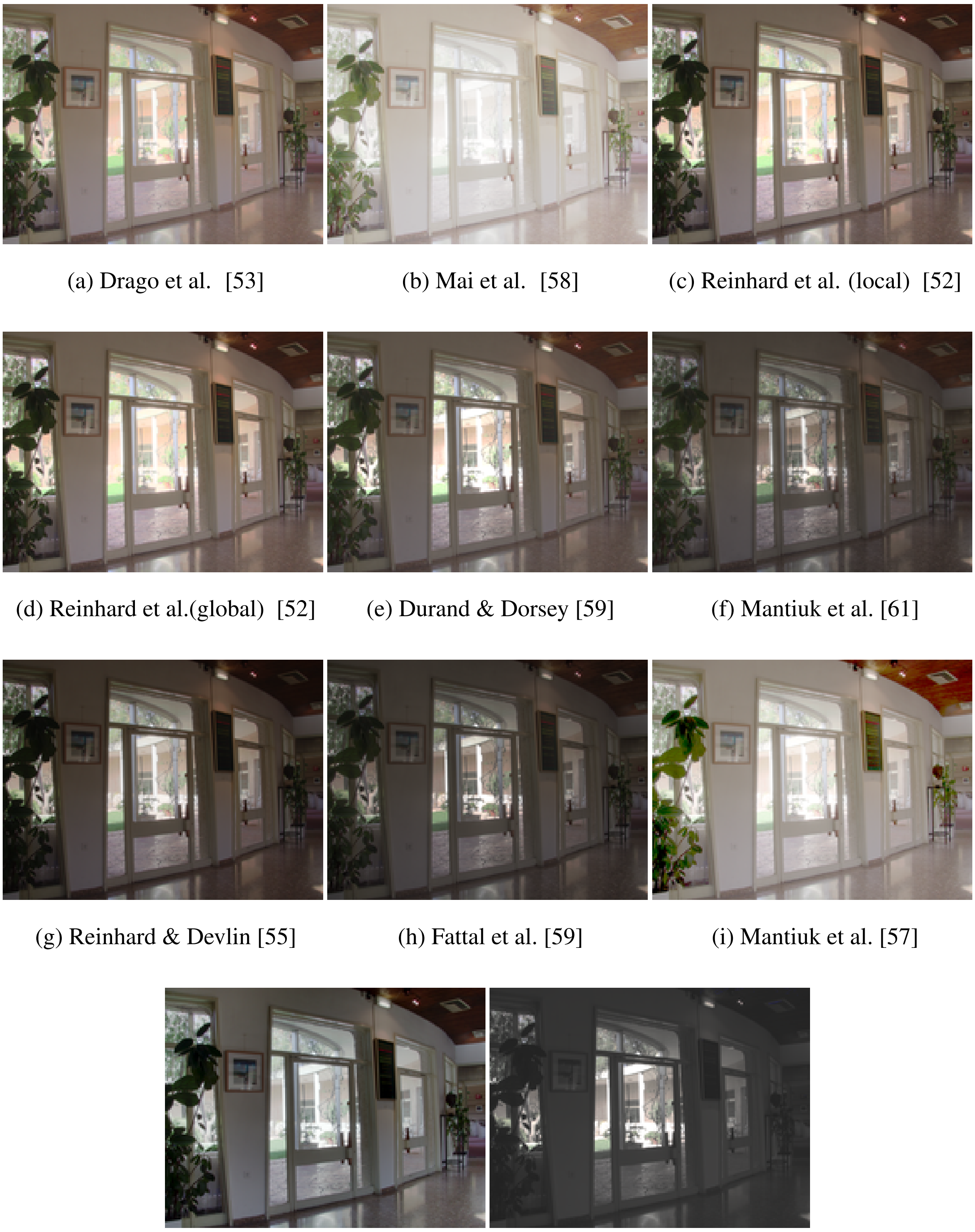
In Figure 2.2, tone mapping results for the "Belgium House" HDR image is given for several tone mapping operators. fpstools [1] implementation is used for the operators with the default parameters for each operator. As depicted in the figure, tone mapping operators give different results for the same scene in terms of image properties like brightness and color.

2.2 Image Similarity

Traditionally, image similarity is measured by measuring the distance between hand crafted features extracted from each image. These hand crafted features include simple descriptors such as color/luminance histograms, or improved ideas, including histogram of oriented gradients [62]. GIST [63], SIFT [64], SURF [65]. These features are compared using several types of distance metrics. Recently, deep convolutional neural networks (DCNNs) became the state of art for image classification. Starting with AlexNet [66] and followed by deeper networks such as VGG [67], GoogleNet[68], and ResNet [69], DCNNs started to perform near human level success for image classification. Their success lead to use feature vectors that have been obtained from DCNNs for image retrieval [70, 71, 72, 73]. Unlike previous approaches that are based on hand-crafted features, DCNNs learn the feature vector itself directly from the image.

The similarity between deep learning features can be calculated directly with euclidean or cosine distance without any learning, or can be learned. [74], [75], [76] and OASIS[77] are famous similarity learning techniques, that can be classified linear metric learning. These works proposed to work on hand-crafted image features but shown in [70] that can also work with deep features. The second group of metric learning is nonlinear metric learning methods, that learn similarity metric directly with deep neural network architectures[78]. Siamese[79][80] and triplet[81] [82] architectures minimizes the classification loss function while a more recent study [83] proposes similarity network architecture to obtain similarity score by minimizing a ranking loss function.

One major drawback of using DCNNs is the need for using very large labeled datasets for training, which is difficult to obtain or not available at all for most problem domains. Transfer learning [84] aims to solve this problem by using pretrained networks



(a) Drago et al. [53]

(b) Mai et al. [58]

(c) Reinhard et al. (local) [52]

(d) Reinhard et al.(global) [52]

(e) Durand & Dorsey [59]

(f) Mantiuk et al. [61]

(g) Reinhard & Devlin [55]

(h) Fattal et al. [59]

(i) Mantiuk et al. [57]

(j) Ferradans et al. [56]

(k) Pattanaik et al. [54]

Figure 2.2: Tone mapping results for the "Belgium House" HDR image (image courtesy of Dani Lischinski) with different tone mapping operators. Note that pfstools [1] implementation is used with default parameters.

on large scale datasets such as ImageNet [85]. The basic method is to give the images to the pre-trained network and use the output of the last fully connected layers as feature vectors [86, 70] – an approach that is also adopted in this thesis.

Visual similarity is a perceptual phenomenon without ground-truth data. This makes collecting data using crowdsourcing experiments valuable. Indeed, there are several crowdsourcing-based works [87, 88, 6] that address shape or style similarity problems and conduct user experiments to either derive or validate models.

Of most related to our work are two similarity studies that also employ subjective experiments. Among these, in Rogowitz et al. [89], human participants are asked to judge image similarity using two different experiments: one involving printouts of images (called table scaling) and the other using a computer based comparison (called computer scaling). These results are compared with computational similarity approaches [90] and simple CIELAB histograms. It was found that both table and computer scaling yield similar results and color is a major factor influencing similarity for human observers.

In another study [91], user experiments are conducted to evaluate the relationship between an image-indexing system and perceived similarity in an LDR setting. The tested image indexing system is based on basic properties of early stages of human vision – chromaticity, luminance, and texture. Two-alternative forced-choice (2AFC) method is used for all experiments. Three images are shown to the observer, the query image and two test images. Of these two images one image is called the target and the other the distractor. These images are selected based on the rankings obtained from the image-indexing system. Then the correlation between the users' preference and index rank is investigated. First, each index, chromaticity, luminance, and texture are calculated separately. From these indexes chromaticity is found to give the best results. Then for the second experiment, combinations of the indexes are evaluated. The combination of chromaticity and texture indices are found to give better results than chromaticity alone and the combination of all indices are found to give the best result.

As mentioned above, although visual image similarity is an extensively studied subject [5], to our knowledge there is no study that directly addresses this problem for

HDR images. Thus, understanding the nature of image similarity for HDR images and developing an objective similarity measure is the primary goal of this thesis.

CHAPTER 3

FEATURES AND METRICS

In this thesis, image representations and the scheme that these representations are related constitutes the basics of the assessment of similarity between images. In this chapter, first, image features that are used to represent images are presented. The image features measure different aspects of the image and also differ in representation. Therefore, the metrics that estimate the distance between image features need to be chosen accordingly. Several distance metrics are presented in the second part of this chapter and for each feature the selected distance metric is given.

3.1 Image Features

As the most fundamental part of many computer vision applications, image feature extraction is a widely studied research topic for the last several decades. There are numerous image features in the literature developed with different purposes which can be divided into two broad categories: global and local features. While global features like histograms of image properties represent the whole image with a single feature vector, local features like SIFT [64] or SURF [65] are a collection of vectors representing small neighborhoods called interest regions. Although local features are used successfully for the applications like image classification or retrieval, this thesis focuses on global features to investigate the image similarity in a general sense. In this section, a review of global image features that have been used in this thesis is given. Table 3.1 lists these features together with their representations and the distance metric used for each feature.

Table 3.1: HDR Image features and distances

Feature	**Model**	**Distance Metric**
Color	2D chromaticity histogram	EMD
Luminance	1D (relative) luminance histogram	EMD
Texture	Histograms of gradients	EMD
GIST	Feature vector	Cosine distance
VGG16/VGG19 - fc6	Fused fc6 layer	Cosine distance
VGG16/VGG19 - fc7	Fused fc7 layer	Cosine distance

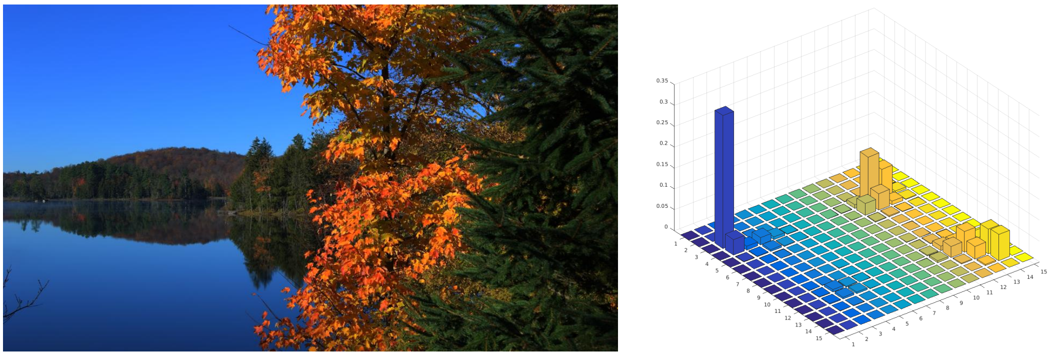


Figure 3.1: Sample image (left), 2D histogram (right).

3.1.1 Color

Since the early days of the image similarity research, color has been used as one of the most discriminative cues [91]. In this thesis, we used the a and b channels of the CIELAB color space [92] to represent chromaticity information. This is an opponent color space, where the a channel represents red/green opponent colors and the b channel yellow/blue opponent colors. We used a 2D chromaticity histogram to represent the distribution of colors in a given image. Each dimension contained 15 bins for a total of 225 bins. Figure 3.1 shows this histogram for the Mason Lake image from the dataset [4].

3.1.2 Texture

Texture is the second most used feature for content based image retrieval systems after chromatic features. This feature is especially helpful for discriminating images that have similar color but different spatial characteristics such as blue sky and sea or sand and buildings. In this thesis, to represent the texture information histogram of gradient magnitudes [93] is used.

3.1.3 Luminance

The main difference between an HDR and an LDR image is the much wider range of luminance distribution for the former. A single HDR image may contain very low luminances corresponding to highly shadowed regions as well as very high luminances corresponding to bright highlights. Therefore, we hypothesized that the luminance distribution of an HDR image may be an important cue for visual similarity. The luminance distribution is modeled using a 1D (relative) luminance histogram with 50 bins.

3.1.4 GIST Features

The GIST descriptor [63] aims to represent the dominant spatial structure of a scene by using low level multi-scale representations. This descriptor defines the scene as a whole rather than focusing on individual objects or regions. Discriminative properties of a scene are listed as naturalness, openness, roughness, expansion, and ruggedness. The class of a scene, e.g., man-made, natural, indoor, outdoor, etc., is determined by these properties.

The procedure for extracting GIST descriptors consists of applying Gabor filters that are scaled and orientated differently to the input image, dividing the filter response map into a grid in order to have spatial information, averaging the filter response in each grid, and concatenating the results to obtain the final feature vector, i.e. the GIST descriptor.

3.1.5 Deeply Learned Features

Recently, DCNNs have started to dominate object recognition and image classification tasks, achieving near human success rates [66, 67, 94]. These models are trained with large prelabeled datasets and develop a hierarchical model that becomes more aware of the content of the image rather than the underlying pixel values. To our knowledge currently there is no DCNN model that is trained on HDR images for the purpose of image indexing, scene classification, or visual similarity tasks. Furthermore, there is no prelabeled large HDR image dataset to use for training a DCNN model from scratch. Therefore in this thesis, we used transfer learning method to employ pretrained DCNNs for our perceptual similarity problem.

For feature extraction, pretrained AlexNet [66] and two variants of VGG networks, VGG16 and VGG19, are used [67]. All networks are trained on the ImageNet [85] dataset, but we also evaluated their performance when trained using different datasets. For transfer learning, the last fully connected layer, which contains classification outputs, is removed and the remaining 4096 dimensional two fully connected layers, **fc6** and **fc7**, are used as feature vectors. As suggested by Simonyan and Zisserman [67], the results obtained from VGG16 and VGG19 are fused (by taking an average) and it is observed that the fused version performs better than both VGG16 and VGG19. The distance between the feature vectors are calculated using cosine distance, which is a commonly used distance metric for deep learning features.

3.2 Distance Metrics

The use of a proper distance metric is as important as the features themselves. Each feature representation may require a different distance metric. In this section, we briefly describe the definitions and properties of the dissimilarity measures that we used for different types of features.

3.2.1 Euclidean Distance

The Euclidean distance between two histograms p and q is calculated as:

$$dist_{euc}(p,q) = \sqrt{\sum_i (p_i - q_i)^2}, \tag{3.1}$$

where i is the bin index. In general, dissimilarity obtained by Euclidean distance for histograms is not satisfactory as it does not take bin proximity into account.

3.2.2 Bhattacharyya Distance

Bhattacharyya distance [95] measures the overlap between two distributions. If p and q are two histograms, it can be calculated as:

$$dist_{bhat}(p,q) = -\ln\left(\sum_i \sqrt{p_i.q_i}\right). \tag{3.2}$$

For our HDR similarity problem Bhattacharyya distance gives slightly better results than Euclidean distance. However, it also suffers from the same problem that the proximity of the bins is not taken into account.

3.2.3 Earth Mover's Distance

Earth Mover's Distance (EMD) is a dissimilarity metric commonly used for image the retrieval problems [96]. EMD aims to capture the perceptual similarity between two distributions by calculating the minimal cost of transforming one distribution to the other. Unlike the other dissimilarity metrics, EMD can be calculated for varying-size partitions of the data, called signatures. Signatures consist of dominant clusters of the data, represented as $si = (m_i, w_i)$ pairs where mi is the cluster center and w_i is the size of the cluster. EMD does not require the signatures to have the same number of clusters – ground distances between cluster centers are sufficient. Histograms are signatures with bin centers corresponding to cluster centers, mi, and normalized bin values to weights, w_i.

The total amount of work to transform distribution p to q with flow f is:

$$WORK(P,Q,F) = \sum_{i}^{m}\sum_{j}^{n} d_{ij} f_{ij}, \tag{3.3}$$

where dij is the ground distance between cluster centers i and j. The optimal flow f that results with the minimum work, can be found by any linear optimization algorithm. When f is calculated, the EMD between p and q is defined as:

$$EMD(p,q) = \frac{\sum\sum d_{ij} f_{ij}}{\sum\sum f_{ij}}. \tag{3.4}$$

In our problem, bin centers correspond to color values (ab values in the CIELAB space) and ground distances are calculated as Euclidean because of the perceptual uniformity of the CIELAB color space.

Figure 3.2 compares the effect of these three distance metrics for a sample image from the dataset. The image on the first column is the query image, and in each row, the most similar five images from the dataset are shown. The distance metric used in first row is Euclidean, the second row is Bhattacharyya, and the last row is the EMD. It can be argued that more similar images are found using the EMD metric.

3.2.4 Cosine Similarity

Cosine distance between two vectors p and q is calculated as:

$$dist_{cosine}(p,q) = 1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}} \tag{3.5}$$

Cosine distance is a widely used distance metric for deep representations. In this thesis, we used cosine distance for calculating the distances between CNN feature vectors and GIST features.

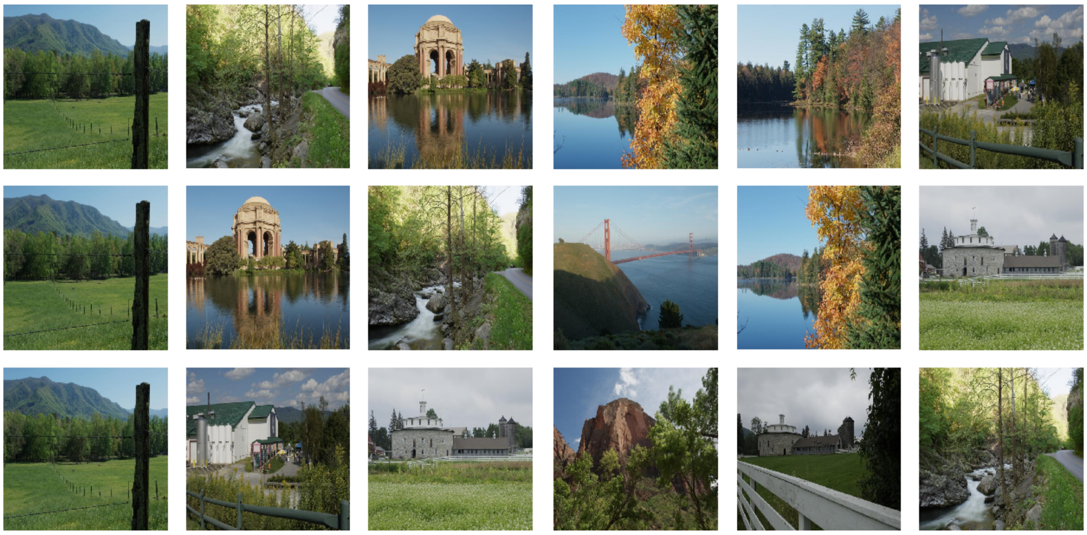


Figure 3.2: A comparison of dissimilarity metrics for histogram-based features. The leftmost image is the query image, the most similar five images from the dataset are shown in each row: Euclidean distance (first row), Bhattacharyya distance (second row), Earth Mover's distance (third row).

CHAPTER 4

VISUAL SIMILARITY EXPERIMENTS

Image similarity is inherently a subjective phenomenon and like other subjective phenomena user experiments play an important role for a better understanding and improved models. Thus, within the scope of this thesis, a user experiment is conducted to investigate HDR image similarity. This chapter presents the image dataset used for the experiment, the experimental setup that allows the users to assess similarity between HDR image triplets, the data gathering through crowd sourcing and then experiment analysis that displays how the image features introduced in Chapter 3 correlates with human responses. Lastly, two models for a combined feature are proposed which is a combination of individual features and the weights of these models are estimated using experiment responses.

4.1 Dataset

The set of images used in visual similarity experiments should be sufficiently diverse. Although such datasets exists for LDR images, there is no specific similarity dataset for HDR images. However, there exists HDR image datasets that were created for various purposes and by different authors. We therefore decided to select 100 HDR images from various such sources to present observers with a diverse set of images2. The used datasets were: Fairchild's HDR Photographic Survey [4], HDR-Eye [97], DEIMOS [98], Empa HDR Image Database [99], and pfstools HDR Image Gallery [100]. Thumbnails for the used images are shown in Figure 4.1.

Figure 4.1: HDR images used in the visual similarity experiments.

4.2 Experiment Setting

To measure perceptual similarity between HDR images, we conducted a 2AFC experiment. The experiment is publicly available[1]. As we needed a large number of responses, we designed a web-based interface to collect crowdsourcing data. We used the HDRHTML technique [101] for visualizing HDR images on web browsers.

This technique uses a windowing approach to select a desired exposure range from the HDR image. Multiple exposures are encoded by combining a small set of basis images with opacity coefficients. The tone-curves of these basis images are approximated as a piece-wise linear function. Instead of finding optimal tone-curves and opacity coefficients, HDRHTML uses a precomputed optimal solution and use these tone-curve points and opacity coefficients for all images for fast processing. After basis images are created using the tone-curves, these basis images are used to reconstruct multiple exposures and gives user the control over exposure settings with a slider. By dynamically adjusting the position of the slider, the user can efficiently view the entire exposure range contained within the HDR image. These sliders are normally overlayed with the image histogram. We removed this overlay to prevent the image histogram from affecting the observers' decisions. Figure 4.2 shows a sample trial from the experiment. An HDR reference image was shown at the top and two HDR test images were shown at the bottom. The sliders, which were mandatory to be adjusted, allowed all images to be inspected at different exposure levels.

In each experimental session, 33 such image triplets were displayed to the observers. Thus, an experimental session consisted of 33 trials. In each trial, the observers were asked to choose which of the two test images was visually more similar to the reference image. All trials, except for the verification ones, were generated randomly from the dataset during the runtime of the experiment.

Three of the experiment triplets were used for verification. They contained an obviously similar reference and test image pair to evaluate the reliability of an observer as shown in Figure 4.3. As seen from the figure, the test image that is similar to the reference image is another image from the same scene. These images are not used

[1] http://user.ceng.metu.edu.tr/ merve/userstudy/

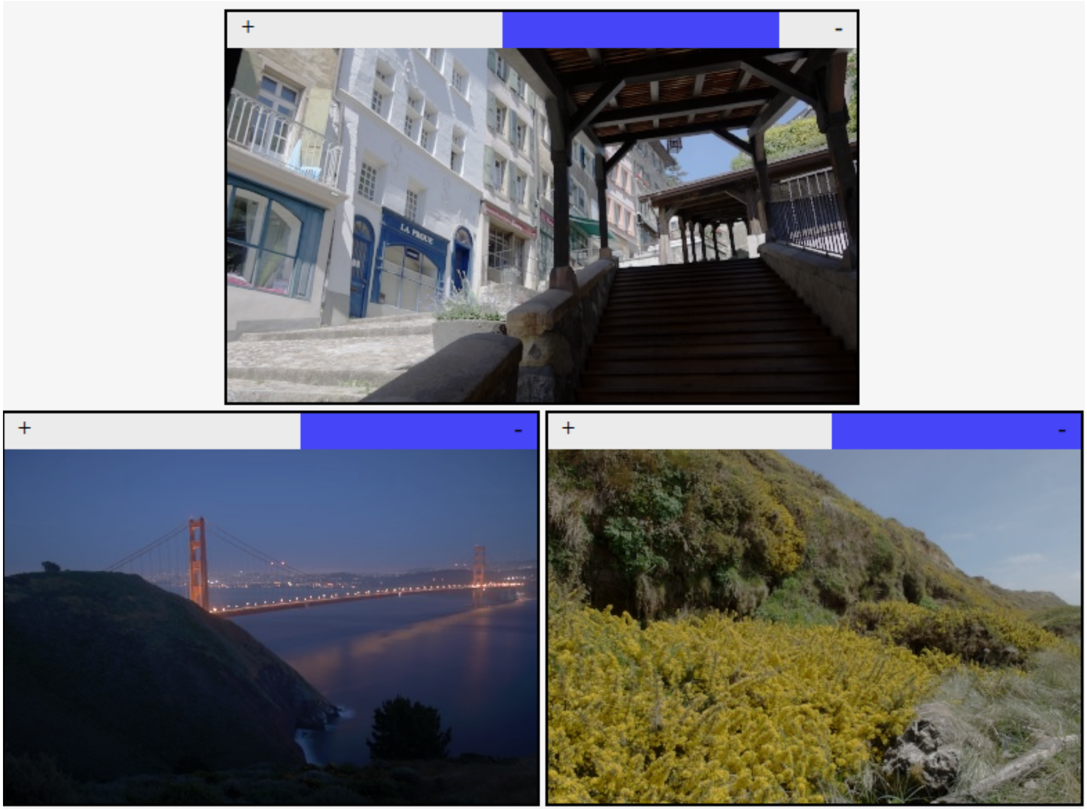


Figure 4.2: A sample trial from the experiment. The observers were asked to choose the most similar image to the reference image (top) from the test images (bottom). All images could be examined at different exposure levels by adjusting their sliders.

in the actual part of the experiment, and do not belong to the dataset of 100 images shown in Figure 4.1.

If an observer failed to provide the correct answer even for one of these trials, his or her data was discarded as being unreliable. These trials were distributed across the experiment to ensure that observers were attentive throughout. Before the experiment began, observers were informed about their task and the expected duration of the experiment, which was at most 20 minutes at a normal pace. During the experiment, observers were required to use the exposure sliders for each image before they made selection. Image selection was done by clicking on one of the test images. The selection was indicated using a green border around the selected image. Observers could change their selection until they pressed the “Next” button. The progress of an observer was indicated using a small progress bar at the bottom center of the screen. At the end of the experiment, observers were informed with a final page confirming the conclusion of the experiment and were presented with unique session ids. They were required to enter this id to the crowdsourcing platform to verify that they have finished the experiment.

4.3 Data Collection

Crowdsourcing has been used in many computer vision problems to collect non-expert data [102]. In this thesis, in order to reach as many people as possible, the experiment was published at Microworkers crowdsourcing platform[2]. For each completed experiment 0.3$ were paid to the participants.

At the beginning of the experiment, an introductory page, given in Figure 4.4, is displayed to the user. This page first describes the task to be fulfilled: complete a set of trials by picking the comparison image that is more similar to the reference image. Here it is important to note that the users are not asked to decide for a specific type of similarity such as object, color, etc. By intentionally leaving the definition of visual similarity vague, it is hoped to achieve a range of responses, which in overall, would converge to a common sense understanding for similarity.

[2] www.microworkers.com

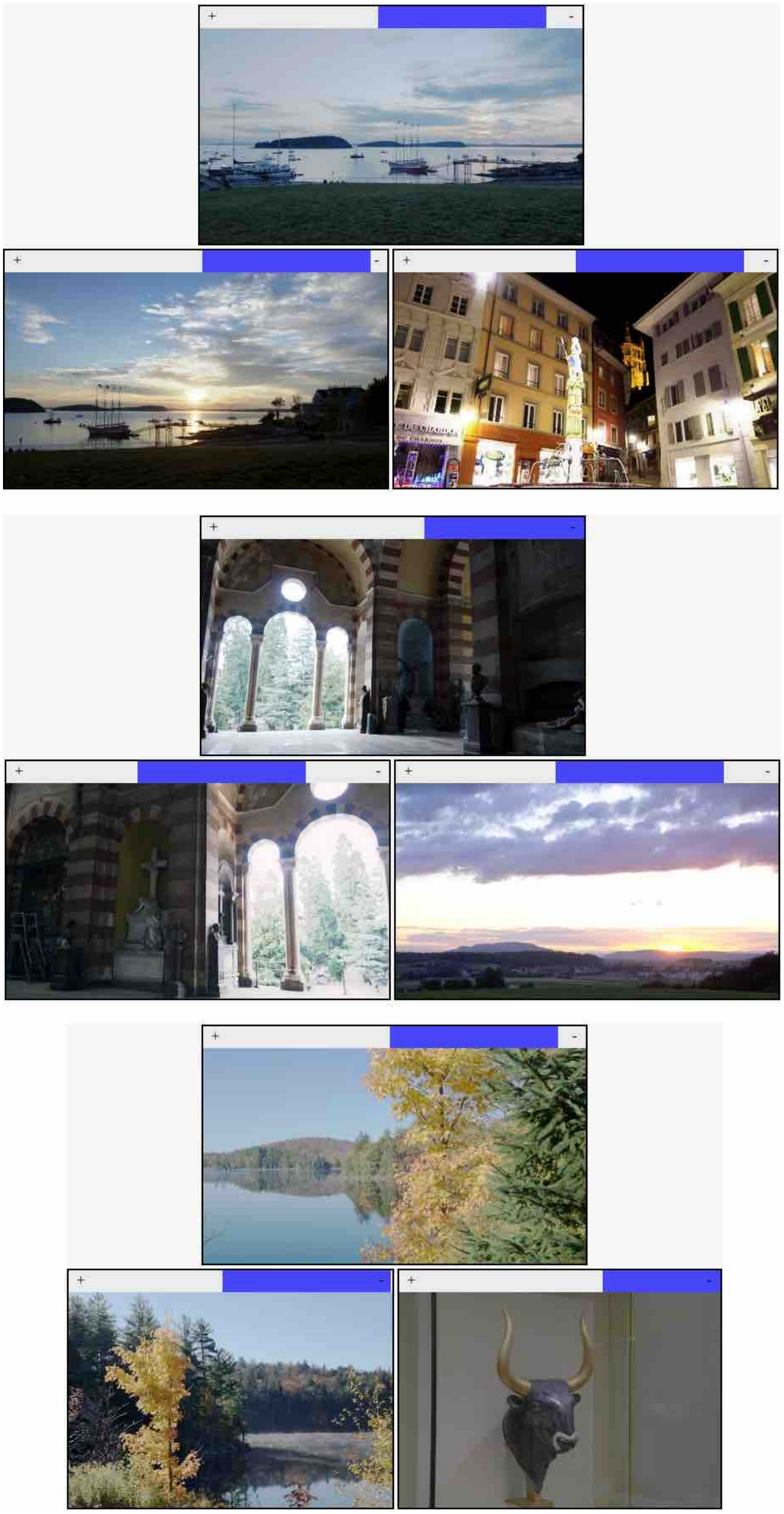
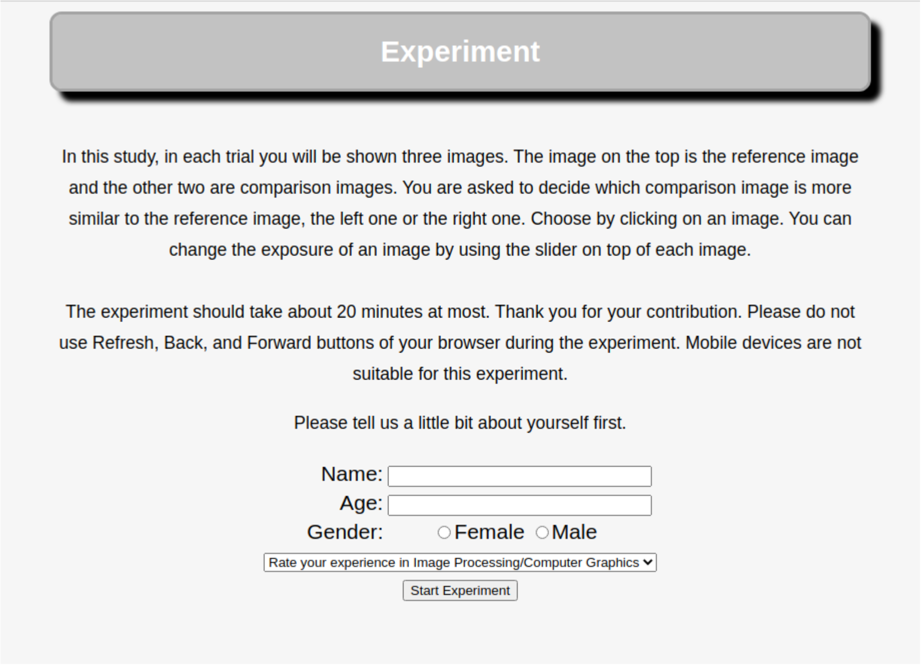


Figure 4.3: Verification triplets, shown as 3rd, 10th and 16th trials in the experiment sessions.


Experiment

In this study, in each trial you will be shown three images. The image on the top is the reference image and the other two are comparison images. You are asked to decide which comparison image is more similar to the reference image, the left one or the right one. Choose by clicking on an image. You can change the exposure of an image by using the slider on top of each image.

The experiment should take about 20 minutes at most. Thank you for your contribution. Please do not use Refresh, Back, and Forward buttons of your browser during the experiment. Mobile devices are not suitable for this experiment.

Please tell us a little bit about yourself first.

Name:
Age:
Gender: ○Female ○Male

Figure 4.4: Start page of the experiment, that describes the task and collects some information about the user.

After giving the instructions about the experiment, the user is informed about the expected time to finish the experiment when done in a normal pace. This is followed by some warnings about browser usage during the experiment and unsuitable devices. Then, a form is presented to collect some information about the participants. In Figure 4.5, the distribution of age, gender, and familiarity with computer graphics/image processing of the participants according to the data entered to this form is shown. Finally, the user can start the experiment on demand by pressing the provided button labeled as "Start Experiment".

One of the challenges of data crowdsourcing is eliminating users that give unreliable responses, this may due to not being qualified for the task or just being a spammer [103]. To minimize the problem of users being not qualified, the crowd group of English speaking, highly qualified users are selected from the crowdsourcing platform. Even though the experiment itself does not require language proficiency, the instructions at the beginning of the experiment is important for users to successfully complete the experiment. The users are in the *highly qualified* crowd, if they have been done other crowdsourcing tasks before and received some positive feedback.

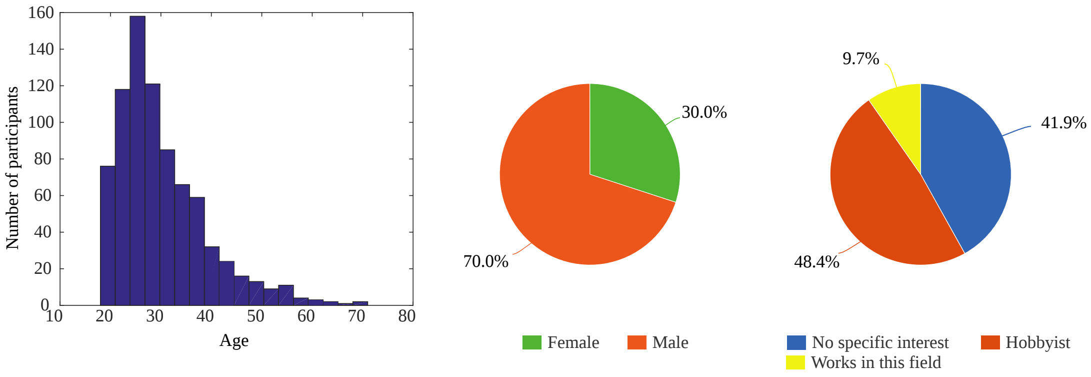


Figure 4.5: Age, gender, and computer graphics/image processing familiarity distribution of the participants.

Another measure taken to achieve reliable responses is using verification triplets. In total, 165 sessions were discarded due to incorrect responses given to the verification trials.

4.3.1 Phase I

In the first phase of the experiment, randomly selected triplets are shown to the users without any restrictions on the image selection. After collecting the experimental results, and eliminating the triplets from invalid sessions, it was found that 18747 unique image triplets were judged by the observers. This amounts to approximately 11.6% of the total possible triplets that can be obtained from 100 images, $C(100, 3)$. Experiment sessions were independent and random for each participant, but it was guaranteed that a single session consisted of only unique triplets.

This design resulted in a single response for the majority of the triplets. Some triplets received two responses and only a few received three or more. As such this first phase of the experiment is considered as a random exploration of all possible comparisons. However, as judging similarity based on a single response could be too subjective, the experiment is extended as discussed below to collect multiple responses for each triplet.

4.3.2 Phase II

The first phase of the experiment was extended to obtain three evaluations per triplet. Unlike the first phase where triplets were generated randomly, the second phase solely used the triplets that had been evaluated in the first phase of the experiment. To achieve this, the triplets sorted from the first phase in descending order by the number of responses collected. If a triplet had more than three responses, three of the responses are randomly selected. The triplets with exactly three responses were used as is. These two cases occurred very rarely. Next, triplets with two responses, and then a single response were presented randomly to obtain a total of 4990 triplets that had been evaluated three times. Among these thrice evaluated triplets, 2170 triplet were judged consistently by all three observers. The remaining 2820 triplets generated two-to-one responses. Similar to the first part of the experiment, the second part also contained the same validity checks to eliminate the responses of inattentive observers.

4.4 Experiment Analysis

Having discussed the details of crowdsourcing study in Section 4.3 and experimented features in Section 3.1, this section investigates how the human judgments and the image features are related. First analysis method for assessing the correlation between individual feature type and the experiment results is given. Then, two possible methods to combine the features for developing a more effective similarity model is discussed. In the evaluations, different formats for HDR images are employed in order to measure the effect of the image format to the proposed methods.

4.4.1 Preprocessing

In evaluations, HDR images used directly, as well as by linear scaling and applying several tone mapping operators. For linear scaling, 5% of the brightest pixels and 5% of the darkest pixels are discarded by setting them to 95th and 5th percentile of the pixels values respectively. The reason behind this step is to eliminate extreme values from the HDR images that may also introduced by hardware.

$$I' = \begin{cases} P_5(I), & \text{if } I < P_5(I) \\ I, & \text{if } P_5(I) \leq I \leq P_{95}(I) \\ P_{95}(I), & \text{if } I > P_{95}(I), \end{cases} \tag{4.1}$$

where I denotes the original image values and I' is after the pixel values outside of the range are removed. Then HDR image is linearly scaled with

$$lin(I) = \frac{I' - I'_{min}}{I'_{max} - I'_{min}}, \tag{4.2}$$

for each color channel separately.

Besides of using original and linearly scaled HDR images, tone mapped version of the images are also evaluated. For this purpose commonly used tone mapping operators Mai et al. [58], Reinhard et al. (local) [52], Reinhard et al. (global) [52], Durand & Dorsey [59], Mantiuk et al. [61], Reinhard & Devlin [55], Fattal et al. [59], Mantiuk et al. [57], Ferradans et al. [56] and Pattanaik et al. [54] are used. The implementations of these TMOs are available in opensource PFStmo software library [100], which provides a reliable implementation of several commonly used TMOs. The images in the experiment dataset are tone mapped using the PFStmo software with the given TMOs using the default parameters.

4.4.2 Individual Feature Correlations

To investigate how individual image features correlate with the human responses collected through the experiment, the feature responses are treated as an observer that are given the same set image triplets used in the experiment. Assume that $t_i = R_i - A_i - B_i$ represents the i^{th} triplet (i.e. trial) with R_i being the reference image, A_i the left test image, and B_i the right test image. This triplet could have been evaluated one or more times by different human observers. Let $n(A_i)$ and $n(B_i)$ represent the number of times that each image was found more similar to R_i than the other. From this information, a binary vector is created to encode the participants' responses:

$$P = (x_1, \ldots, x_N), \tag{4.3}$$

where each element is defined as:

$$x_i = \begin{cases} 1, & \text{if } n(A_i) > n(B_i), \\ 0, & \text{otherwise.} \end{cases} \tag{4.4}$$

For each feature type f, the feature representations of each image is computed as $f(R_i)$, $f(A_i)$, $f(B_i)$. Then their similarity to each other is calculated to obtain the following binary vector:

$$F = (y_1, \cdots, y_N), \tag{4.5}$$

where

$$y_i = \begin{cases} 1, & \text{if } d(f(R_i), f(A_i)) < d(f(R_i), f(B_i)) \\ 0, & \text{otherwise.} \end{cases} \tag{4.6}$$

In this equation d represents the distance metric that was chosen to be used for feature f. This encoding gave rise to two binary vectors, P and F, with the former computed from user responses and the latter from feature similarities. There are many approaches to compute the correlation between two such vectors. In this thesis, the Sokal-Michener correlation is used, which is a simple, intuitive, and effective way to correlate two binary vectors [104]. This correlation is defined as

$$s = \frac{S_{11}(P, F) + S_{00}(P, F)}{N}, \tag{4.7}$$

with S_{11} and S_{00} representing the total count of matching ones and zeros respectively:

$$S_{11}(P, F) = P \cdot F, \tag{4.8}$$

$$S_{00}(P, F) = \neg P \cdot \neg F, \tag{4.9}$$

Note that the correlation coefficient s can take a value in range [0, 1]. In the following, this coefficient is multiplied by 100 to represent the correlations as percentages.

The raw feature correlations with the first (Section 4.3.1) and the second phase of the experiment (Section 4.3.2) are reported in Tables 4.1 and 4.2, respectively. In these tables, the leftmost column indicates the processing type applied to the images before the computation of features.

"HDR-original" represents the unaltered HDR image whereas "HDR-linear" represents its linearly scaled version. The other processing types all include the application of a certain tone mapping operator. For all processing types, except the original, the images were gamma-corrected and scaled to [0, 255] range.

4.4.3 Feature Learning with Triplet Networks

For both of the phases of the user experiments and for different tone mapping operators the deeply learned features correlate better than the other features as given in Table 4.1 and Table 4.2. The deep learning features are the last fully connected layer of two deep CNN architectures, VGG16 and VGG19, that are trained on ImageNet dataset. There are many studies that report better results achieved with transfer learning by taking a pretrained network on a different dataset and then training this dataset on the actual dataset [105]. For most of the studies including this thesis, the reason of choosing finetuning over training a CNN from scratch is the insufficient amount of labeled data and the required training time for the huge datasets like ImageNet that has millions of images.

Since, both VGG16 and VGG19 are classification networks, the input for these network is a single image and the output is the class labels. The collected user data is not suitable for finetuning these networks. However, there are architectures that consists of multiple CNNs for training comparison data. One example is Siamese architecture [106]. This network takes two images as input and contains two different CNNs that can be trained for classification or feature learning. A more recent architecture is Triplet Network [107], which takes three images as input, compares triplets simultaneously and outputs the probability of the given triplet is a valid triplet, the reference image is closer to the first image. In this thesis, a triplet network architecture called DeepRanking [81] is trained on the experimental data. This network has three identical deep neural networks with shared parameters. These deep neural networks learns

Table 4.1: Individual feature correlations with the first phase of the experiment. The numbers indicate the Sokal-Michener correlation scaled by 100 to represent percentages.

Processing Type	**VGG16**	**VGG19**	**Color**	**Luminance**	**Texture**	**GIST**
HDR-original	56.79	58.09	55.10	53.14	52.39	56.82
HDR-linear	63.54	63.31	55.69	54.07	54.36	58.18
Drago et al. [53]	65.88	65.74	56.73	57.45	51.17	58.23
Mai et al. [58]	65.28	65.13	56.01	56.77	51.90	57.57
Reinhard et al. (local) [52]	65.82	65.63	56.58	54.77	51.43	57.89
Reinhard et al. (global) [52]	65.75	65.52	56.59	54.68	51.39	57.92
Durand & Dorsey [59]	66.17	65.43	55.77	55.12	51.79	57.85
Mantiuk et al. [61]	65.42	65.33	56.29	55.38	52.08	58.03
Reinhard & Devlin [55]	65.28	65.20	57.15	55.89	54.85	58.33
Fattal et al. [59]	65.90	65.72	56.39	57.46	51.92	58.19
Mantiuk et al. [57]	65.71	65.74	55.98	56.99	51.84	57.86
Ferradans et al. [56]	66.02	65.90	55.18	56.51	51.99	58.33
Pattanaik et al. [54]	64.46	64.38	53.04	54.61	53.06	57.84

Table 4.2: Individual feature correlations with the second phase of the experiment. The numbers indicate the Sokal-Michener correlation scaled by 100 to represent percentages.

Processing Type	**VGG16**	**VGG19**	**Color**	**Luminance**	**Texture**	**GIST**
HDR-original	64.88	67.14	60.23	58.39	54.42	63.50
HDR-linear	75.58	76.13	60.78	57.79	57.97	65.71
Drago et al. [53]	80.88	81.80	62.58	62.72	53.87	65.39
Mai et al. [58]	80.00	79.95	61.11	61.66	53.46	64.06
Reinhard et al. (local) [52]	80.92	81.61	62.21	58.16	53.87	64.88
Reinhard et al. (global) [52]	80.92	81.57	62.21	57.97	54.75	64.75
Durand & Dorsey [59]	81.75	81.34	62.07	59.22	53.00	64.19
Mantiuk et al. [61]	80.41	80.65	61.15	59.59	52.49	64.47
Reinhard & Devlin [55]	80.37	80.41	64.15	61.43	60.55	65.44
Fattal et al. [59]	80.51	80.92	62.30	64.24	52.90	65.02
Mantiuk et al. [57]	80.00	80.78	62.12	61.71	54.56	64.19
Ferradans et al. [56]	81.38	82.21	58.39	61.61	55.02	65.25
Pattanaik et al. [54]	78.66	78.11	57.33	58.66	55.71	64.52

an image embedding f that will be the input of the last layer called ranking layer which minimizes the hinge loss:

$$l(p_i, {p_i}^+, {p_i}^-) = max\{0, g + D(f(p_i), f({p_i}^+)) - D(f(p_i), f({p_i}^-))\}, \tag{4.10}$$

where p_i is the input image, ${p_i}^+$ is the similar image and ${p_i}^-$ is the less similar image, D is the Euclidean distance between learned image features and g is a regularizer parameter of the gap between image pairs. In the Table 4.3, the correlation scores obtained from this network trained on the experiment data is given.

Table 4.3: Correlation results obtained with a triplet network trained on Phase I of the experiment.

Training	**Number of Triplets**	**Epocs**	**Correlation**
Training I	1500	25	60.50
Training II	1500	50	59.72
Training III	15000	25	61.09

For the training of this network Phase I of the experiment is used. Although Phase I is less reliable compared to Phase II, it has more triplets to train with. The selected tone mapping operator is the tone mapping operator of Durand & Dorsey et. al [59]. For the pretrained deep neural network for DeepRanking VGG16 is used. In total three trainings are evaluated, in the first training 1500 triplets are used and the network is trained for 25 epocs. The correlation is calculated similarly as Table 4.1 but on the triplets that has not been used for training. The first evaluation is resulted with ~5% lower correlation than VGG16 feature. For the second training, the tranining duration is doubled with 50 epocs but this did not increase the correlation, on the contrary it is slightly decreased possibly due to overfitting. For the third training, the number of triplets used for training increased 10 times and a slight increase in correlation is observed. However, it was not possible to train the network with more data since the Phase I of the user experiment has approximately 18K triplets.

Table 4.4: p values for statistical significance between individual image features, calculated with Fisher's exact test.

	VGG16	VGG19	Color	Luminance	Texture	GIST
VGG16	1.0	0.504	$3.832x10^{-62}$	$1.123x10^{-47}$	$4.828x10^{-79}$	$1.963x10^{-33}$
VGG19	0.504	1.0	$3.083x10^{-67}$	$3.670x10^{-52}$	$9.416x10^{-85}$	$3.302x10^{-37}$
Color	$3.832x10^{-62}$	$3.083x10^{-67}$	1.0	0.033	0.027	$3.720x10^{-06}$
Luminance	$1.123x10^{-47}$	$3.670x10^{-52}$	0.033	1.0	$1.222x10^{-05}$	0.014
Texture	$4.828x10^{-79}$	$9.416x10^{-85}$	0.027	$1.222x10^{-05}$	1.0	$6.947x10^{-12}$
GIST	$1.963x10^{-33}$	$3.302x10^{-37}$	$3.720x10^{-06}$	0.014	$6.947x10^{-12}$	1.0

4.4.4 Statistical Analysis

In Table 4.4, the statistical significance of the difference between each individual feature correlations are given. Fisher's exact test [108] is used for the calculations, image features are calculated on images tone mapped with Ferradans et al. [56] TMO, and correlation is calculated against Phase II of the experiment. This analysis shows that, the difference for the correlations of VGG16 and VGG19 features are not that significant and also Color - Luminance - Texture performances are rather close. On the other hand, deep learning features are significantly better than the other features and similarly, GIST feature is significantly better than Color and Texture features.

Another variable in the experiment analysis is the selection of tone mapping operators. Results given in Table 4.1 and Table 4.2 shows that for a given image feature tone mapping operators are performing comparably. Table 4.5 shows the p values calculated with Fisher's exact test [108] between the VGG19 feature's correlation with the Phase II of the user study on images tone mapped with different tone mapping operators. While for most of the tone mapping operators the performance difference is not significant. For this feature, the differences between the best performing Ferradans et. al.'s TMO [56] and the successors Drago et al.'s [53] and Reinhard et al. [52] TMOs are not significant ($p > 0.6$). However, the difference to the worst performing TMO, Pattanaik et al.'s TMO is significant with $p = 0.001$.

4.4.5 Combined Feature Correlation

Given the individual correlations reported in the tables Table 4.1 and Table 4.2, a natural question that follows is if the features can be combined to develop a single objective metric that better correlates with human's assessment of similarity for HDR images. To this end, two types of logistic regression analysis are performed yielding two related but different models.

4.4.5.1 Triplet Model

In the first analysis, the aim is to develop a model that predicts which of the two test images is more similar to the reference image using the pairwise distances between the test and reference images. Assuming that j is a feature index, one can compute these pairwise differences as follows:

$$a_j = d_j(f_j(R), f_j(A)), \tag{4.11}$$

$$b_j = d_j(f_j(R), f_j(B)). \tag{4.12}$$

Here d_j represents the distance metric chosen for the j^{th} feature. The model takes as input these differences for all features (i.e. $j \in 1, 2, 3, 4, 5, 6$) and computes their weighted average as its response:

$$r = c_0 + c_1(a_1 - b_1) + c_2(a_2 - b_2) + c_3(a_3 - b_3) + c_4(a_4 - b_4) + c_5(a_5 - b_5) + c_6(a_6 - b_6) \tag{4.13}$$

To compute the unknown coefficients logistic regression is used as the dependent data (i.e. user responses) is binary: given one reference and two test images, the user selects either the left image or the right one, encoded as 1 and 0.

The regression is performed between the two vectors, namely the P vector from Equation 4.3, and the model response R comprised of the following elements:

$$R = (r_1, \cdots, r_N), \tag{4.14}$$

where

$$r_i = [a_{i1} - b_{i1} \cdots a_{i6} - b_{i6}]. \tag{4.15}$$

The logistic regression models the logarithm of the odds as the response of the model:

$$ln\left(\frac{Pr(x=1)}{1-Pr(x=1)}\right) = r. \tag{4.16}$$

From this equation, it can be derived that the probability of a user responding 1 (i.e. selecting the left image) is equal to

$$Pr(x=1) = \frac{1}{1+e^{-r}} \tag{4.17}$$

If $Pr(x = 1) > 0.5$, it is assumed that the model has selected the left image. Otherwise, the model's response was taken as the right image.

To measure the effectiveness of this model 10-fold cross validation is used. In each fold, 90% of the trials were selected for training and the remaining 10% for testing. This process was repeated 10 times while ensuring that each test fold is mutually exclusive from each other. Similar to the analysis of individual features, the success of this model is assessed against both the Phase I and the Phase II of the experiment. The results are shown in Table 4.6. It can be seen that the feature combination, on average, improves the success of each presentation type by about 3% to 4%. The best three results are obtained by Ferradans et al.'s [56], Drago et al.'s [53], and Reinhard et al.'s [52] TMO algorithms. The reported coefficients are computed by using the entire dataset from the Phase II of the experiment due to its higher correlation with the combined features.

4.4.5.2 Duplet Model

Despite the first regression model yielding high correlations exceeding 80% for most algorithms, it has an important drawback. It requires a triplet of images, one reference and two test, as input to the model. While this matches the presentation type in our experiment, a more desirable model should be able to take only two images (e.g., a

Table 4.5: p values for statistical significance between correlations of VGG19 features extracted on images tone mapped with different tone mapping operators, calculated with Fisher's exact test. User responses are used from Phase II of the experiment.

	Dra. [53]	Mai. [58]	Rein.(local) [52]	Rein. (global) [52]	Dur. [59]
Dra. [53]	1.0	0.132	0.906	0.875	0.725
Mai. [58]	0.132	1.0	0.178	0.19	0.265
Rein.(local) [52]	0.906	0.178	1.0	1.0	0.845
Rein.(global) [52]	0.875	0.19	1.0	1.0	0.876
Dur. [59]	0.725	0.265	0.845	0.876	1.0
Man. [61]	0.351	0.593	0.438	0.461	0.588
Rein. [55]	0.261	0.732	0.333	0.353	0.463
Dur. [59]	0.483	0.444	0.586	0.613	0.756
Man. [57]	0.414	0.516	0.509	0.534	0.67
Fer. [56]	0.752	0.063	0.636	0.608	0.479
Pat. [54]	0.003	0.146	0.005	0.005	0.009

	Man. [61]	Rein. [55]	Dur. [59]	Man. [57]	Fer. [56]	Pat. [54]
Dra. [53]	0.351	0.261	0.483	0.414	0.752	0.003
Mai. [58]	0.593	0.732	0.444	0.516	0.063	0.146
Rein.(local) [52]	0.438	0.333	0.586	0.509	0.636	0.005
Rein.(global) [52]	0.461	0.353	0.613	0.534	0.608	0.005
Dur. [59]	0.588	0.463	0.756	0.67	0.479	0.009
Man. [61]	1.0	0.878	0.847	0.939	0.198	0.043
Rein. [55]	0.878	1.0	0.701	0.788	0.139	0.067
Dur. [59]	0.847	0.701	1.0	0.939	0.291	0.024
Man. [57]	0.939	0.788	0.939	1.0	0.241	0.032
Fer. [56]	0.198	0.139	0.291	0.241	1.0	0.001
Pat. [54]	0.043	0.067	0.024	0.032	0.001	1.0

Table 4.6: The correlations of the first regression model with the user responses. Phase I and Phase II represent the initial and extended experiments respectively. The coefficients are reported for the Phase II of the experiment only due to its higher correlation with the user data.

Processing Type	**Phase I**	**Phase II**	**c0**	**c1**	**c2**	**c3**	**c4**	**c5**	**c6**
HDR-original	60.67	70.76	0.0573	0.0768	-3.3241	-0.0028	-0.0124	-0.2921	-10.7505
HDR-linear	64.81	78.83	0.0005	-5.4801	-5.9902	-0.0074	-0.0635	-0.3289	-10.1782
Drago et al. [53]	67.36	83.49	-0.0423	-7.8751	-7.6339	-0.0506	-0.0958	0.0043	-7.3615
Mai et al. [58]	66.70	81.78	0.0085	-5.2932	-7.9526	-0.0601	-0.1078	-0.0358	-4.9275
Reinhard et al. (local) [52]	67.19	83.21	-0.0154	-7.3838	-8.7207	-0.0688	-0.0853	0.0145	-7.8380
Reinhard et al. (global) [52]	67.34	83.16	-0.0230	-7.0932	-8.8856	-0.0687	-0.0783	0.0101	-7.4470
Durand & Dorsey [59]	66.92	83.03	-0.0604	-8.1694	-7.3044	-0.0977	-0.0147	0.0082	-6.8549
Mantiuk et al. [61]	66.64	81.74	0.0220	-6.1999	-8.0462	-0.1081	-0.0286	-0.0102	-10.0494
Reinhard & Devlin [55]	66.72	82.75	-0.0332	-5.6555	-8.8871	-0.1284	-0.0144	-0.0254	-7.9970
Fattal et al. [59]	67.25	82.56	-0.0025	-6.2320	-8.3176	-0.1120	-0.0272	-0.0143	-7.9175
Mantiuk et al. [57]	66.91	82.15	-0.0005	-5.7555	-8.4433	-0.0777	-0.0548	-0.0041	-6.8189
Ferradans et al. [56]	67.21	83.53	0.0226	-5.7782	-9.8899	-0.0801	-0.0432	-0.0060	-7.4090
Pattanaik et al. [54]	65.02	79.89	-0.0365	-7.5194	-5.6052	0.0132	-0.0565	0.0211	-6.1389

query image and a test image) and produce a relative similarity score between them. This may allow, for instance, ranking the similarity of multiple images with a query image as in image-based search applications.

In order to allow for this possibility, our second regression model was designed in the following manner. For each trial, $ti = R_i - A_i - B_i, i \in 1, \cdots, N$, two elements are inserted to our user response vector:

$$x_{2i-1} = \begin{cases} 1, & \text{if } n(A_i) > n(Bi) \\ 0, & \text{otherwise,} \end{cases} \tag{4.18}$$

$$x_{2i} = \neg x_{2i-1}, \tag{4.19}$$

yielding a vector of size $2N$:

$$P = (x_1, x_2, \cdots, x_{2N}). \tag{4.20}$$

As for the model's inputs each element of the feature vector was computed as

$$y_{2i-1} = [a_1 \cdots a_6], \tag{4.21}$$

$$y_{2i} = [b_1 \cdots b_6], \tag{4.22}$$

yielding

$$F = (y_1, y_2, \cdots, y_{2N}). \tag{4.23}$$

In summary, the elements of the feature vector always followed the A, B order, whereas the corresponding elements in the user vector were 1 for the selected image and 0 for the other image. This second regression model learns to produce the following response given the feature differences between a reference and test image:

$$r_a = c_0 + c_1 a_1 + c_2 a_2 + c_3 a_3 + c_4 a_4 + c_5 a_5 + c_6 a_6 \tag{4.24}$$

By converting this response to probability values as in Equation 4.17, one can compute a relative degree of similarity between the two images. To validate this model, the model response is computed twice by using $R_i - A_i$ and $R_i - B_i$ image pairs:

$$Pr(x = left) = \frac{1}{1 + e^{-r_a}} \tag{4.25}$$

$$Pr(x = right) = \frac{1}{1 + e^{-r_b}} \tag{4.26}$$

Given a triplet, if $Pr(x = left) > Pr(x = right)$ it is assumed the model to have selected the left image. Otherwise, it was assumed that the model selects the right one. The correlation of this model with the user responses was calculated as in the previous model yielding the results in Table 4.7. The best result of the second model was found for Drago et al.'s [53] TMO in the second phase of the experiment. The model achieved a correlation of 83.81% with the user responses. Similarly to the Triplet Model, the reported coefficients are computed by using the entire dataset from the Phase II of the experiment due to its higher correlation with the combined features.

In this chapter, a crowdsourcing user study with the aim of HDR image similarity assessment is presented. The experiment data is collected in two phases with in total more than 1200 participants. While in the first phase, single response per image triplet is sought, the second phase extends the first phase by collecting three responses per image triplet. These responses analyzed separately for feature correlation and found out for the image triplets that have three consistent responses, the correlation scores are higher for both individual and combined features. Besides that, the effect of image format is also investigated and it is observed that tone mapped versions have higher correlations for hand crafted features and deeply learned features than raw HDR data and linearly scaled image. On the other hand, tone mapping operators perform comparably. Lastly, two models learnt from user data are proposed for the combined feature perform similarly with each other and better than the individual features. Among these models, duplet model is more suitable for applications in the sense that it does not require image triplets but estimates image similarity between image pairs. In the next chapter, it is shown how this model can be used to improve the style-based tone mapping application.

Table 4.7: The correlations of the second regression model with the user responses. Phase I and Phase II represent the initial and extended experiments respectively. The coefficients are reported for the Phase II of the experiment only due to its higher correlation with the user data.

Processing Type	**Phase I**	**Phase II**	**c0**	**c1**	**c2**	**c3**	**c4**	**c5**	**c6**
HDR-original	60.75	70.80	2.9323	-0.1531	-2.8191	-0.0024	-0.0063	-0.2494	-7.1569
HDR-linear	64.65	78.50	5.5623	-3.9224	-3.7111	-0.0149	-0.0048	-0.2164	-5.5490
Drago et al. [53]	67.52	83.81	8.3967	-5.5248	-4.0845	-0.0280	-0.0587	-0.0054	-3.3575
Mai et al. [58]	66.72	81.73	7.4594	-4.0822	-5.0859	-0.0196	-0.0532	-0.0326	-0.9064
Reinhard et al. (local) [52]	67.35	83.53	8.2123	-5.6104	-4.3743	-0.0249	-0.0290	0.0063	-3.6705
Reinhard et al. (global) [52]	67.20	83.16	8.2162	-5.3915	-4.5828	-0.0259	-0.0264	0.0049	-3.6673
Durand & Dorsey [59]	66.81	82.61	8.2396	-5.9298	-4.0539	-0.0560	-0.0081	0.0077	-3.1001
Mantiuk et al. [61]	66.50	82.10	7.6833	-4.3626	-4.8232	-0.0658	-0.0212	-0.0082	-3.4889
Reinhard & Devlin [55]	66.56	82.61	8.5676	-4.6656	-4.9324	-0.0936	-0.0075	-0.0262	-2.8843
Fattal et al. [59]	67.07	82.79	8.0671	-4.4119	-4.8215	-0.0716	-0.0191	-0.0141	-2.8938
Mantiuk et al. [57]	66.57	82.01	7.7805	-3.9872	-5.3899	-0.0228	-0.0319	-0.0084	-2.6625
Ferradans et al. [56]	67.33	83.16	8.5911	-5.0432	-4.9965	-0.0541	-0.0258	-0.0089	-2.3825
Pattanaik et al. [54]	64.94	79.84	6.6735	-5.6657	-3.4601	0.0109	-0.0295	0.0241	-1.8922

CHAPTER 5

STYLE-BASED TONE MAPPING

HDR image similarity has many applications, and one of these applications on tone mapping is presented in this chapter. The tone mapping operator that is depicted in the following sections named as style-based tone mapping, consists of a methodology that employs image similarity to tone map HDR images consistently according to a style without manual parameter adjustments. In this chapter, first the problem definition and related work is given followed by the details of the method and the obtained tone mapping results. Lastly, the improvements made to the initial method using the findings of the user experiment is presented.

5.1 Problem Definition

Tone mapping operators aim to reduce the dynamic range of an HDR image to display it on an low dynamic range display devices. The existence of numerous tone mapping operators that are available paved the way for many studies that are conducted for selecting the best one [109]. However, tone mapping can be conducted for different purposes, and rendering the resulting images to follow a consistent style can be one of them. For example, in a movie production process, making all frames consistently tone mapped, regardless of the content of the frames, can be a desired operation to impart a certain look and feel to the viewers. Although obtaining different renderings from tone mapping operators is partly achievable by using different tone mapping parameters, some tone mapping operators have none [110] or few parameters [111] while others [112] have too many. Besides, even though one set of parameters are found for a certain image to depict a certain style of rendering, the same set of parameters would not yield with the same look when applied to other images.

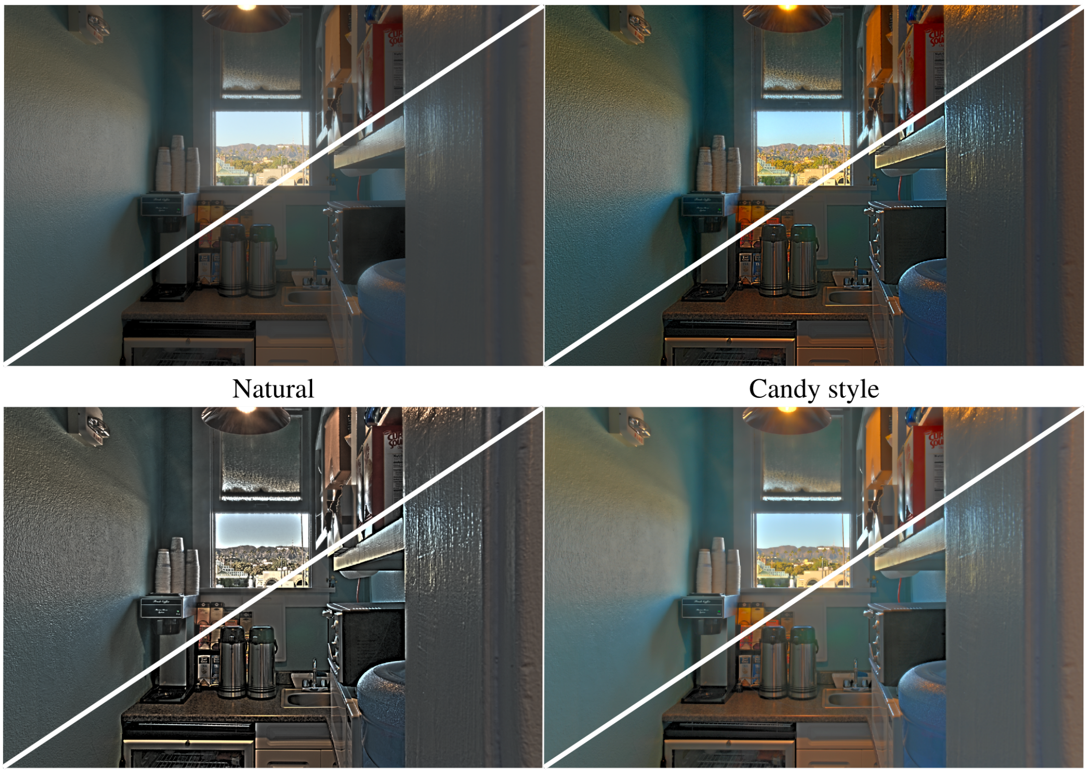
In this chapter, style based tone mapping operator that aims to consistently tone map different HDR images with the defined style is proposed. Figure 5.1 shows different styles for the same images created by an artist. The application depicted in this thesis approaches the problem of defining a style by presenting the user an HDR image and with the help of the tool developed, ask the user to change the parameters until the image matches the desired look. By repeating this process for a small set of calibration images, we aim to *learn* the style. Applying the learned style from calibration images to the previously unseen images is essentially a visual image similarity problem.

5.2 Related Work

Image reproduction is ultimately a subjective process where photographers and artists of all persuasions are free to reflect their own interpretation into the final rendering. For any task of reproduction, there is generally a vast range of choices from literal or realistic reproduction to those that significantly depart from reality. Since the early days of photography, many tools and systems have been developed to enable artists express their creativity [113, 114, 115, 116].

Tone mapping, in the context of HDR imaging, is no different in creative expression than the art of photography. There is essentially an infinite amount of choices that an artist can make when tone mapping an HDR image for display purposes (Figure 5.1). Perhaps, this is underscored by the large number of TMOs that have so far been developed that produce different outputs from the same HDR image [10].

Most TMOs fail to provide sufficient freedom to artists to express their creativity and rather focus on visibility or photographic look. However, there are a few TMOs that are designed to enable artistic freedom such as the stroke-based interface that was proposed by Lischinski et al. [117]. This algorithm allows the user to select image regions using brush strokes and set different tone mapping parameters for each region. These parameters are then extrapolated to the entire image using an edge-preserving energy minimization method. Paris et al. [118] gives users the possibility



Natural

Candy style

Grittily

Painterly

Figure 5.1: Different artists may prefer different tone reproductions of the same HDR image. The same artist may also choose to produce different styles based on the situational/contextual considerations. The images reproduced by a professional artist using a different tool (top half) are replicated using our operator (bottom half). Presented algorithm can also learn the styles of an artist and produce results consistent with them.

of enhancing details or edge-preserving smoothing while tone mapping. This method first applies detail modifying local Laplacian filters on the log intensity of the image and then map the intensities to displayable range by scaling. Aubry et al. [119] offers a fast implementation of local Laplacian filters and also demonstrates how this filter can be used for photographic style transfer. Given a model image, this iterative method uses local Laplacian filters for gradient histogram matching for local contrast modification and intensity histogram matching for global contrast modification. Tone mapping is achieved through the intensity matching since the method matches the same dynamic range of the model image.

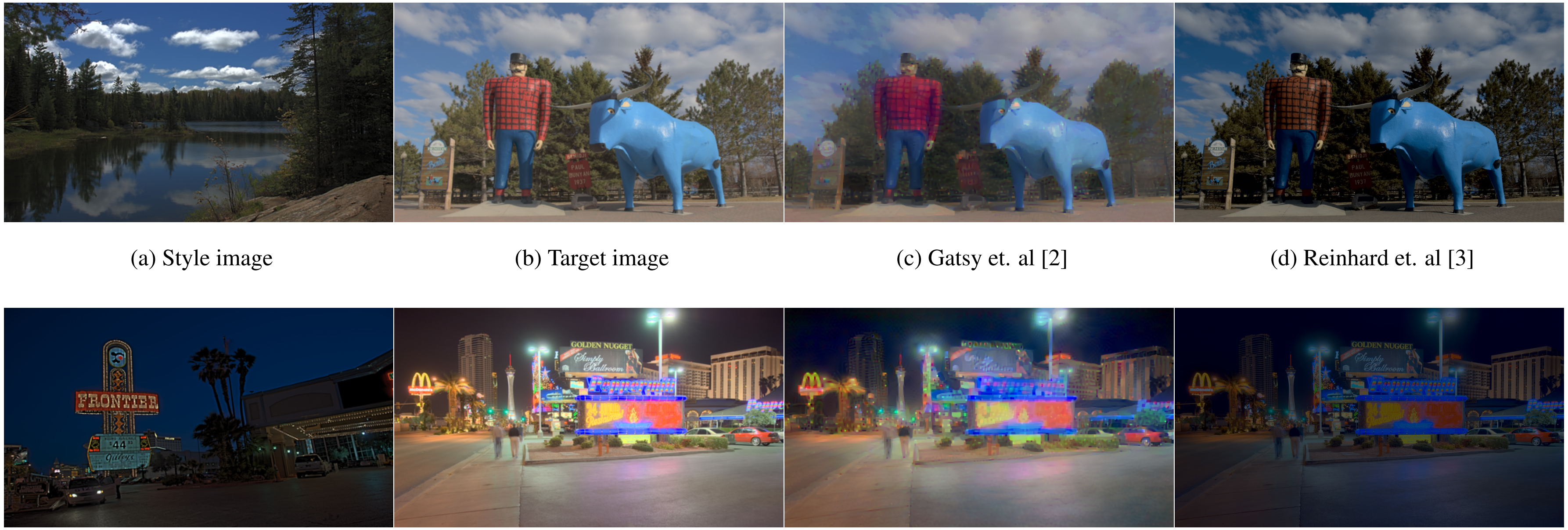
Another algorithm that allows a wide range of possibilities during tone mapping is the generic TMO [120]. This operator hinges on the idea that although a large number of TMOs exist, the majority of them can be modeled using a generic tone curve followed by local modifications. The authors have demonstrated that by using a small set of parameters, a skilled artist can match the output of any TMO with the generic TMO. Although these algorithms offer expressive freedom to the artist, they have no notion of *learning* the artist's preference. In fact, it has been shown by the authors that the same set of parameters can give rise to vastly different outputs for different images [120]. As such, using these operators to automatically tone map a large number of images is impractical.

There are several LDR retouching studies that are similar to the proposed style based tone mapping operator in the sense of enhancing images by using a set of training images [121, 122] or by automatic exposure correction [123] or artistic enhancement [124]. However, these studies do not focus on HDR image tone mapping.

Recently, deep neural networks are also used for learning image adjustment parameters. Some of these studies uses a training set of high quality images and learns the enhancement parameters in an unsupervised manner [125, 126, 127]. These methods aim to enhance the general quality of the given image instead of applying a specific style. Yan et al. [128] on the other hand, proposes to learn a computational model of a style from image pairs, an untouched image and a stylized version. Their method incorporates high level semantic information of the pixels as well as local and global features of the image. Another study with a similar goal, Hu et al. [129] uses unpaired data, a set of stylized images, for learning an operation sequence that will style an input image when applied. The benefit of learning modification steps is it makes the style explainable, which most of the deep learning based image enhancements methods lack. While these methods aim to learn predefined style, Lee et al. [130] uses deep learning based semantic search to find similar images from a large collection and applies the styles of these images to the given image, by using a transfer function for color and luminance statistics, which brings the benefit of automatically selecting candidate styles. Similar to the previous LDR retouching studies, these methods also do not aim to tone map HDR images.

Besides of these studies that learns image adjustment parameters using deep neural networks, initiated with the study of Gatsy et. al [2], deep neural networks are also used for style transfer between images. The algorithm introduced in [2] separates and recombines the content and style of images using different layers of pretrained VGG19 network. After Gatsy et. al [2], neural style transfer gained popularity and many studies focused on improving this method. Gatsy et. al [131] introduced control methods such as color, scale and spatial control, Gupta et. al [132] improves stability, Liu and Lai [133] added depth awareness, Huang and Belongie [134] made the neural style transfer more practical with a faster method and flexible user controls.

In Figure 5.2, the neural style transfer method [2] is applied to the tone mapped HDR images. All images are taken from Fairchild's HDR dataset [4]. Content images are tone mapped with Reinhard et. al [52], and style images are tone mapped manually with the method that is described in Section 5.3.1 according to the *candy* style shown in Figure 5.1. For neural style transfer method, *conv5_2* is used as content layer, while *conv1_1* is used as style layer, content weight $\alpha = 0.1$ and style weight $\beta = 10^4$. Style transfer results are given in Figure 5.2c and Figure 5.2g. As shown in the resulting images and stated in the original study [2], it is not possible to completely separate the content and style of the images, and this method performs pleasing when artistic images like paintings are used as style images and not so well when photographs are given as style images. It is possible to see this effect in the resulting images, the edges and some objects are deformed, it is not possible to read the writings in the images, and the images looks more like a painting although given style images do not have this property.



(a) Style image

(b) Target image

(c) Gatsy et. al [2]

(d) Reinhard et. al [3]

(e) Style image

(f) Target image

(g) Gatsy et. al [2]

(h) Reinhard et. al [3]

Figure 5.2: Style transfer results of Gatsy et. al [2] and color transfer results of Reinhard et. al [3].

This drawback of neural style transfer is addressed by Luan et. al [135] by preventing spatial distortions and allowing transformations only in color space. Also by including semantic segmentation, the style transfer between unrelated contents are prevented. While this method produces very pleasing results for style transfer between images depicting similar scenes but different color palette or time of the day, it is not fully free of distortions such as creating patterns on uniformly colored objects or changing the illumination on an object in an unnatural way. Also the success of the method depends on the underlying segmentation method. On the other hand, color transfer method of Reinhard et. al [3] is a global method that uses simple image statistics to transfer color from one image to another. In Figure 5.2d and Figure 5.2h, the result of this method on tone mapped HDR images are given. In both results, overall brightness and contrast decreased, although both style images have bright regions similar to the content images, i.e. clouds, small light sources.

In summary, style based tone mapping differs from the previous methods by means of learning different tone mapping styles from a small set of tone mapped calibration images. Then, new HDR images, which may be created from vastly different scenes, can be tone mapped according to these learnt styles. In this sense, style based tone mapping is the first method that allows to batch process a set of HDR images consistently according to the created style.

5.3 Method

Style based tone mapping [136] consists of two consecutive phases, namely *calibration* and *operation*. Calibration is the phase where the user defines a style and by manually adjusting the provided parameters tone maps a small set of images until the tone mapped image depicts the newly defined style. The next phase, operation is the phase that the user given HDR image is tone mapped automatically with the selected predefined style in the calibration phase by estimating the tone mapping parameters from the calibration image tone mapping parameters. Figure 5.3 shows the style based tone mapping algorithm steps and these steps are explained in detail in the following sections.

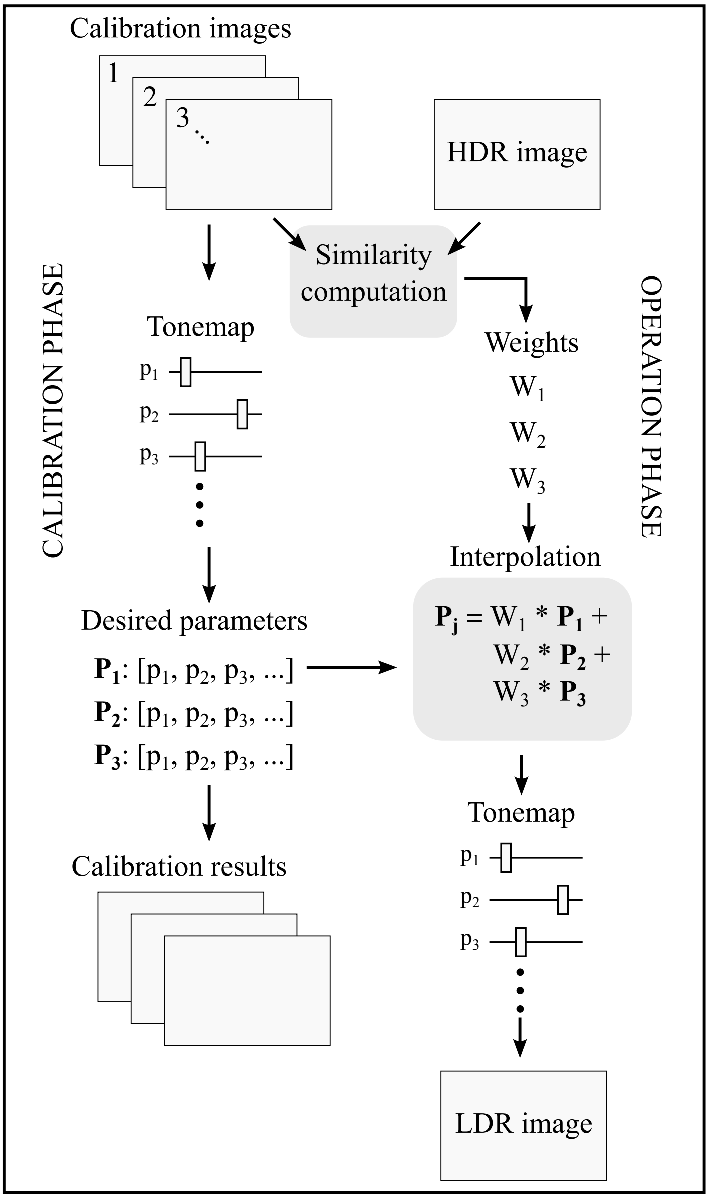


Figure 5.3: Style based tone mapping is comprised of a calibration and operation phase. The artist tone maps several images during calibration which results in a set of parameters. During operation, these parameters are interpolated based on image similarity to find the tone mapping parameters for the given image.

5.3.1 Calibration

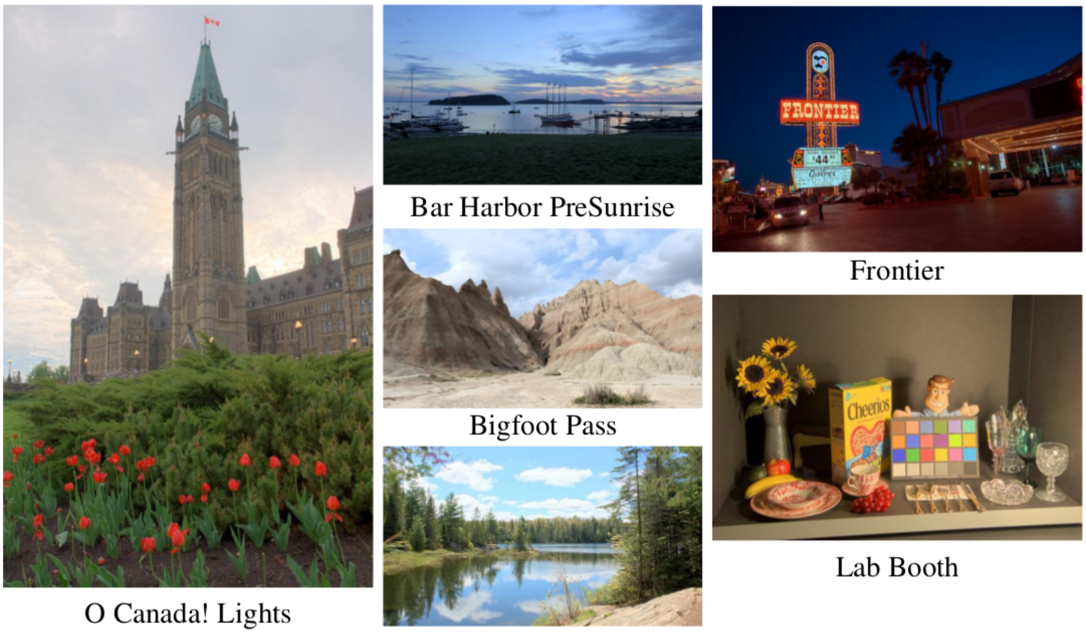
In the calibration phase, first the user is asked to pick a name for the new style to be created, and then the user is asked to tone map a fixed set of calibration images. Assigning a name to the style would be helpful for the user to stay consistent with the style while tone mapping the calibration images. In addition, once the style is created and saved, this will allow the user to have many predefined styles in preset library and reuse them.

Calibration images should be representative enough for different environments that has different characteristics and at the same time distinctive from each other as much as possible to keep the number of the calibration images low. The number of calibration images directly affects the time spent in the calibration phase, so although more calibration images would represent different environments better, in order not to overwhelm the user, the number of calibration images is chosen in a way that the user can finish the calibration phase in a reasonable time.

Calibration images are selected in a semi automatic way. First, all images are taken from Fairchild's HDR image dataset [4] and then converted a feature space and using k-means algorithm each image in the dataset assigned to a cluster, which resulted in six clusters given in Appendix A. After clusters are obtained, one calibration image is hand-selected from each cluster, yielding six calibration images in total, which are shown in Figure 5.4.

The calibration phase starts with naming the style, and then the user is presented with the tone mapping interface. With using this interface the user is required to tone map the calibration images one after another with the desired style. The manual tone mapping procedure is basically adjusting tone mapping parameters given in Table 5.1 with the provided sliders.

The tone mapping operator is a modified version of Generic TMO [137]. Generic TMO is able to model many existing tone mapping operators, both local and global with a tone curve followed by a spatial modulation function. It is noted in [137] that same set of parameters yields very different results for different images. Style based tone mapping uses Generic TMO with the following modifications.



Bar Harbor PreSunrise

Frontier

Bigfoot Pass

Lab Booth

O Canada! Lights

Peck Lake

Figure 5.4: Calibration images

Table 5.1: Tone mapping parameters for style-based tone mapping

Brightness (b)	Prct. mapped to half-max intensity
Contrast (c)	Slope of the tone curve at b
Black point (bp)	Prct. clamped to min intensity
White point (wp)	Prct. clamped to max intensity
Color saturation (c)	Saturation control exponent
Small detail strength (λ_s)	UM factor for small details
Medium detail strength (λ_m)	UM factor for medium details
Large detail strength (λ_l)	UM factor for large details

The tone mapping parameters of Generic TMO are replaced with their percentile counterparts in order to make the algorithm less image dependent and make it easier for user to have an understanding about the parameters. For example, the parameter *Brightness* in style based tone mapping with the value 50 would correspond to the median brightness value of the HDR image in Generic TMO's *b* parameter. Likewise, for the parameter *White point*, the value 95 would mean 5% of the brightest pixels will be burned out. As one may imagine, representing the tone mapping parameters as percentiles is not sufficient to achieve the same effect on different images. In order to achieve this, style based tone mapping uses parameter interpolation to use *similar* parameters to *similar* images as described in Section 5.3.2.

The second modification to Generic TMO belongs to spatial modulation. In [137], a linear combination of band-pass filters are used as spatial modulation function. These filters are from modified Cortex transform and applied after global tone curve modulation. On the other hand, in style based tone mapping, local modulation is applied in multiple scales and afterwards global operation is performed. This has the benefit of adjusting detail level before the HDR compression is applied and it gives better results.

For detail modulation several approaches has been tested, unsharp masking (UM), bileteral filtering [138] and gradient reversal removed BF [139]. While BF-based filters results with less halo, they are computationally expensive, unlike UM, which is prone to halos but computationally efficient. Besides, UM is shown to be improve sharpness and local contrast in an earlier study [140]. It is decided to use UM even though it may introduce halos. For some cases, halos may be also introduced by the user in order to create an unrealistic style.

The detail modulation is achieved by first creating three low-pass images in the logarithmic domain for $small$, $medium$ and $large$ details.

$$L'_{\sigma_s} = g_{\sigma_s} * L', \tag{5.1}$$

$$L'_{\sigma_m} = g_{\sigma_m} * L', \tag{5.2}$$

$$L'_{\sigma_l} = g_{\sigma_l} * L', \tag{5.3}$$

where $L' = logL$ and g_σ are 2D Gaussian filters are different scales. σ_s, σ_m, and σ_l are set to 0.0625%, 0.3125% and 0.625% of the minimum image dimension respectively.

Then these low pass images are used to enhance different scales with the chosen detail factor parameters λ.

$$L_{sm} = e^{L'+\lambda_s(L'-L'_{\sigma_s})+\lambda_m(L'_{\sigma_s}-L'_{\sigma_m})+\lambda_l(L'_{\sigma_m}-L'_{\sigma_l})} \tag{5.4}$$

L_{sm}, spatially modulated luminance image, then fitted to the tone curve as in [137].

$$TC(L_{sm}) = \begin{cases} 0 & \text{if } L'_{sm} \leq b - d_l \\ \frac{1}{2}c\frac{L'_{sm}-b}{1-a_l(L'_{sm}-b)} + \frac{1}{2} & \text{if } b - d_l < L'_{sm} \leq b \\ \frac{1}{2}c\frac{L'_{sm}-b}{1+a_h(L'_{sm}-b)} + \frac{1}{2} & \text{if } b < L'_{sm} \geq b + d_h \\ 1 & \text{if } L'_{sm} > b + d_h \end{cases} \tag{5.5}$$

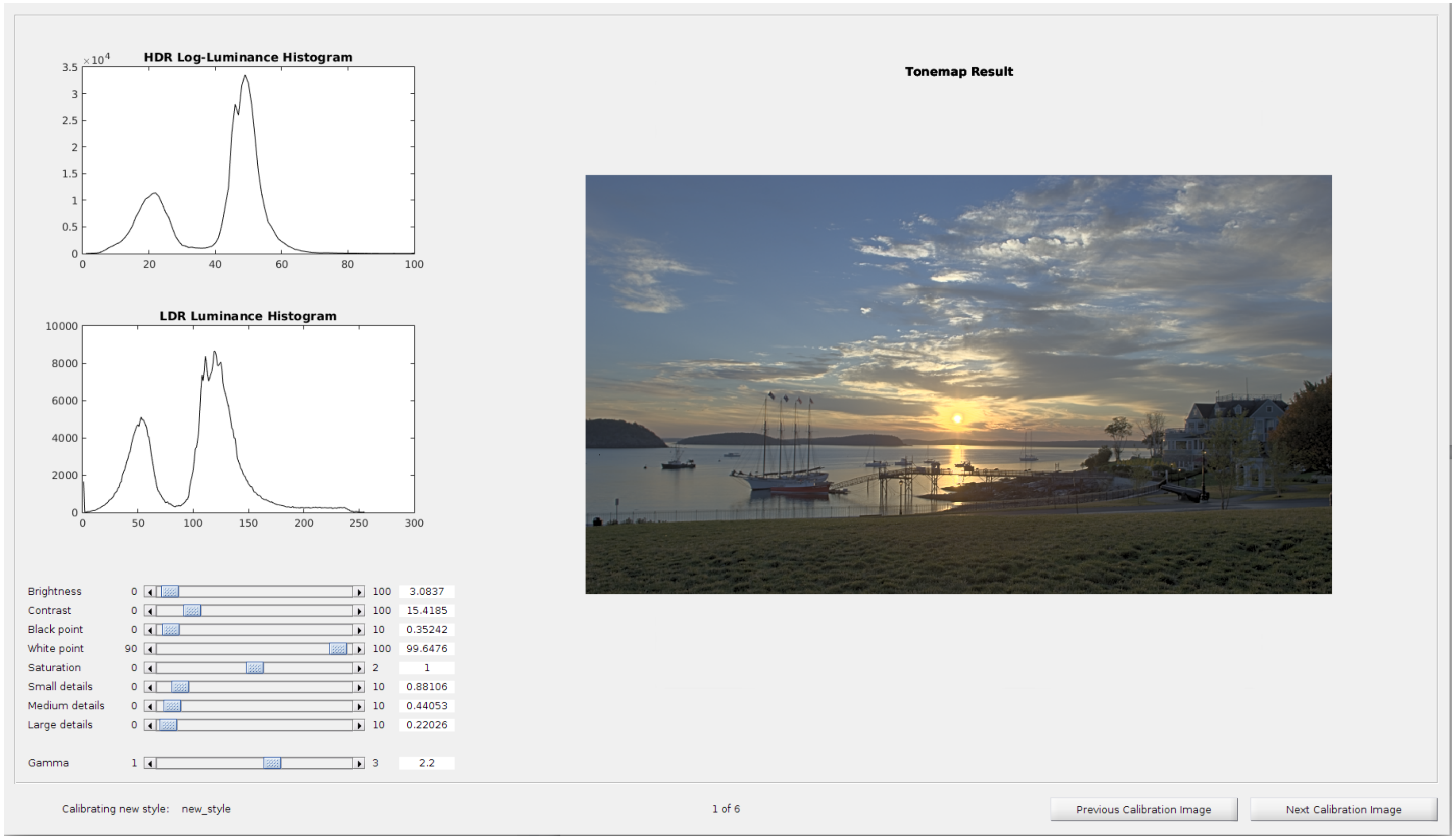
where L'_{sm} is the logarithm of the spatially modulated luminance c is the contrast, and parameters b, d_l and d_h are the absolute values of the user given parameters in percentiles for brightness, black point and white point respectively. Parameters a_h and a_l are contrast compression for light and dark areas computed from [137].

$$a_l = \frac{c.d_l - 1}{d_l} \text{ and } a_h = \frac{c.d_h - 1}{d_h} \tag{5.6}$$

In Figure 5.5, the user interface that allows the user to define a style for tone mapping during the calibration phase is shown. The tone mapping parameters can be easily adjusted with the sliders and the calibration image will be tone mapped and shown to the user in real time. The luminance histograms of log HDR and LDR images are also shown to aid the user and show the effect of the changes. After all of the calibration images are tone mapped, the style parameters are saved and the calibration phase is completed. Manually stylized calibration images following the four styles presented in Figure 5.1 are given in Appendix B.

5.3.2 Operation

In the operation phase, the user selects a style from the preset library that has been created in the calibration phase. Given a new HDR image to be tone mapped with the selected style, the tone mapping parameters *t* (given in Table 5.1) must be determined



Brightness	0	100	3.0837
Contrast	0	100	15.4185
Black point	0	10	0.35242
White point	90	100	99.6476
Saturation	0	2	1
Small details	0	10	0.88106
Medium details	0	10	0.44053
Large details	0	10	0.22026
Gamma	1	3	2.2

Calibrating new style: new_style

1 of 6

Figure 5.5: User interface for calibration phase

based on calibration images tone mapping parameters t_i. In this thesis, this problem is approached as an image similarity problem. If two images are similar according to a similarity metric, their tone mapping parameters should be also similar. After the distances between input image and the calibration images are obtained, the tone mapping parameters are calculated as inverse distance transform [141]:

$$\mathbf{t} = \frac{\sum_{i=1}^{N} \frac{1}{d(\mathbf{f},\mathbf{f_i})}\mathbf{t_i}}{\sum_{i=1}^{N} \frac{1}{d(\mathbf{f},\mathbf{f_i})}} \tag{5.7}$$

Here, $\mathbf{f}$ is the feature vector of the current input image and $\mathbf{f_i}$ the feature vector for the calibration image i and $\mathbf{t}$ its computed tone mapping parameters. Lastly, the function d calculates the similarity between two feature vectors.

In [136], images are represented with HSV histograms [142] and histograms of gradients [62] to capture colorimetric and structural properties of the images. HDR images varies highly on pixel values and it is hard to compare them directly. To overcome this, images are tone mapped to the interval $[0, 1]$:

$$L_{out} = \frac{L_{in}}{1 + L_{in}}, \tag{5.8}$$

and color channels are transformed with:

$$\mathbf{C_{out}} = \frac{\mathbf{C_{in}}}{L_{in}} L_{out}. \tag{5.9}$$

The feature vector is then computed using transformed values, as a 60 dimensional vector, $3x15$ bins for HSV histogram and 15 bins for gradient histogram.

Unfortunately, treating histograms as high dimensional points and computing their Euclidean distances does not yield correct results as this ignores the proximity information of the bins. For instance, although the histogram $H1 = (1, 0, 0, \cdots, 0)$ is closer to $H2 = (0, 1, 0, \cdots, 0)$ than $H3 = (0, 0, 1, \cdots, 0)$, their Euclidean distances are equal. To circumvent this problem, each histogram is convolved with a 1-D Gaussian $(\sigma = 0.7)$ prior to computing their distances [142]. Circular similarity of the hue histogram is also accounted. Thus the final distance metric between two feature vectors f_i and f_j become:

$$d(\mathbf{f_i}, \mathbf{f_j}) = (\mathbf{f_i} - \mathbf{f_j})^T (\mathbf{f_i} - \mathbf{f_j}), \tag{5.10}$$

where f is the combined histogram. This metric is used to measure the similarity between input HDR image and calibration images. Also the same distance metric is used to cluster Fairchild dataset in order to pick calibration images.

After tone mapping parameters are calculated with parameter interpolation, HDR image is tone mapped and presented to the user in a similar user interface like Figure 5.4. User can do the final adjustments and save the tone mapped LDR image. Figure 5.6 shows a gallery of the obtained tone mapping results. Note that despite the large variation of the image content, the selected styles are successfully applied to each image.

To sum up, tone mapping can be an artistic process that would result with different look and feel of the same scene instead of achieving a single "correct" look. The tone mapping methodology given in this chapter is versatile enough to create new styles with a small set of calibration images. And more importantly, with the leverage of image similarity, it is possible to apply the created styles to new HDR images automatically which can have quite different characteristics. However the similarity approach given in this chapter is rather straight forward and there is room for improvement. In the following section, two different similarity models backed up by a user study are suggested to improve style-based tone mapping.

5.4 Improvements to Style-based Tone Mapping

Although style based tone mapping has achieved some success for consistently tone mapping different images, the image similarity method given in Section 5.3.2 can be improved with the findings from the conducted user study. In this section, two modifications of this method that are made possible by the experimental findings of the user study is given. The first technique uses all of the image features utilized in Section 5.3.2 with different weights to estimate tone mapping parameters in the operation phase. Meanwhile, the second technique relates tone mapping parameters

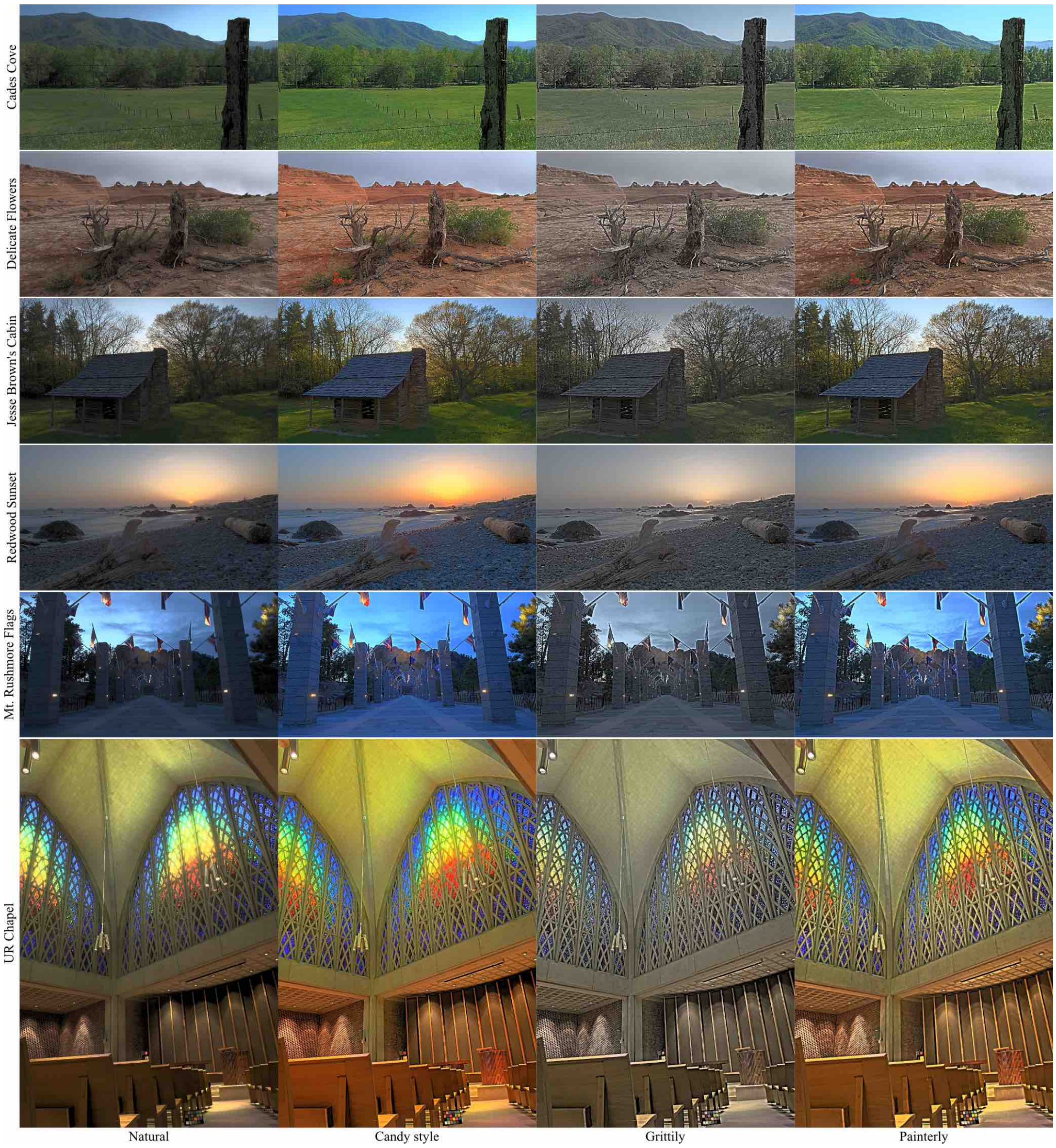


Figure 5.6: All of these images are automatically tone mapped using the four styles that are generated.

and image features for the estimation.

5.4.1 Parameter Interpolation with All Features

In the first modification, the model features given in Table 3.1 are extracted from the selected HDR image and calibration images. Then, the distances between these features are calculated separately using the corresponding distance metrics given in the same table. After that, the weighted average of these feature distances are calculated. The weights used are the coefficients of the logistic regression model (Equation 4.24) obtained from the user experiment, with the idea that less important features should also contribute less to the distance. This operation can be summarized with the following equation:

$$d_i = \sum_{j=1}^{6} c_j d_j(\mathbf{f_i}, \mathbf{f_{ij}}) \tag{5.11}$$

where c_j is the coefficient of the j^{th} feature, $\mathbf{f_j}$ is the j^{th} feature of the input image, $\mathbf{f_{ij}}$ is the same for the i^{th} calibration image, and finally d_j is the distance metric for the j^{th} feature. The result d_i represents the combined distance between the input image and the corresponding calibration image. These combined distances are calculated between the selected HDR image and all calibration images. The tone mapping parameters for the selected HDR image are then interpolated using inverse distance transform as in Equation 5.7. This method differs from the initially presented approach in several aspects:

- Using a different and more representative set of features, luminance which is one of the most important features of HDR images has a separate feature vector,
- More suitable distance metrics for feature types, for example, EMD takes into account bin proximity for calculating differences between histograms,
- Instead of using a single fused feature vector, each feature distance calculated separately,

Table 5.2: Model features used for interpolation of tone mapping parameters.

Tone mapping parameter	**Model feature**
Brightness (t_b)	Luminance
Contrast (t_c)	Luminance
Black point (t_{bp})	Luminance
White point (t_{wp})	Luminance
Color saturation (t_s)	Color
Small detail strength (t_{λ_s})	Texture
Medium detail strength (t_{λ_m})	Texture
Large detail strength (t_{λ_l})	Texture

- Employing a weighted average of feature vector distances with weights obtained from a user experiment, compared to using equal weight.

In Figure 5.7, the modified style-based tone mapping algorithm is shown. Note that while the calibration phase has not changed, the operation phase has been adjusted with the changes listed above compared to the initial version of the algorithm given in Figure 5.3.

5.4.2 Parameter Interpolation with Related Features

While the previous approach calculates a single distance value between the given image and calibration images and use this value to interpolate all tone mapping parameters, the second modification described in this section relates the model features with the tone mapping parameters and interpolates individual tone mapping parameters with different weights. In order to achieve this, the relationships defined in Table 5.2 are used.

As an example, the brightness parameter t_b is computed by interpolating the t_{b_i} parameters of the calibration images by using the similarity between the luminance

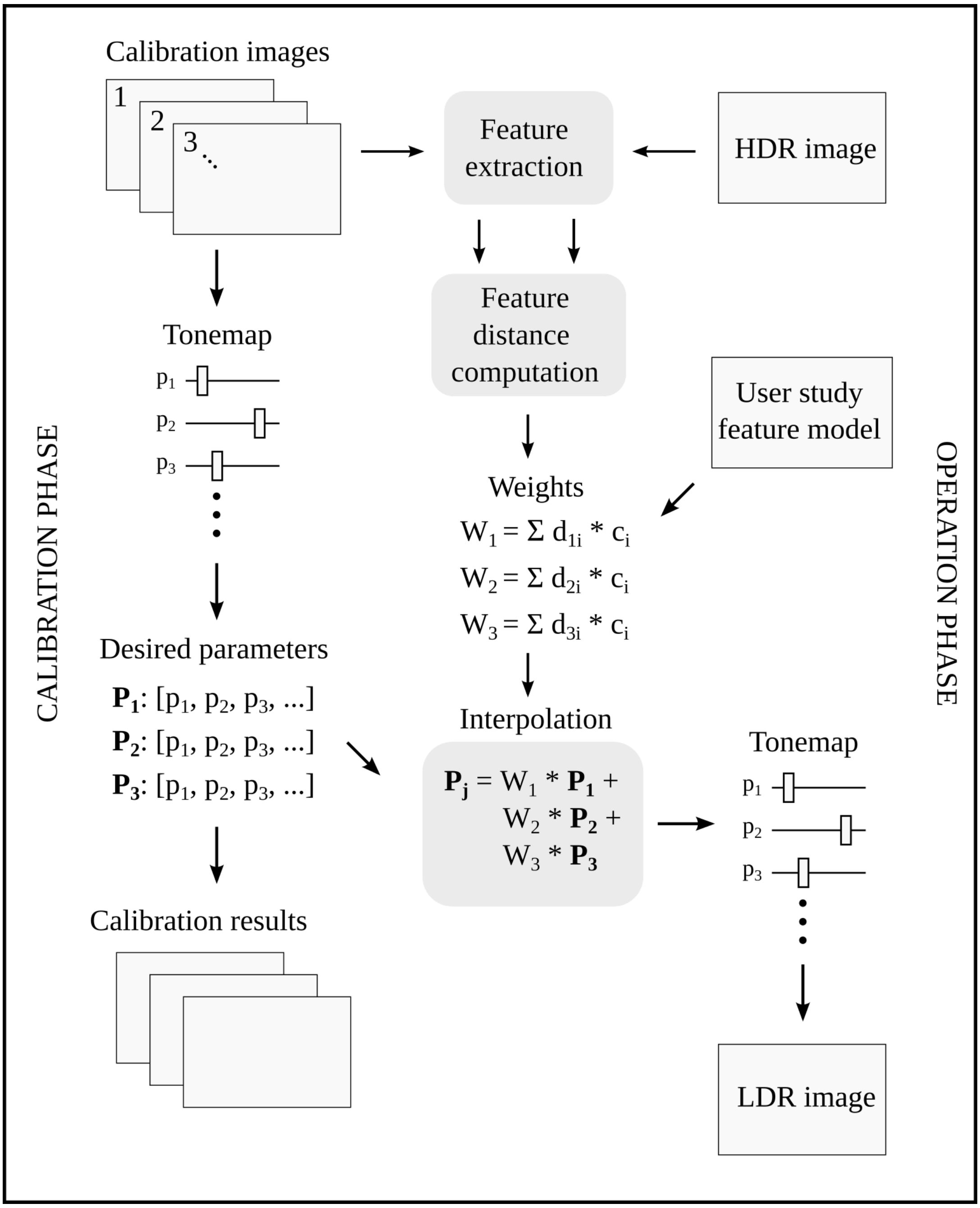

Calibration images
1
2
3
Feature extraction
HDR image
Tonemap
p_1
p_2
p_3
Feature distance computation
User study feature model
CALIBRATION PHASE
OPERATION PHASE
Weights
$W_1 = \Sigma\, d_{1i} * c_i$
$W_2 = \Sigma\, d_{2i} * c_i$
$W_3 = \Sigma\, d_{3i} * c_i$
Desired parameters
$\mathbf{P_1}$: [p_1, p_2, p_3, ...]
$\mathbf{P_2}$: [p_1, p_2, p_3, ...]
$\mathbf{P_3}$: [p_1, p_2, p_3, ...]
Interpolation
$\mathbf{P_j} = W_1 * \mathbf{P_1} + W_2 * \mathbf{P_2} + W_3 * \mathbf{P_3}$
Tonemap
p_1
p_2
p_3
Calibration results
LDR image

Figure 5.7: Modified style-based tone mapping algorithm with the findings of the user experiment. Compared to the initial method given in Figure 5.3, more representative set of features are extracted and distances between these features are computed separately. Besides, instead of using equal weights on feature distances, the weights estimated from user study are employed.

features:

$$t_b = \frac{\sum_{i=1}^{N} \frac{1}{d_{lum}(lum,lum_i)} t_{b_i}}{\sum_{i=1}^{N} \frac{1}{d_{lum}(lum,lum_i)}} \tag{5.12}$$

Other tone mapping parameters that are related to the model features are interpolated analogously. Because GIST and deep learning features are not directly linked to a specific appearance phenomenon but are measures of overall similarity between the given images, they are not directly linked to specific tone mapping parameters. Instead these features are experimented with merging them using the individually interpolated parameters as follows:

$$\mathbf{t} = w_0\mathbf{t_0} + w_1\mathbf{t_1} + w_2\mathbf{t_2}, \tag{5.13}$$

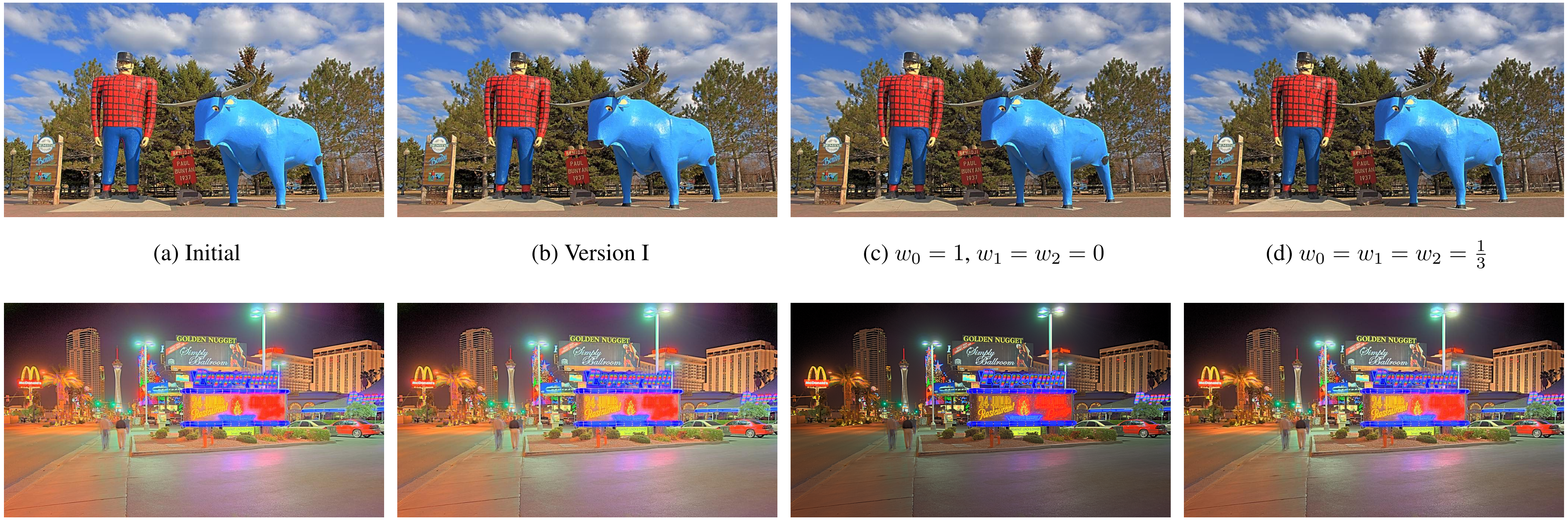
where $\mathbf{t_0}$ represents individually interpolated TMO parameters (given in Table 5.2), $\mathbf{t_1}$ TMO parameters interpolated as a whole using GIST similarity only, and $\mathbf{t_2}$ TMO parameters interpolated as a whole using solely deep learning feature similarity. The weights control the influence of TMO parameters that are computed by using these different approaches.

5.4.3 Results

In Figure 5.8, several results are compared that obtained by using the initial style based tone mapping method as well as with the two modifications proposed in the previous sections, Section 5.4.2 and Section 5.4.1.

In the first row of the figure, the tone mapping results of the "Paul Bunyan" scene from the HDR Photographic Survey [4] is shown. This scene depicts a bright outdoors environment with colorful foreground objects. It may be noted that all results are similar but the individual parameter interpolation with equally weighted GIST and deep learning features (d) has slightly higher contrast (please refer to Appendix C with high resolution images for better comparison). The overall *candy* style is preserved in all images.

In the second row of the figure, “Peppermill” night scene from the same dataset is shown. This is a night scene of a street with some bright lights and banners. For this scene the difference of the second modification, parameter interpolation with related features, is more clear as images in (c) and (d) exhibit a darker rendering, which is more suitable for a night scene. The reason for this darkening effect is that the brightness parameter, t_b, for tone mapping becomes more similar to the t_b parameter of the night image in the calibration images due to the similarity of the *luminance* features between these images. The addition of GIST and deep learning features with equal weight in (d) yields a slightly brighter image compared to (c). Similar to the results of the previous image overall *candy* style is preserved also in this image. Appendix C has higher resolution versions of the results for more clear observation of the differences.



(a) Initial

(b) Version I

(c) $w_0 = 1, w_1 = w_2 = 0$

(d) $w_0 = w_1 = w_2 = \frac{1}{3}$

(e) Initial

(f) Version I

(g) $w_0 = 1, w_1 = w_2 = 0$

(h) $w_0 = w_1 = w_2 = \frac{1}{3}$

Figure 5.8: Application of the user study findings for the style-based tone mapping problem. Initial results are shown in the first column, followed by Version I in the second column and two variants of Version II in the last two columns.

CHAPTER 6

DISCUSSION

In this thesis, visual similarity problem for HDR images is experimentally investigated. Two user experiments are conducted through a crowdsourcing platform and several image features are evaluated with respect to the data collected via these two experiments. Evaluation is performed both on individual features and on their combination. When combining features, two models both of which using logistic regression are considered. Although both models are found to perform comparably, the second one permits direct one-to-one comparison, which makes it more suitable for practical applications that relies on assessment of similarity between image pairs.

One such application, that uses image similarity to tone map similar images in a similar way according to a style, namely style-based tone mapping is given. The first step of this tone mapping scheme is the creation of an artistic style by manually tone mapping a set of calibration images, then when an input HDR image is given, the similarity between this new image and calibration images are calculated, and the tone mapping parameters that will achieve the same style for the new images are estimated automatically from the parameters of the calibration images. Calibration images are picked from an HDR image dataset in a way that this set of images are dissimilar from each other as much as possible to cover the HDR image space as much as possible. At the beginning, a basic image similarity model are used for parameter estimation and then two new methods are proposed that improves this similarity model by incorporating the findings that are obtained from the user study.

Key observations obtained in this thesis are the following:

1. When properly tone mapped images are used as compared to using either orig-

inal or linearly scaled HDR images, higher correlations with human responses are obtained.

2. Most tone mapping operators (TMOs) yield comparable performance.
3. Deeply learned features, in comparison to hand-crafted features, correlate better with the human responses.
4. Among hand-crafted features, GIST yields the highest correlation, followed by color, luminance, and texture.
5. Combination of features performs better than individual features.
6. All of the estimated correlations for the second experiment are higher in comparison to those for the first experiment.
7. Applying the same tone mapping parameters directly to different HDR images yields different and mostly unpleasing tone mapping results.
8. It is time consuming to manually tone map HDR images according to a style.
9. Using the similarities to a small but diverse set of images, it is possible to consistently tone map HDR images according to a style, even if the images have different characteristics.
10. In the context of style-based tone mapping, better results are obtained when more descriptive features are used, feature distances calculated separately and suitable weights are used for features while calculating image similarities.

The first observation highlights the importance of using tone mapped data for HDR image similarity. While tone mapping is a lossy process, it brings the data to a more meaningful range for the computation of most features. However, some features are less dependent on tone mapping. For instance the texture feature represented by the histogram of oriented gradients is found to produce about the same correlation regardless of whether HDR or tone mapped data is used. This is followed by the color feature represented by 2D chromaticity histogram. Among the hand-crafted features the largest difference is observed for luminance feature when tone mapped data is used for representation. This can be expected as non-linear luminance compression often eliminates large gaps in luminance histogram where little useful information is present.

Perhaps unexpectedly, the second observation suggests that TMOs perform comparably. Although there exists a large number of TMO evaluation studies, we are not aware of any work that compares TMOs for the task of HDR image similarity. The lowest performing operator is found to be Pattanaik et al.'s [54] algorithm. It is, however, known that this algorithm highly depends on calibrated input data and viewing conditions as it tries to accurately model the human visual system.

As for the third observation, it is not surprising to find that features obtained from a DCNN [67] trained over a large image dataset [85] outperform simple hand-crafted features. Similar findings are reported by image retrieval studies conducted for low dynamic images [70, 71]. For HDR images, our findings indicate that deep features are mostly useful if the images are tone mapped to the 8-bit per color channel domain first. This is also expected as the training data of DCNNs are comprised of such images.

The fourth observation indicates that the GIST descriptor surpasses the other hand-crafted features, texture, luminance, and color for HDR image similarity. In addition to outperforming them, in fact, it performs surprisingly consistently across different processing types. Despite having a smaller correlation with the user responses than the deep features, it exhibits less variability overall. This may be a desirable property for different applications as it appears to be minimally affected by how an HDR image is processed.

As expected, the fifth observation points to the findings that a combined feature with the learned weights performs better than individual features, which holds true for both of the logistic regression models. Although deeply learned features outperform the hand crafted features, and that is also observed with higher weights in the models, other features also contribute to the performance of the models.

Regarding to the sixth observation, it can be argued that seeking multiple consistent responses by the participants are important; not only for developing a more reliable model but also for assessing the correlation of different features with user responses.

For instance, inspection of Tables 4.1 and 4.2 reveals that while deep features correlate better with the user responses, this difference is clearly magnified for the second phase of the experiment. In other words, as the experimental findings become more reliable the merits and drawbacks of different features become more noticeable.

The seventh observation is, due to the vastly varying nature of HDR images, using the same set of tone mapping parameters that gives pleasing results for an image, may result with very poor rendering for another image. For example, since HDR images luminance values are boundless, a high cut-off value for a night scene might be very low for a day time scene, yielding with a really dark image that loses many details otherwise could be rendered.

The eighth observation states that, since tone mapping parameters can not be applied directly to different images, for each new image that depicts a different scene, the parameters need to be adjusted in a way that the resulting image follows the same style. This is a time consuming and not so straight forward process, searching for a set of parameters manually by adjusting parameters one by one, observing the effect on the resulting image and updating the parameters in a way that hopefully results with a rendering that is consistent with the other images. Keeping this observation in mind, the number of calibration images that need to be manually tone mapped are kept as small as possible.

The ninth observation features that although HDR images can be very diverse and for uncalibrated HDR images the same values do not correspond to the same physical or perceptual brightness, it is possible to tone map these images in a consistent way following a user defined style by using similar parameters for similar images. In style-based tone mapping, in order to achieve this, a small set of calibration images are used and the tone mapping parameters are estimated with the similarities between these calibration images and the input image. Although the number of calibration images are quite low, this is preferred since it directly effects the duration of manual style creation, the method is able to capture the style and consistently tone map different HDR images as presented in Figure 5.6.

However, it should be noted that the selection of calibration images are important for the performance of the style-based tone mapping. The set of selected images should be as large as possible to be able to find a similar image in this set to the input HDR image. On the other hand, with a large the number of the calibration images, calibration phase, the style generation, will take too much time that the method become unpractical. Therefore, there is a trade off between the expressiveness and the number of calibration images. In order to keep the number of images as low as possible while keeping the expressiveness high, the calibration images are picked in a way that these images are dissimilar to each other as much as possible. In this thesis calibration images are hand selected from a clustered HDR dataset of natural images to be able to handle the most common HDR image types. Also, the calibration images capture varying sceneries like night, bright sunny day, cloudy day, sunset, indoor, outdoor etc. However, with this set of calibration images, this tone mapping scheme may not give successful results for computer generated scenes or modern art. For different domains, domain specific calibration images may give better results.

Finally, as the last observation, not surprisingly, the used similarity model has a direct effect on the performance of the application. In the initial version of style-based tone mapping which is described in Section 5.3.2, the image representation consists of a single fused vector of histograms calculated from HSV color space and magnitude of gradients. To improve the image representation, these features are replaced with the features introduced in Chapter 3 and analyzed in Chapter 4 against the user responses. The first difference is using luminance as a separate feature represented as a higher resolution histogram with the motivation of luminance being one of the most important features in HDR tone mapping. Secondly, Lab color space is preferred over HSV color space due its perceptually uniform characteristic. Lastly, besides of low level image features, high performing features like GIST descriptor and deeply learned features are also included.

In addition to the image representation, the metrics that are used to estimate the distance between features are also plays an important role in image similarity. The distance metric used in the initial version of style-based tone mapping, given in Equation 5.10 is used to calculate the distance between image features that are fused into a single vector. Instead, as given in Section 5.4, calculating the distances between features separately using the proper distance metrics yields better results. For example, EMD takes into account of bin proximity when calculating distances between features that are represented with a histogram, which makes it a more suitable metric.

Another point that needs consideration is the contribution of the image features to the similarity model when multiple image features are employed. In the initial version of the style-based tone mapping operator, all features are fused using the equal weights. On the other hand, as showed in Chapter 4, a similarity model that using different weights for different features can be derived. With the conducted user experiment data, the weights for the features are estimated with a logistic regression and these weights are used for similarity calculation in style-based tone mapping, as described in Section 5.4.1. With all these modifications that are made to improve the similarity calculation of the style-based tone mapping operator, results that are more compatible with the given style and the image characteristics are obtained as shown in Figure 5.8.

While these modifications that improves the similarity model in a general sense yield with better tone mapping results, like in many other applications, it is possible to refine the results further by adding problem specific heuristics. In Section 5.4.2, instead of using all feature distances to estimate each tone mapping parameter, image features that are directly related with the tone mapping parameters are used for the estimation of tone mapping parameter and the features that do not directly relate to a tone mapping parameter added to the overall estimation. As shown in Figure 5.8 this approach results with capturing image characteristics better.

Although, several style-based tone mapping results are provided in Section 5.3.2 and comparison of results with improved similarity models in Section 5.4, it is not straight forward to measure the performance of the proposed tone mapping methodology nor the improvements to it. From the style perspective, there is no ground truth that will allow to calculate the differences or compare image statistics. On the other hand, for improvements to the style-based tone mapping, some image statistics between resulting images can be calculated, like the change in mean brightness or contrast. However, image statistics do not directly translate to perception and even if the measured effects are also visible when results are compared, it does not imply better tone mapping results or compliance with the given style. As a future research direction, some perceptual experiments can be conducted to measure the performance of the proposed tone mapping methodology, in terms of depicting the given style or image quality of the tone mapped images or evaluating the effect of employed similarity models.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

In this thesis, HDR image similarity problem is investigated through an experimental user study and two similarity models are proposed to model this subjective phenomenon. In the user experiment, one reference image and two test images are presented and the participants are asked to choose the test image that is similar to the given reference image. The definition of similarity is not given in order not to constrain the users and introduce bias to the data. The experiment is conducted through a web-based interface which makes it possible to collect the responses of more than 1200 users that are reached through a crowdsourcing platform. To keep the data quality high, several measures are taken such as including verification trials through the experiment.

In order to gain insight about the subjective evaluation of HDR image similarity, the experiment data is analysed with different TMOs, image features and distance metrics. The selected TMOs are heavily used TMOs that prioritize different aspects of tone mapping. The image features include commonly used low level features as well as deep learning features. The distance metrics are chosen in a way that is suitable for the selected image feature representations. Correlations between human judgements and these quantitative features are computed to assess how much each feature contributes to visual similarity. Lastly, two combined features which perform better than individual image features are also proposed with the weights of contributing features are estimated from the user data.

Reliable assessment of image similarity lies at the hearth of many computer vision applications. In this thesis, an application is given to demonstrate how HDR image similarity can be used for consistently tone mapping various HDR images following a created style. This tone mapping method, namely style based tone mapping, estimates the tone mapping parameters for the given image from the tone mapping parameters of a set of manually tone mapped calibration images using the similarity between the given image and calibration images. The initial basic similarity model of the operator is then improved with two different approaches derived from the findings of the user experiment. The improvements yielded with better tone mapping results in terms of style imitation while keeping the image characteristics intact.

Although a small number of calibration images are used, the tone mapping operator is shown to depict the given style and produce satisfying tone mapping results for HDR images that has different image statistics. It should be also noted that, one of the main benefits of the presented operator is once the calibration images are tone mapped, the new HDR images can be automatically tone mapped with the created style without any intervention such as manual search of optimal parameters. Hence, it makes this operator a suitable choice for the applications where batch tone mapping of HDR images is necessary.

One of the limitations of the proposed tone mapping operator is the performance of the method can be affected by calibration image selection. The operator tone maps similar images similarly, if the input HDR image is too different than the calibration images, the estimated parameters may not give pleasing results. This problem can be prevented by selecting a different set of calibration images for specific domains.

The conducted image similarity experiment also has certain limitations and drawbacks. Firstly, it relies on crowdsourcing, which was necessary to reach a wider audience to collect as much as data possible but made it impossible to control the viewing conditions of the participants. Different results could have been obtained if the experiments were done in a laboratory environment with controlled display and lighting conditions.

Secondly, the participants compared the HDR images on standard monitors and used sliders to visualize different image regions that are visible on different exposures. Similarly, different results could have been obtained if participants viewed the images on an HDR display.

Lastly, in the experiment, the meaning of similarity is not given intentionally and left to the interpretation of the participants. To reduce this uncertainty, future studies may explicitly define what is meant by similarity such as object similarity, color similarity, indoor-outdoor similarity or time-of-day similarity.

7.1 Future Work

Although image similarity is an extensively studied topic, HDR image similarity has not gained as much attention yet. To investigate this subjective phenomenon, further experiments which consider ranking and rating tasks as well as pairwise comparisons can be conducted. Also, the proposed models can be extended with different types of features. Evaluations may include DCNNs that are either fine-tuned or trained with HDR data from the ground up.

Given the large number of image quality datasets and subjective evaluations in the form of mean opinions scores (MOS), whether image quality and similarity correlate with each other in the context of HDR imaging can be investigated. Image saliency can also be taken into account for similarity judgements as it was found to improve performance in some other domains [143]. Perhaps most importantly, the effect of calibrated HDR images for image similarity and retrieval tasks can be studied. As objects are represented with their true luminances in calibrated data, this may simplify the similarity assessment between the images. Also, with emerging standards for HDR video streaming such as HDR10+ and Dolby Vision, the HDR video similarity problem will gain importance in near future.

Finally, in the recent years, HDR video tone mapping is gaining popularity and more and more studies are conducted to tone map videos coherently [144]. Style-based tone mapping can be extended for HDR video tone mapping, where the number of images are too many to pick tone mapping parameters manually and often different scenes needs to have the same look of the general style of the video.

REFERENCES

[1] "Hdr image gallery." `http://pfstools.sourceforge.net/hdr_gallery.html`. Accessed: 2017-08-26.

[2] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423, 2016.

[3] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34–41, 2001.

[4] M. D. Fairchild, "The hdr photographic survey," in *Color and Imaging Conference*, pp. 233–238, Society for Imaging Science and Technology, 2007.

[5] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern recognition*, vol. 40, no. 1, pp. 262–282, 2007.

[6] Y. Kleiman, G. Goldberg, Y. Amsterdamer, and D. Cohen-Or, "Toward semantic image similarity from crowdsourced clustering," *The Visual Computer*, vol. 32, no. 6-8, pp. 1045–1055, 2016.

[7] S. Rawat, S. Gairola, R. Shah, and P. Narayanan, "Find me a sky: A data-driven method for color-consistent sky search and replacement," in *International Conference on Multimedia Modeling*, pp. 216–228, Springer, 2018.

[8] A. S. Glassner, *Principles of digital image synthesis: Vol. 1*, vol. 1. Elsevier, 1995.

[9] G. Ward and R. Shakespeare, "Rendering with radiance: the art and science of lighting visualization," 1998.

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

[11] J. Tumblin and H. Rushmeier, “Tone reproduction for computer generated images,” *IEEE Computer Graphics and Applications*, vol. 13, pp. 42–48, November 1993.

[12] G. Ward, H. Rushmeier, and C. Piatko, “A visibility matching tone reproduction operator for high dynamic range scenes,” *IEEE Trans. on Visualization and Comp. Graphics*, vol. 3, no. 4, 1997.

[13] J. A. Ferwerda, S. Pattanaik, P. Shirley, and D. P. Greenberg, “A model of visual adaptation for realistic image synthesis,” in *SIGGRAPH 96*, pp. 249–258, August 1996.

[14] R. Mantiuk, S. Daly, and L. Kerofsky, “Display adaptive tone mapping,” *ACM Trans. Graph.*, vol. 27, pp. 68:1–68:10, 2008.

[15] P. E. Debevec and J. Malik, “Recovering high dynamic range radiance maps from photographs,” in *SIGGRAPH 97 Conference Proceedings*, pp. 369–378, August 1997.

[16] P. Sen, N. K. Kalantari, M. Yaesoubi, S. Darabi, D. B. Goldman, and E. Shechtman, “Robust patch-based hdr reconstruction of dynamic scenes.,” *ACM Trans. Graph.*, vol. 31, no. 6, pp. 203–1, 2012.

[17] N. K. Kalantari and R. Ramamoorthi, “Deep high dynamic range imaging of dynamic scenes.,” *ACM Trans. Graph.*, vol. 36, no. 4, pp. 144–1, 2017.

[18] J. M. Theodor and R. S. Furr, “High dynamic range imaging as applied to paleontological specimen photography,” *Palaeontologia Electronica*, vol. 12, no. 1, pp. 1–30, 2009.

[19] J. Happa, A. Artusi, S. Czanner, and A. Chalmers, “High dynamic range video for cultural heritage documentation and experimental archaeology,” in *Proceedings of the 11th International conference on Virtual Reality, Archaeology and Cultural Heritage*, pp. 17–24, 2010.

[20] E. Grinzato, G. Cadelano, P. Bison, and A. Petracca, “Seismic risk evaluation aided by ir thermography,” in *Thermosense XXXI*, vol. 7299, p. 72990C, International Society for Optics and Photonics, 2009.

[21] H. Cai, “High dynamic range photogrammetry for synchronous luminance and geometry measurement,” *Lighting Research & Technology*, vol. 45, no. 2, pp. 230–257, 2013.

[22] S. Harifi and A. Bastanfard, “Efficient iris segmentation based on converting iris images to high dynamic range images,” in *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)*, pp. 115–119, IEEE, 2015.

[23] A. Rizzi, B. R. Barricelli, C. Bonanomi, L. Albani, and G. Gianini, “Visual glare limits of hdr displays in medical imaging,” *IET Computer Vision*, vol. 12, no. 7, pp. 976–988, 2018.

[24] K. C. Brown, T. Bryant, and M. D. Watkins, “The forensic application of high dynamic range photography,” *Journal of Forensic Identification*, vol. 60, no. 4, p. 449, 2010.

[25] H.-H. P. Wu, Y.-P. Lee, and S.-H. Chang, “Fast measurement of automotive headlamps based on high dynamic range imaging,” *Applied optics*, vol. 51, no. 28, pp. 6870–6880, 2012.

[26] “Hdr10+ system whitepaper.” `https://hdr10plus.org/wp-content/uploads/2019/08/HDR10_WhitePaper.pdf`. Accessed: 2021-01-28.

[27] C. Chinnock, “Dolby vision and hdr10,” *White Paper of Insight Media*, 2016.

[28] H. Yeganeh and Z. Wang, “Objective quality assessment of tone-mapped images,” *IEEE Transactions on Image processing*, vol. 22, no. 2, pp. 657–667, 2012.

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

[30] A. Chalmers, P. Campisi, P. Shirley, and I. G. Olaizola, *High dynamic range video: concepts, technologies and applications*. Academic Press, 2016.

[31] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile hdr video production system," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, pp. 1–10, 2011.

[32] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays," in *Digital Photography X*, vol. 9023, p. 90230X, International Society for Optics and Photonics, 2014.

[33] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," in *ACM SIGGRAPH 2004 Papers*, pp. 760–768, 2004.

[34] S. STANDARD, "Dynamic metadata for color volume transform–core components," 2016.

[35] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for hdr image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, pp. 1–18, 2015.

[36] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Transactions on graphics (TOG)*, vol. 30, no. 4, pp. 1–14, 2011.

[37] M. Narwaria, M. P. Da Silva, and P. Le Callet, "Hdr-vqm: An objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.

[38] A. Grimaldi, D. Kane, and M. Bertalmío, "Statistics of natural images as a function of dynamic range," *Journal of vision*, vol. 19, no. 2, pp. 13–13, 2019.

[39] A. Rana, G. Valenzise, and F. Dufaux, "Evaluation of feature detection in hdr based imaging under changes in illumination conditions," in *2015 IEEE International Symposium on Multimedia (ISM)*, pp. 289–294, IEEE, 2015.

[40] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced high dynamic range imaging*. CRC press, 2017.

[41] H. Zhao, B. Shi, C. Fernandez-Cull, S.-K. Yeung, and R. Raskar, "Unbounded high dynamic range photography using a modulo camera," in *2015 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10, IEEE, 2015.

[42] O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, "The state of the art in hdr deghosting: A survey and evaluation," *Computer Graphics Forum*, vol. 34, no. 2, pp. 683–707, 2015.

[43] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar, "Optical splitting trees for high-precision monocular imaging," *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 32–42, 2007.

[44] A. O. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do hdr displays support ldr content? a psychophysical evaluation," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, pp. 38–es, 2007.

[45] B. Masia, S. Agustin, R. W. Fleming, O. Sorkine, and D. Gutierrez, "Evaluation of reverse tone mapping through varying exposure conditions," in *ACM SIGGRAPH Asia 2009 papers*, pp. 1–8, 2009.

[46] L. Wang, L.-Y. Wei, K. Zhou, B. Guo, and H.-Y. Shum, "High dynamic range image hallucination.," in *Rendering Techniques*, pp. 321–326, 2007.

[47] Y. Huo, F. Yang, L. Dong, and V. Brost, "Physiological inverse tone mapping based on retina response," *The Visual Computer*, vol. 30, no. 5, pp. 507–517, 2014.

[48] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "Hdr image reconstruction from a single exposure using deep cnns," *ACM transactions on graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.

[49] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1375–1385, 2020.

[50] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, “Large-scale crowdsourced study for tone-mapped hdr pictures,” *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4725–4740, 2017.

[51] L. Krasula, M. Narwaria, K. Fliegel, and P. Le Callet, “Preference of experience in image tone-mapping: Dataset and framework for objective measures comparison,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 64–74, 2016.

[52] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, “Photographic tone reproduction for digital images,” in *ACM Transactions on Graphics (TOG)*, vol. 21, pp. 267–276, ACM, 2002.

[53] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, “Adaptive logarithmic mapping for displaying high contrast scenes,” in *Computer Graphics Forum*, vol. 22, pp. 419–426, Wiley Online Library, 2003.

[54] S. N. Pattanaik, J. Tumblin, H. Yee, 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*, pp. 47–54, ACM Press/Addison-Wesley Publishing Co., 2000.

[55] E. Reinhard and K. Devlin, “Dynamic range reduction inspired by photoreceptor physiology,” *IEEE transactions on visualization and computer graphics*, vol. 11, no. 1, pp. 13–24, 2005.

[56] S. Ferradans, M. Bertalmio, E. Provenzi, and V. Caselles, “An analysis of visual adaptation and contrast perception for tone mapping,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 2002–2012, 2011.

[57] R. Mantiuk, S. Daly, and L. Kerofsky, “Display adaptive tone mapping,” in *ACM SIGGRAPH 2008 papers*, pp. 1–10, 2008.

[58] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, “Optimizing a tone curve for backward-compatible high dynamic range image and video compression,” *IEEE transactions on image processing*, vol. 20, no. 6, pp. 1558–1571, 2010.

[59] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM transactions on graphics (TOG)*, vol. 21, pp. 257–266, ACM, 2002.

[60] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *ACM transactions on graphics (TOG)*, vol. 21, pp. 249–256, ACM, 2002.

[61] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Transactions on Applied Perception (TAP)*, vol. 3, no. 3, pp. 286–308, 2006.

[62] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *international Conference on computer vision & Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, IEEE Computer Society, 2005.

[63] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[64] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[65] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[67] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[68] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

[69] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[70] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 157–166, ACM, 2014.

[71] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep image retrieval: Learning global representations for image search," in *European conference on computer vision*, pp. 241–257, Springer, 2016.

[72] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, "Large-scale image retrieval with attentive deep local features," in *Proceedings of the IEEE international conference on computer vision*, pp. 3456–3465, 2017.

[73] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning cnn image retrieval with no human annotation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1655–1668, 2018.

[74] A. Frome, Y. Singer, and J. Malik, "Image retrieval and classification using local distance functions," in *Advances in neural information processing systems*, pp. 417–424, 2007.

[75] B. McFee and G. R. Lanckriet, "Metric learning to rank," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 775–782, 2010.

[76] R.-Z. Liang, L. Shi, H. Wang, J. Meng, J. J.-Y. Wang, Q. Sun, and Y. Gu, "Optimizing top precision performance measure of content-based image retrieval by learning similarity function," in *2016 23rd International conference on pattern recognition (ICPR)*, pp. 2954–2958, IEEE, 2016.

[77] G. Chechik, V. Sharma, U. Shalit, and S. Bengio, "Large scale online learning of image similarity through ranking," *Journal of Machine Learning Research*, vol. 11, no. Mar, pp. 1109–1135, 2010.

[78] P. O. Pinheiro, "Unsupervised domain adaptation with similarity learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8004–8013, 2018.

[79] S. Chopra, R. Hadsell, Y. LeCun, *et al.*, "Learning a similarity metric discriminatively, with application to face verification," in *CVPR (1)*, pp. 539–546, 2005.

[80] S. Bell and K. Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 98, 2015.

[81] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, "Learning fine-grained image similarity with deep ranking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1386–1393, 2014.

[82] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5297–5307, 2016.

[83] N. Garcia and G. Vogiatzis, "Learning non-metric visual similarity for image retrieval," *Image and Vision Computing*, 2019.

[84] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, pp. 3320–3328, 2014.

[85] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[86] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *International conference on machine learning*, pp. 647–655, 2014.

[87] Z. Lun, E. Kalogerakis, and A. Sheffer, “Elements of style: learning perceptual shape style similarity,” *ACM Transactions on Graphics (TOG)*, vol. 34, no. 4, p. 84, 2015.

[88] B. Saleh, M. Dontcheva, A. Hertzmann, and Z. Liu, “Learning style similarity for searching infographics,” in *Proceedings of the 41st graphics interface conference*, pp. 59–64, Canadian Information Processing Society, 2015.

[89] B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. B. Kalin, “Perceptual image similarity experiments,” in *Photonics West’98 Electronic Imaging*, pp. 576–590, International Society for Optics and Photonics, 1998.

[90] T. Frese, C. A. Bouman, and J. P. Allebach, “Methodology for designing image similarity metrics based on human visual system models,” in *Human Vision and Electronic Imaging II*, vol. 3016, pp. 472–483, International Society for Optics and Photonics, 1997.

[91] D. Neumann and K. R. Gegenfurtner, “Image retrieval and perceptual similarity,” *ACM Transactions on Applied Perception (TAP)*, vol. 3, no. 1, pp. 31–47, 2006.

[92] E. ISO, “11664-4 colorimetry—part 4: Cie 1976 l* a* b* colour space,” *CEN (European Committee for Standardization): Brussels, Belgium*, 2011.

[93] M. Sharma and H. Ghosh, “Histogram of gradient magnitudes: a rotation invariant texture-descriptor,” in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 4614–4618, IEEE, 2015.

[94] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, “Scene parsing through ade20k dataset,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 633–641, 2017.

[95] A. Bhattacharyya, “On a measure of divergence between two multinomial populations,” *Sankhyā: The Indian Journal of Statistics*, pp. 401–406, 1946.

[96] Y. Rubner, C. Tomasi, and L. J. Guibas, “The earth mover’s distance as a metric for image retrieval,” *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.

[97] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, “Visual attention in ldr and hdr images,” in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015.

[98] M. Klíma, K. Fliegel, P. Pata, S. Vitek, M. Blažek, P. Dostal, L. Krasula, T. Kratochvíl, V. Rícnỳ, M. Slanina, *et al.*, “Deimos–an open source image database.,” *Radioengineering*, vol. 20, no. 4, 2011.

[99] “Empa hdr image database.” `http://www.empamedia.ethz.ch/hdrdatabase/`. Accessed: 2017-08-26.

[100] R. Mantiuk, “High dynamic range imaging: towards the limits of the human visual perception,” *Forsch. Wiss. Rechnen*, vol. 72, pp. 11–27, 2007.

[101] R. Mantiuk and W. Heidrich, “Visualizing high dynamic range images in a web browser,” *Journal of Graphics, GPU, and Game Tools*, vol. 14, no. 1, pp. 43–53, 2009.

[102] A. Kovashka, O. Russakovsky, L. Fei-Fei, and K. Grauman, “Crowdsourcing in computer vision,” *Foundations and Trends in Computer Graphics and Vision*, vol. 10, pp. 177–243, Jan. 2016.

[103] H. Garcia-Molina, M. Joglekar, A. Marcus, A. Parameswaran, and V. Verroios, “Challenges in data crowdsourcing,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 4, pp. 901–911, 2016.

[104] B. Zhang and S. N. Srihari, “Properties of binary vector dissimilarity measures,” in *Proc. JCIS Int’l Conf. Computer Vision, Pattern Recognition, and Image Processing*, vol. 1, 2003.

[105] W. Ouyang, X. Wang, C. Zhang, and X. Yang, “Factors in finetuning deep model for object detection with long-tail distribution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 864–873, 2016.

[106] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, vol. 2, pp. 1735–1742, IEEE, 2006.

[107] N. N. Vo and J. Hays, "Localizing and orienting street views using overhead imagery," in *European conference on computer vision*, pp. 494–509, Springer, 2016.

[108] R. A. Fisher, "On the interpretation of χ 2 from contingency tables, and the calculation of p," *Journal of the Royal Statistical Society*, vol. 85, no. 1, pp. 87–94, 1922.

[109] C. A. Parraga, X. Otazu, *et al.*, "Which tone-mapping operator is the best? a comparative study of perceptual quality," *JOSA A*, vol. 35, no. 4, pp. 626–638, 2018.

[110] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 257–266, 2002.

[111] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 249–256, 2002.

[112] HDRsoft, "Photomatix pro," 2010.

[113] A. Adams, *The camera.* The Ansel Adams Photography series, Little, Brown and Company, 1980.

[114] A. Adams, *The negative.* The Ansel Adams Photography series, Little, Brown and Company, 1981.

[115] A. Adams, *The print.* The Ansel Adams Photography series, Little, Brown and Company, 1983.

[116] M. White, R. Zakia, and P. Lorenz, *The new zone system manual.* Morgan & Morgan, Inc., 1984.

[117] D. Lischinski, Z. Farbman, M. Uyttendaele, and R. Szeliski, "Interactive local adjustment of tonal values," in *ACM Trans. Grap.*, vol. 25, pp. 646–653, 2006.

[118] S. Paris, S. W. Hasinoff, and J. Kautz, "Local laplacian filters: Edge-aware image processing with a laplacian pyramid.," *ACM Trans. Graph.*, vol. 30, no. 4, p. 68, 2011.

[119] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, “Fast local laplacian filters: Theory and applications,” *ACM Transactions on Graphics (TOG)*, vol. 33, no. 5, pp. 1–14, 2014.

[120] R. Mantiuk and H.-P. Seidel, “Modeling a generic tone-mapping operator,” *Comp. Graph. Forum*, vol. 27, no. 2, pp. 699–708, 2008.

[121] S. B. Kang, A. Kapoor, and D. Lischinski, “Personalization of image enhancement,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1799–1806, IEEE, 2010.

[122] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, “Learning photographic global tonal adjustment with a database of input/output image pairs,” in *CVPR 2011*, pp. 97–104, IEEE, 2011.

[123] L. Yuan and J. Sun, “Automatic exposure correction of consumer photographs,” in *European Conference on Computer Vision*, pp. 771–785, Springer, 2012.

[124] M. Son, Y. Lee, H. Kang, and S. Lee, “Art-photographic detail enhancement,” in *Computer graphics forum*, vol. 33, pp. 391–400, Wiley Online Library, 2014.

[125] J. Park, J.-Y. Lee, D. Yoo, and I. So Kweon, “Distort-and-recover: Color enhancement using deep reinforcement learning,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5928–5936, 2018.

[126] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, “Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6306–6314, 2018.

[127] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool, “Wespe: weakly supervised photo enhancer for digital cameras,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 691–700, 2018.

[128] Z. Yan, H. Zhang, B. Wang, S. Paris, and Y. Yu, "Automatic photo adjustment using deep neural networks," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 2, pp. 1–15, 2016.

[129] Y. Hu, H. He, C. Xu, B. Wang, and S. Lin, "Exposure: A white-box photo post-processing framework," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 2, pp. 1–17, 2018.

[130] J.-Y. Lee, K. Sunkavalli, Z. Lin, X. Shen, and I. So Kweon, "Automatic content-aware color and tone stylization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2470–2478, 2016.

[131] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling perceptual factors in neural style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3985–3993, 2017.

[132] A. Gupta, J. Johnson, A. Alahi, and L. Fei-Fei, "Characterizing and improving stability in neural style transfer," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4067–4076, 2017.

[133] X.-C. Liu, M.-M. Cheng, Y.-K. Lai, and P. L. Rosin, "Depth-aware neural style transfer," in *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering*, pp. 1–10, 2017.

[134] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1501–1510, 2017.

[135] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4990–4998, 2017.

[136] A. O. Akyüz, K. Hadimli, M. Aydinlilar, and C. Bloch, "Style-based tone mapping for hdr images," in *SIGGRAPH Asia 2013 Technical Briefs*, p. 23, ACM, 2013.

[137] R. Mantiuk and H.-P. Seidel, "Modeling a generic tone-mapping operator," in *Computer Graphics Forum*, vol. 27, pp. 699–708, Wiley Online Library, 2008.

[138] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proc. of the 1998 IEEE International Conference on Computer Vision*, pp. 839–846, IEEE, 1998.

[139] S. Bae, S. Paris, and F. Durand, “Two-scale tone management for photographic look,” in *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 637–645, ACM, 2006.

[140] M. Trentacoste, R. Mantiuk, W. Heidrich, and F. Dufrot, “Unsharp masking, countershading and halos: Enhancements or artifacts?,” *Comp. Graph. Forum*, vol. 31, no. 2pt3, pp. 555–564, 2012.

[141] D. Shepard, “A two-dimensional interpolation function for irregularly-spaced data,” in *Proceedings of the 1968 23rd ACM national conference*, pp. 517–524, ACM, 1968.

[142] N. Ben-Haim, B. Babenko, and S. Belongie, “Improving web-based image search via content based clustering,” in *CVPR*, pp. 106–106, IEEE, 2006.

[143] D. Amirkhani and A. Bastanfard, “Inpainted image quality evaluation based on saliency map features,” in *2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, pp. 1–6, IEEE, 2019.

[144] G. Eilertsen, R. K. Mantiuk, and J. Unger, “A comparative review of tone-mapping algorithms for high dynamic range video,” in *Computer Graphics Forum*, vol. 36, pp. 565–592, Wiley Online Library, 2017.

APPENDIX A

CLUSTERS OF THE HDR IMAGE DATASET

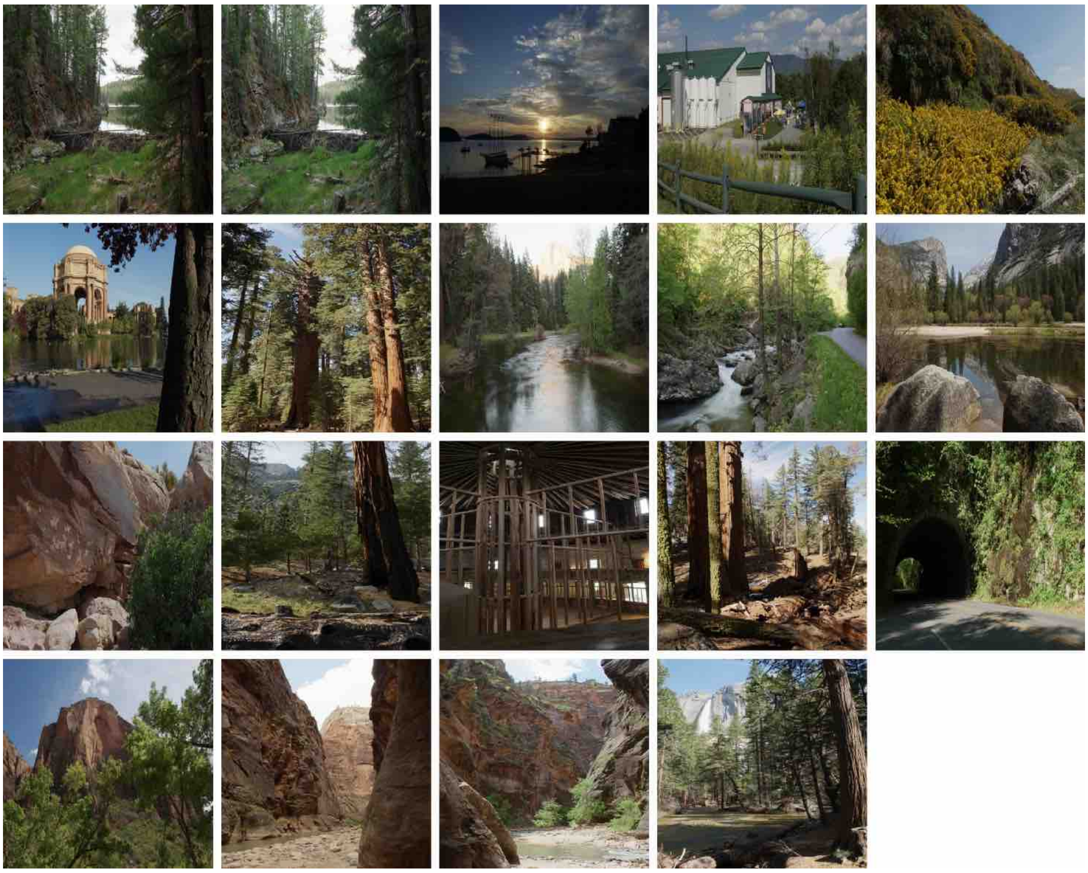


Figure A.1: Cluster I, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

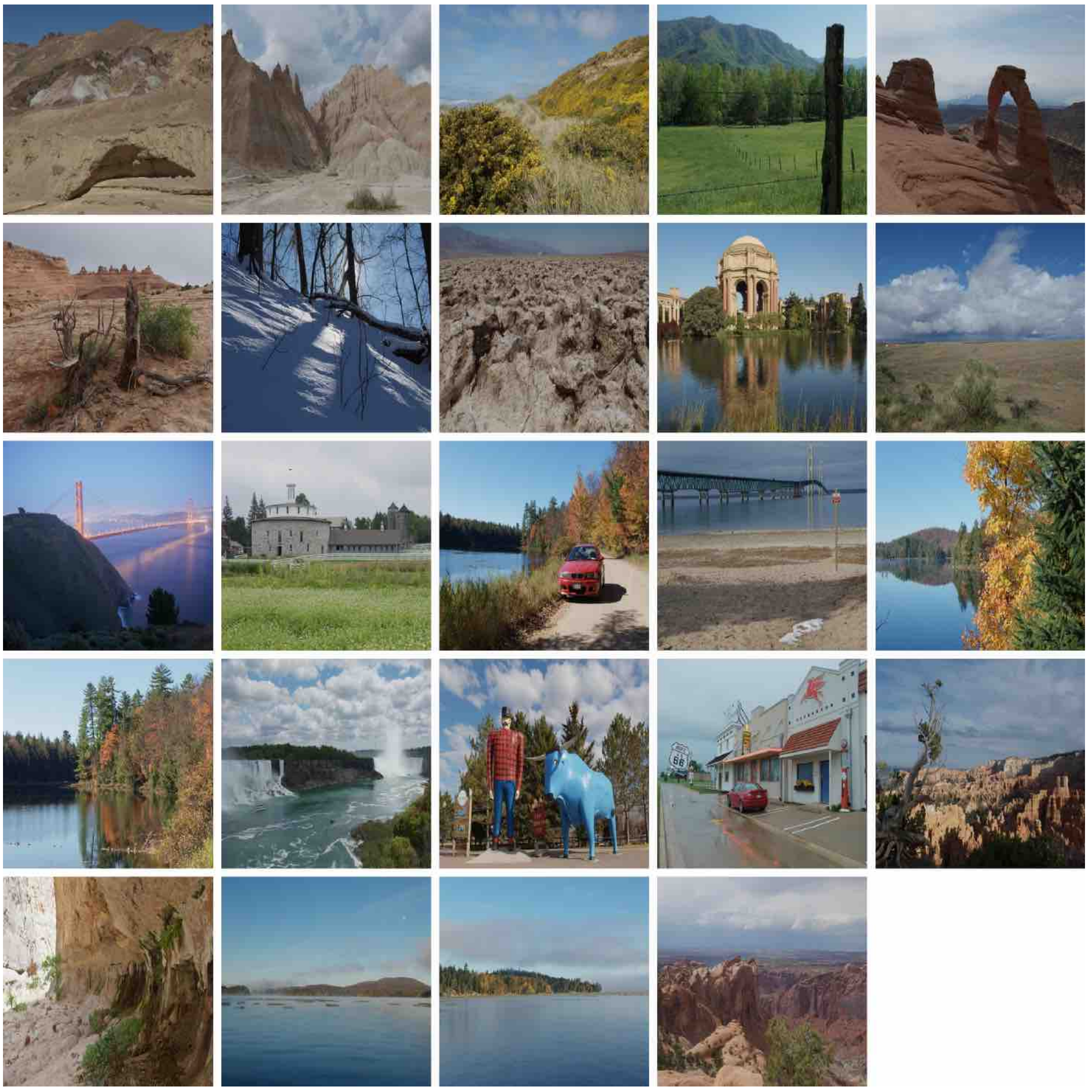


Figure A.2: Cluster II, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

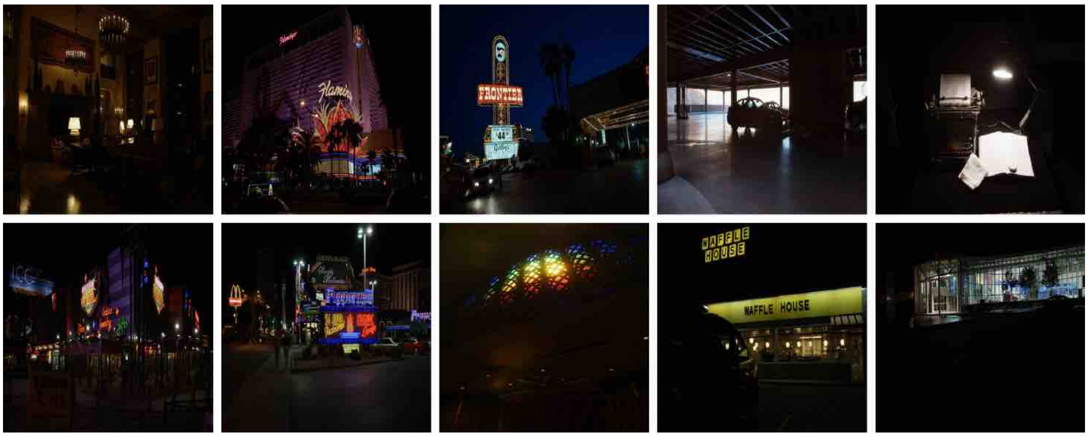


Figure A.3: Cluster III, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

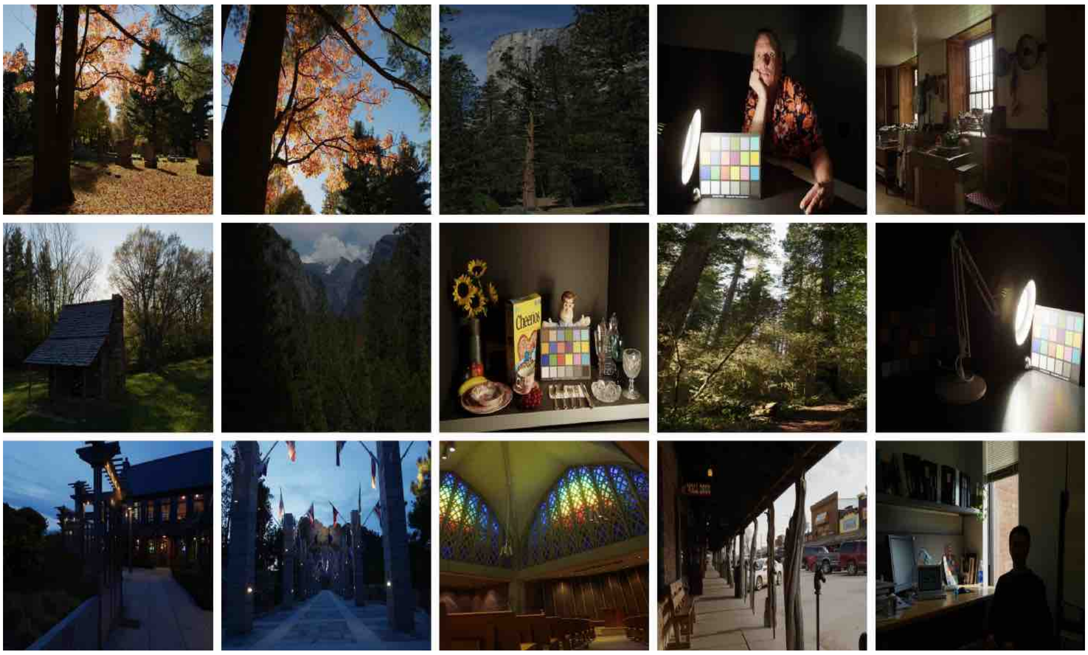


Figure A.4: Cluster IV, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

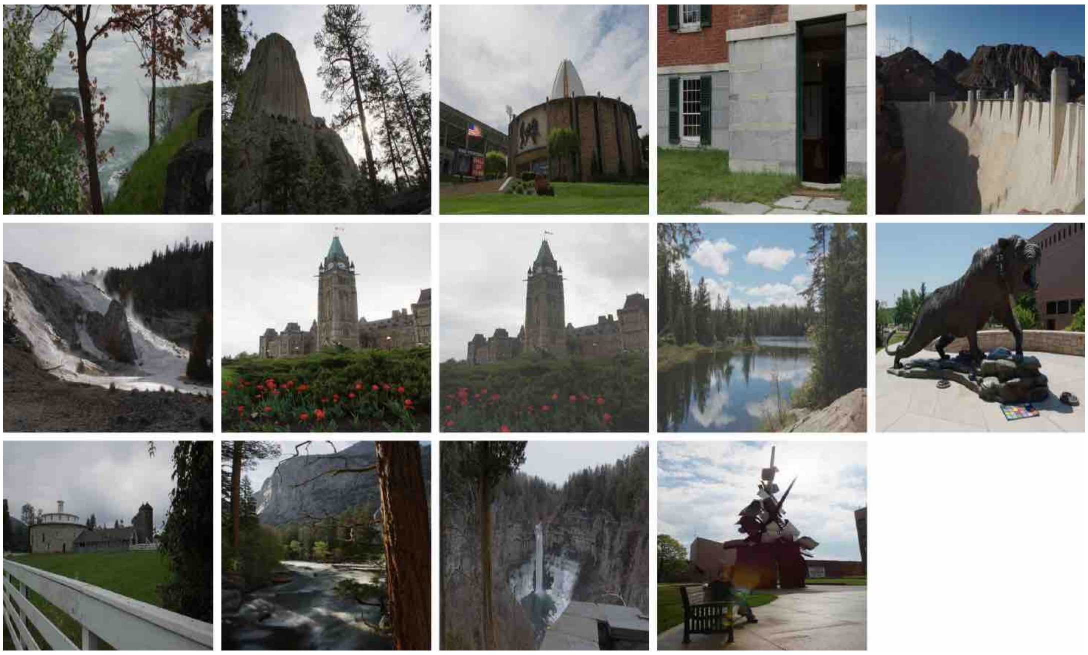


Figure A.5: Cluster V, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

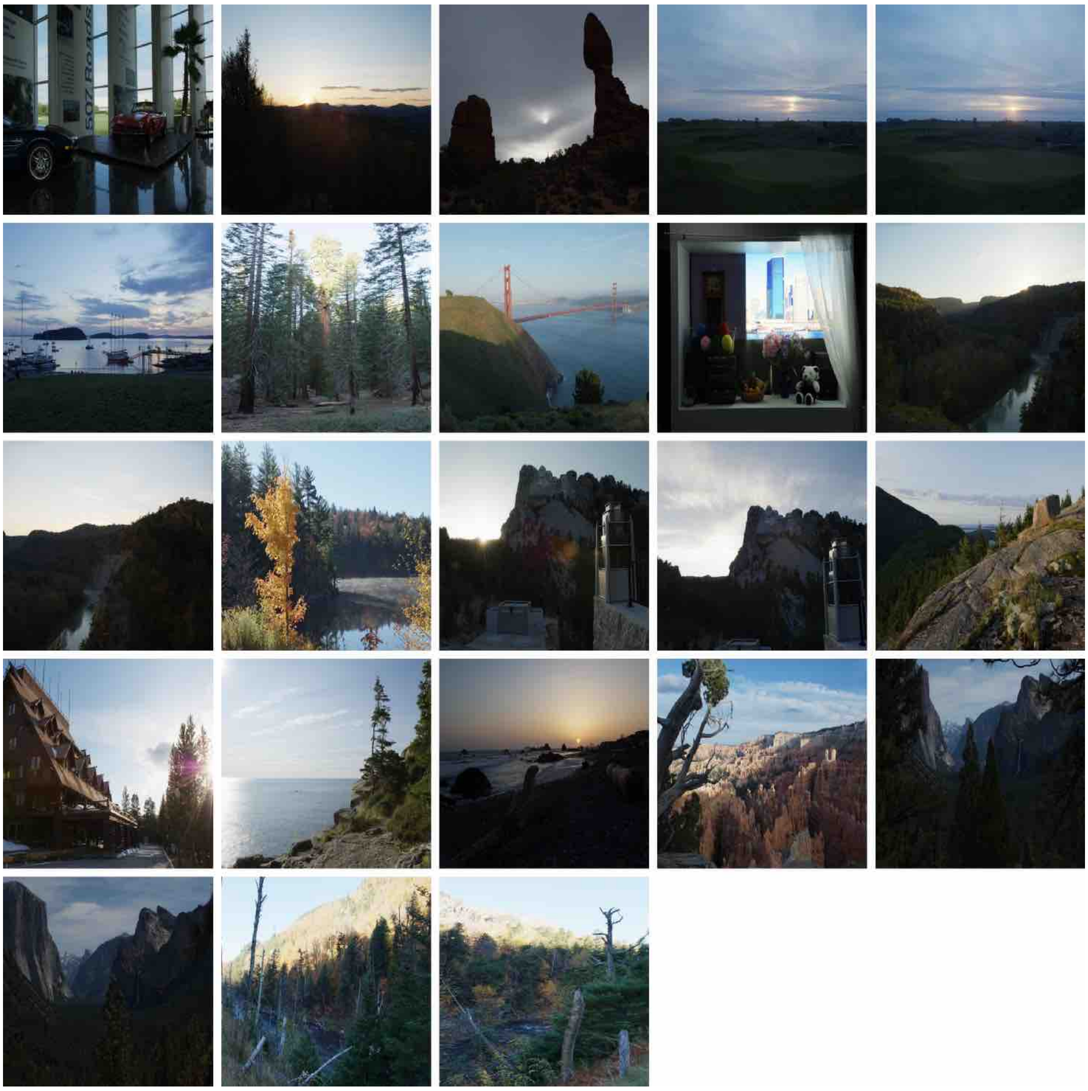


Figure A.6: Cluster VI, clustering result of Fairchild's HDR Dataset [4] for calibration image selection.

APPENDIX B

MANUALLY TONE MAPPED CALIBRATION IMAGES

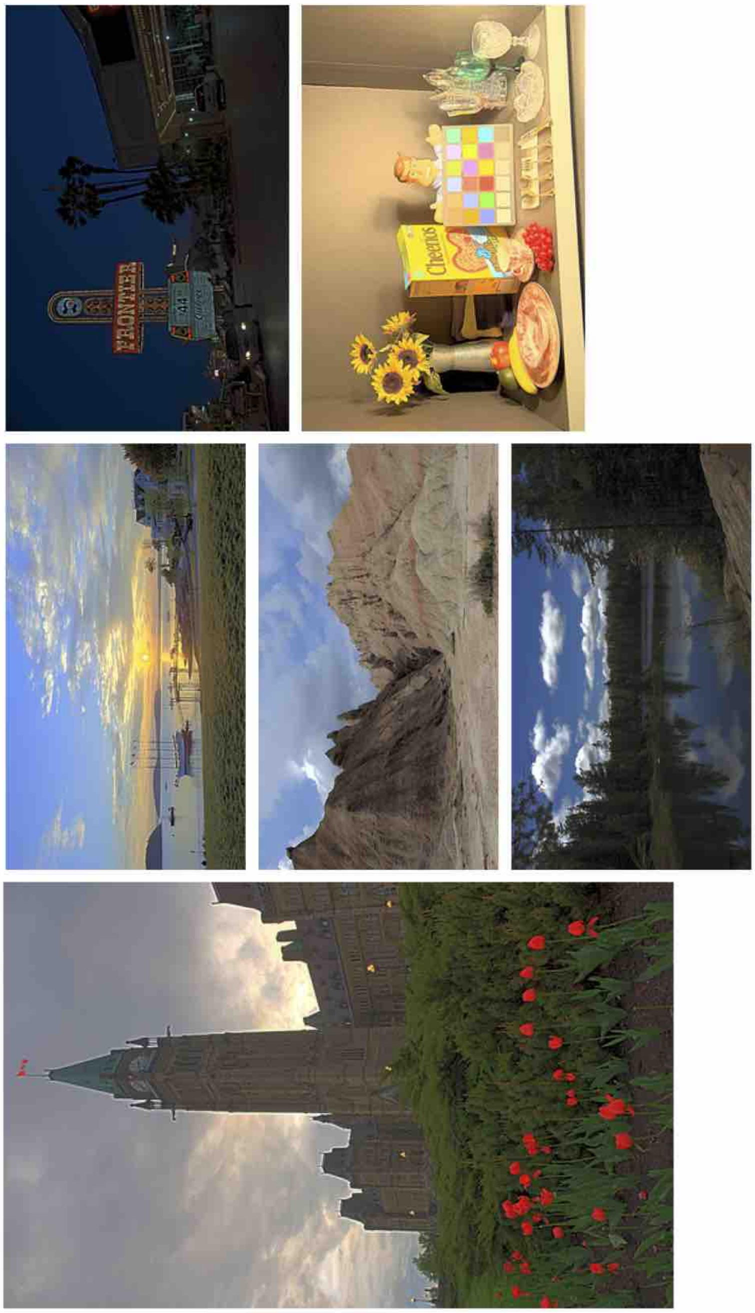


Figure B.1: Manually tone mapped calibration images for *candy* style.

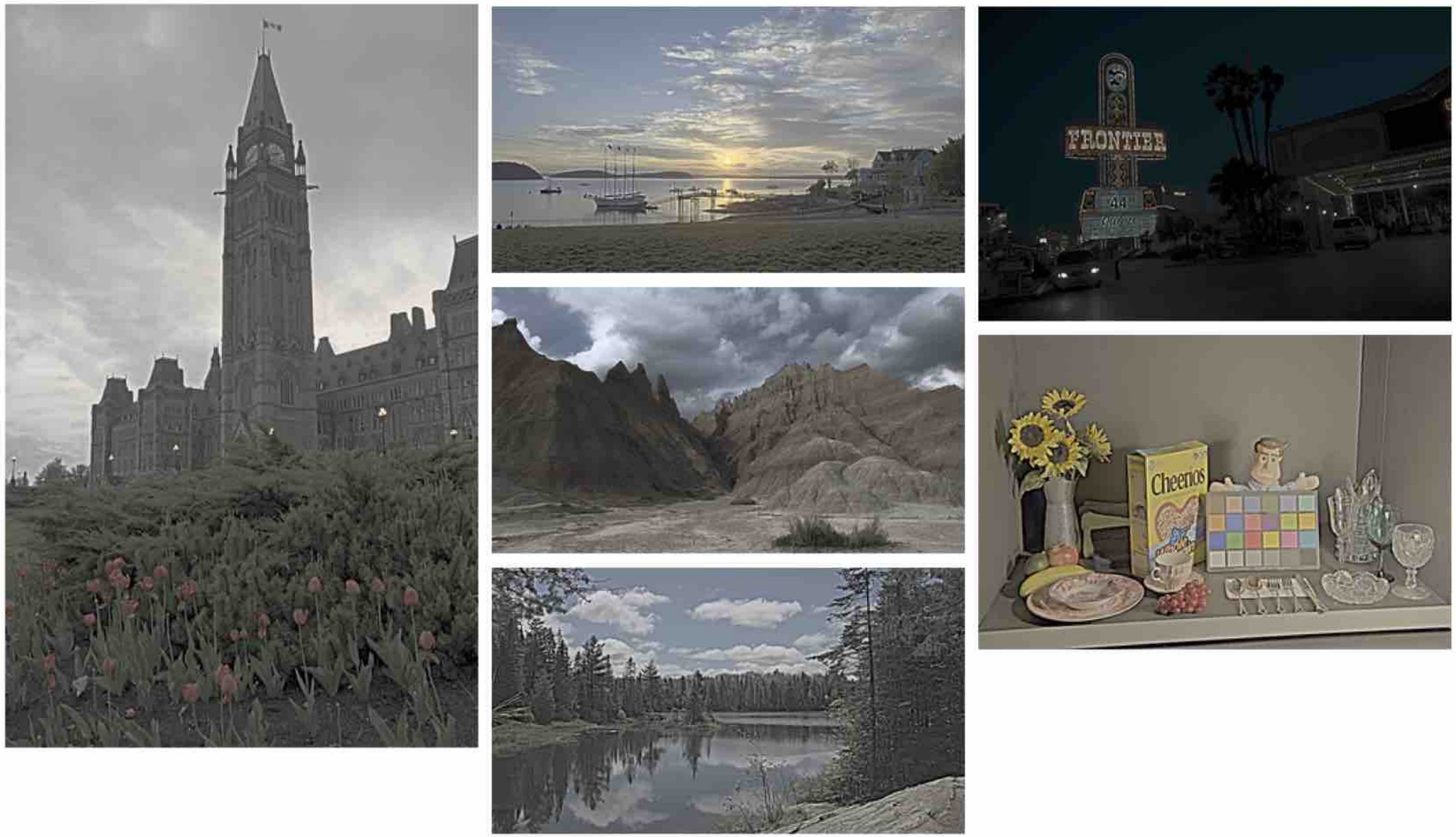


Figure B.2: Manually tone mapped calibration images for *grittily* style.

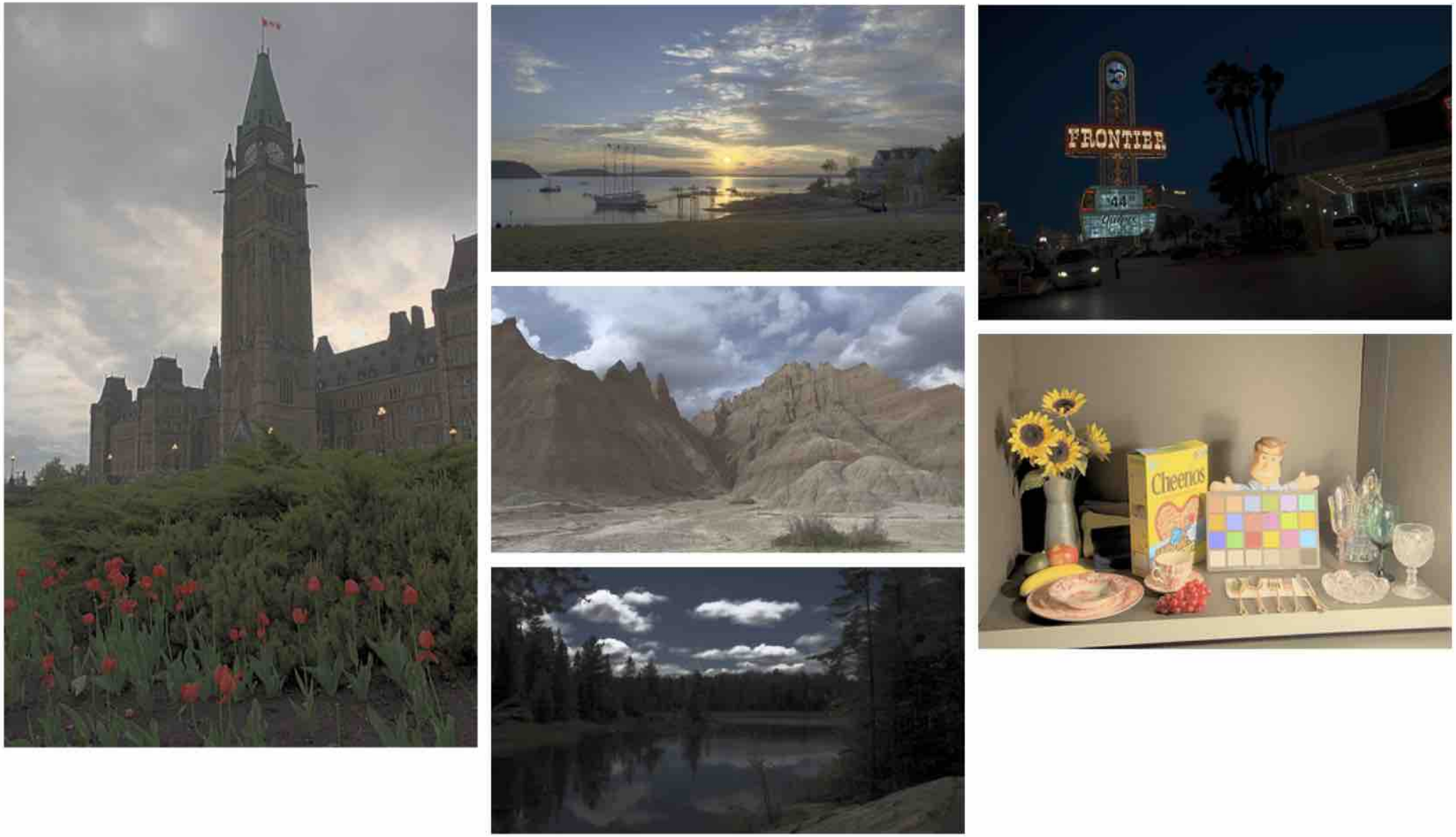


Figure B.3: Manually tone mapped calibration images for *natural* style.

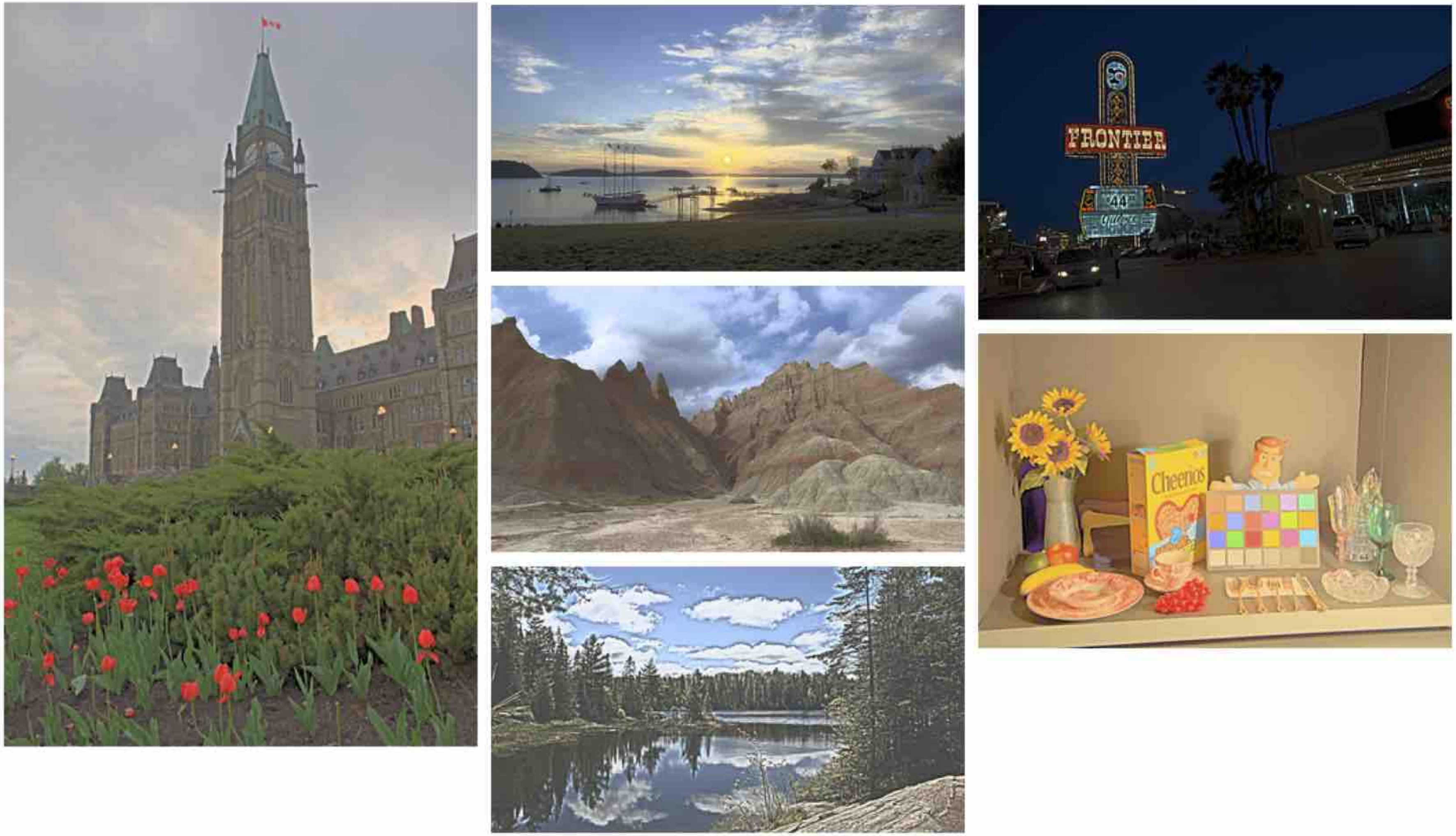


Figure B.4: Manually tone mapped calibration images for *painterly* style.

APPENDIX C

RESULTS OF STYLE BASED TONE MAPPING

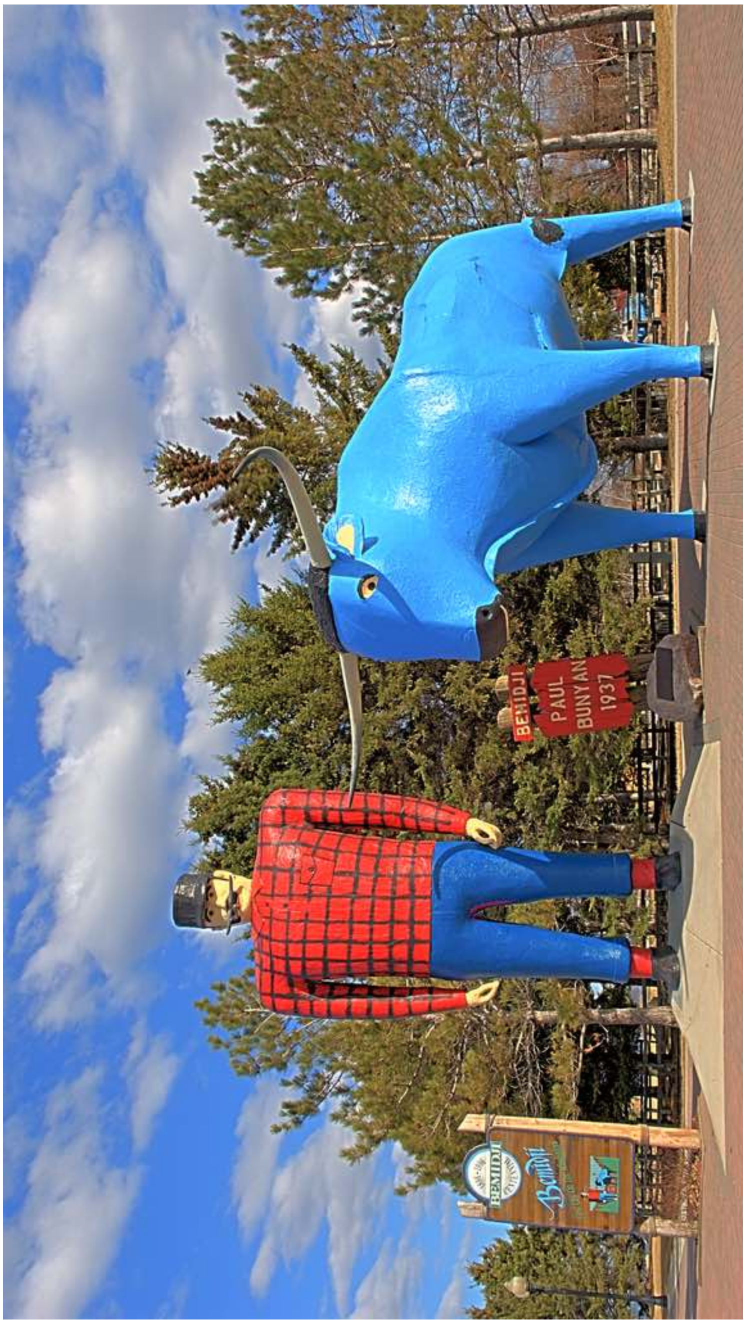


Figure C.1: High resolution version of Figure 5.8(a).

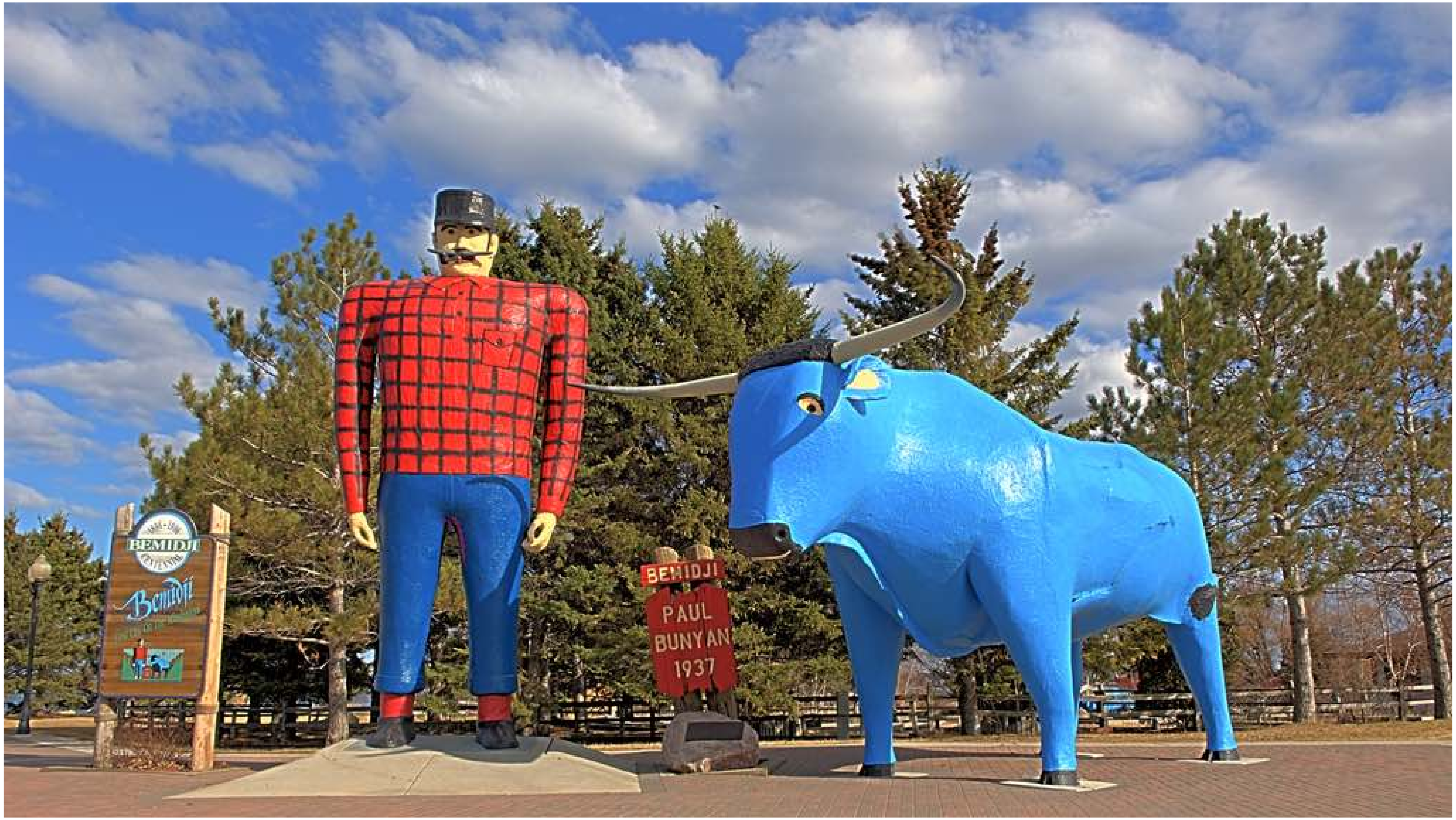


Figure C.2: High resolution version of Figure 5.8(b).

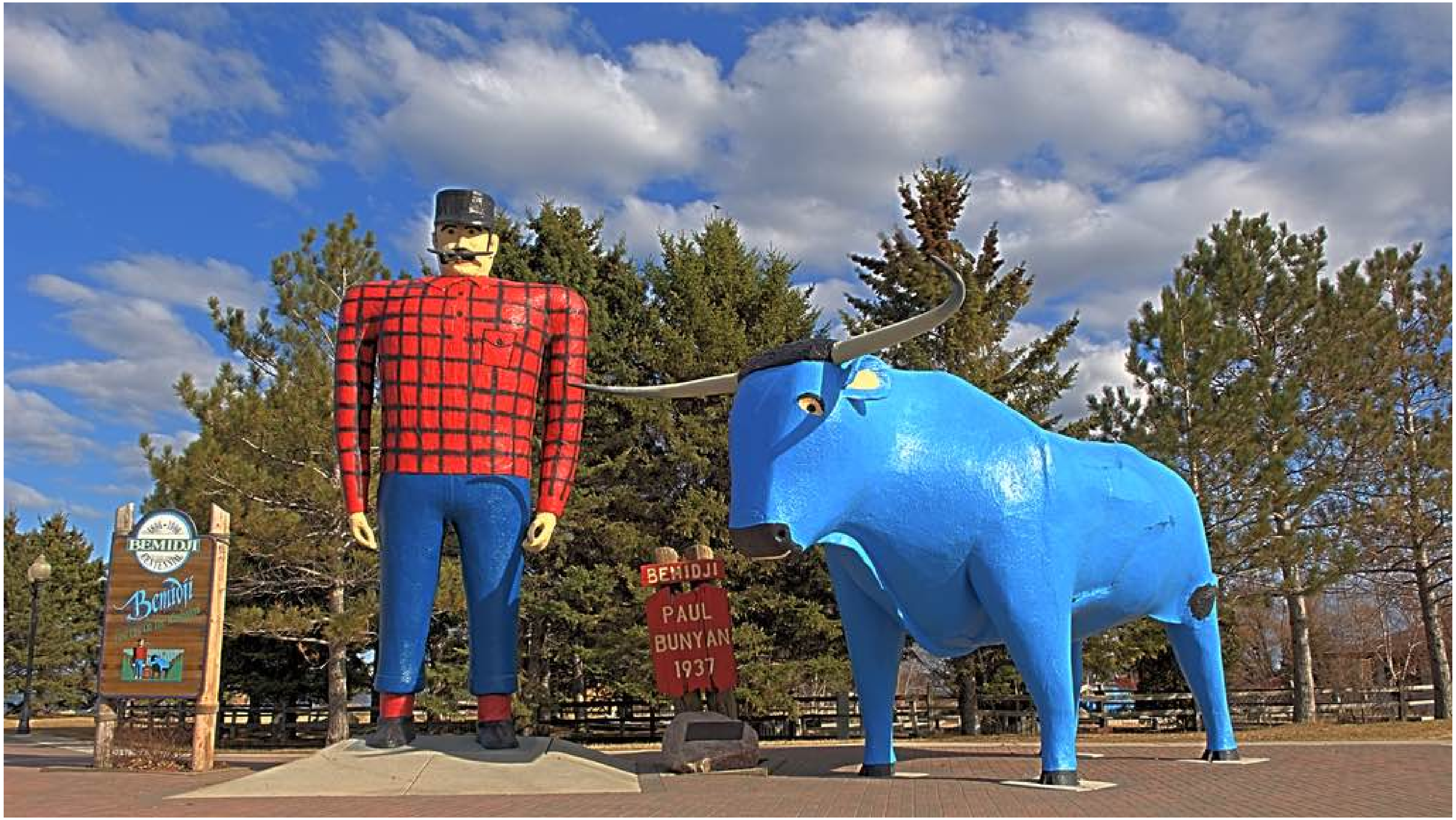


Figure C.3: High resolution version of Figure 5.8(c).

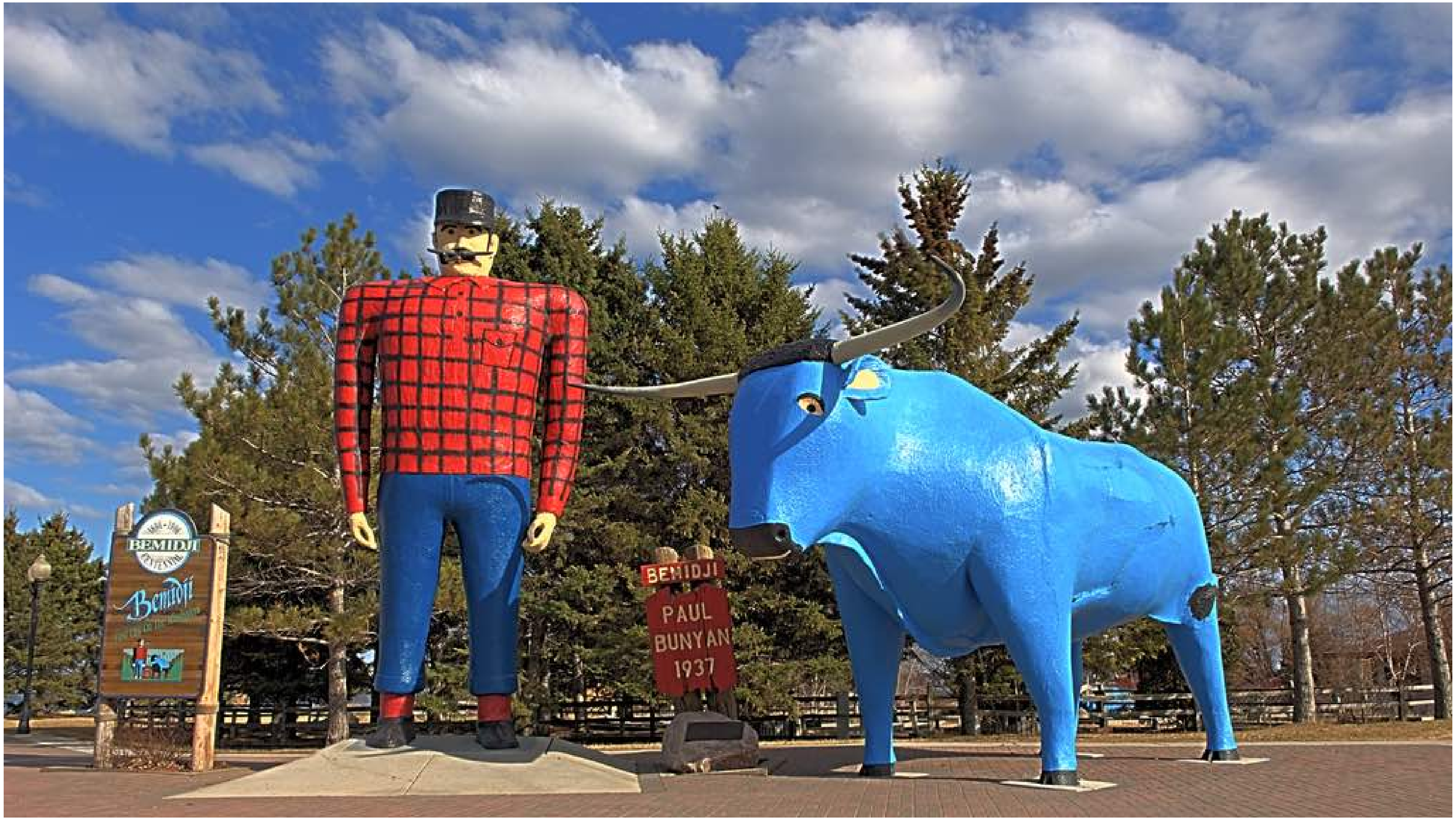


Figure C.4: High resolution version of Figure 5.8(d).

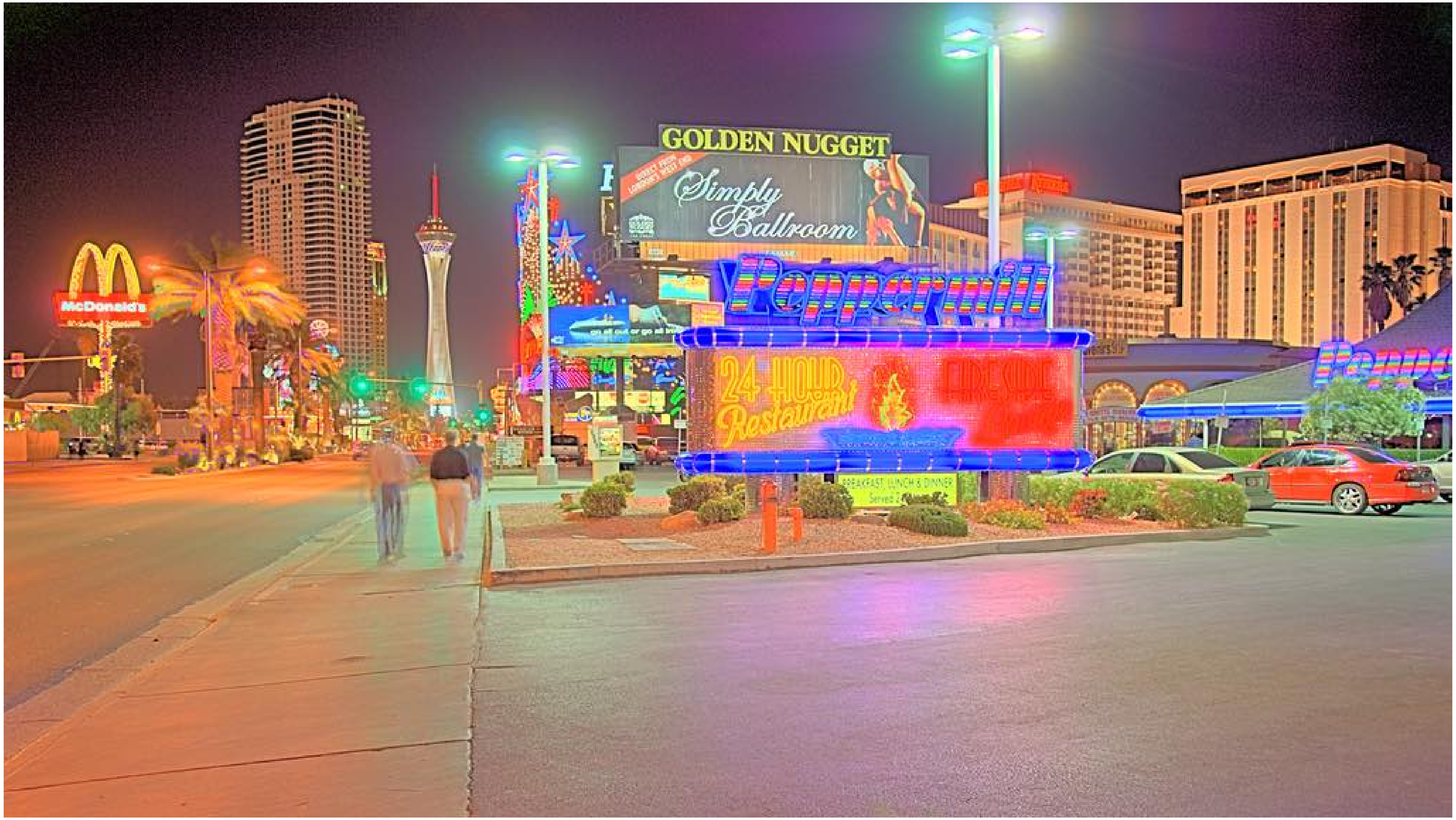


Figure C.5: High resolution version of Figure 5.8(e).

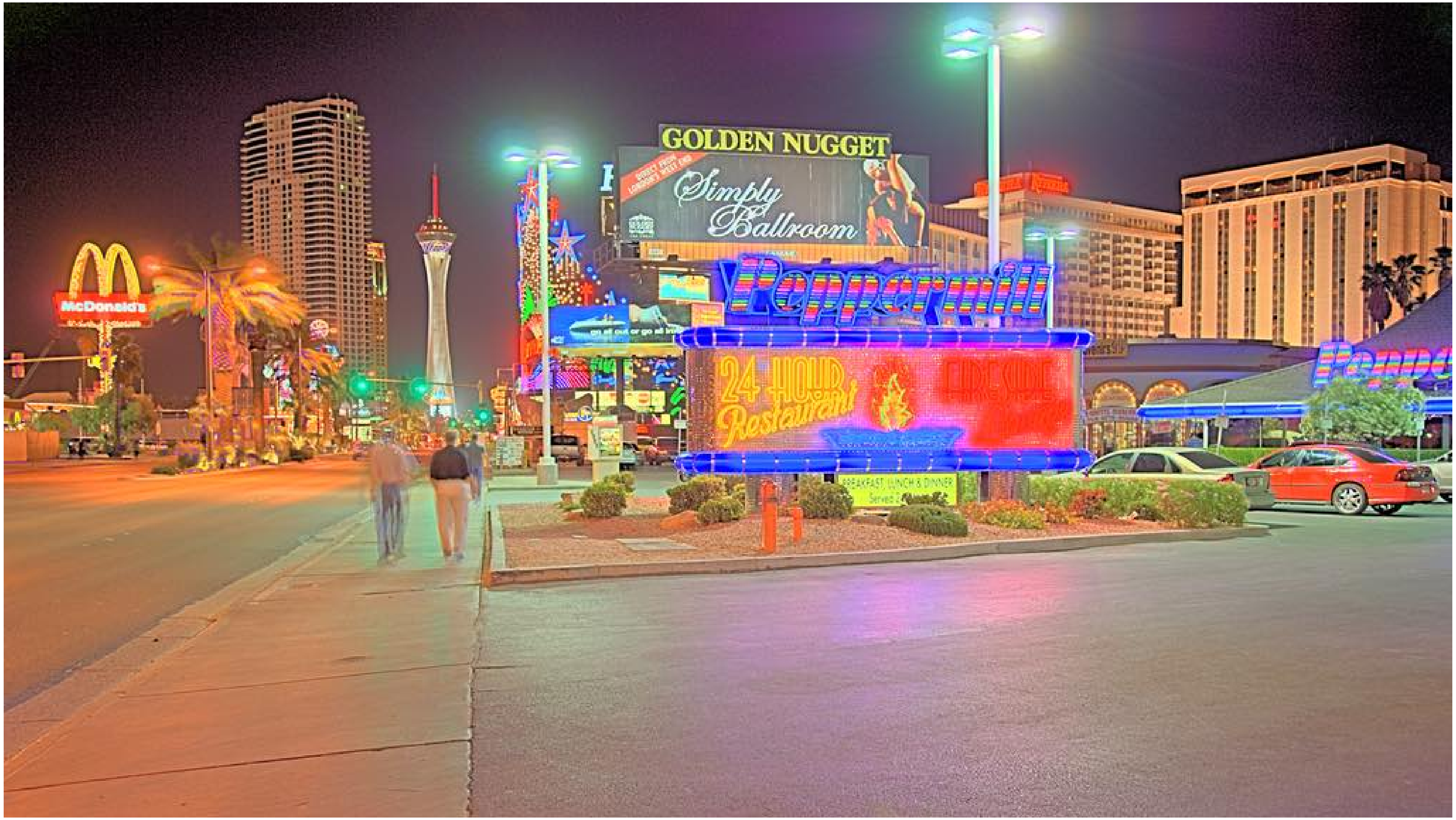


Figure C.6: High resolution version of Figure 5.8(f).

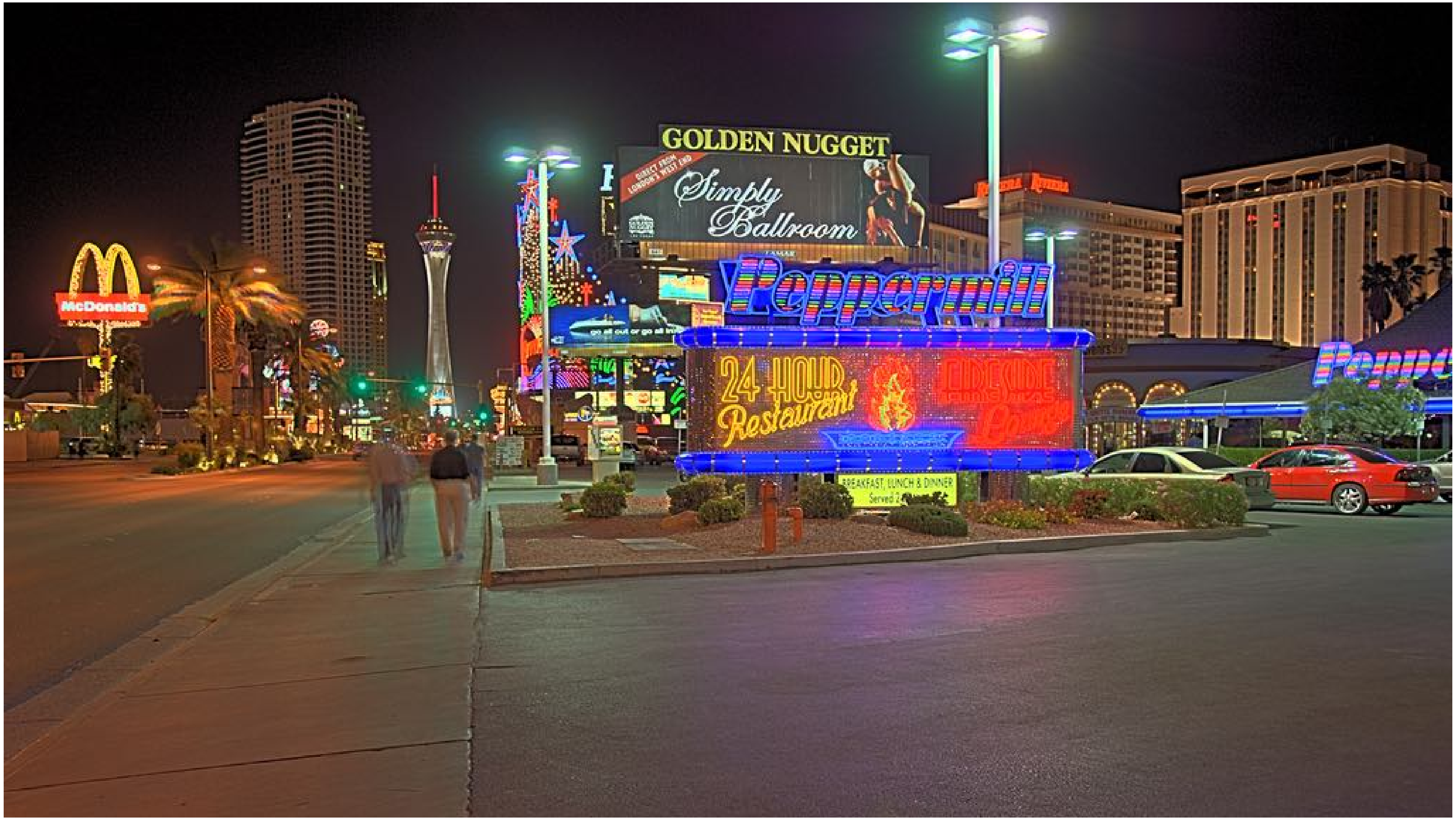


Figure C.7: High resolution version of Figure 5.8(g).

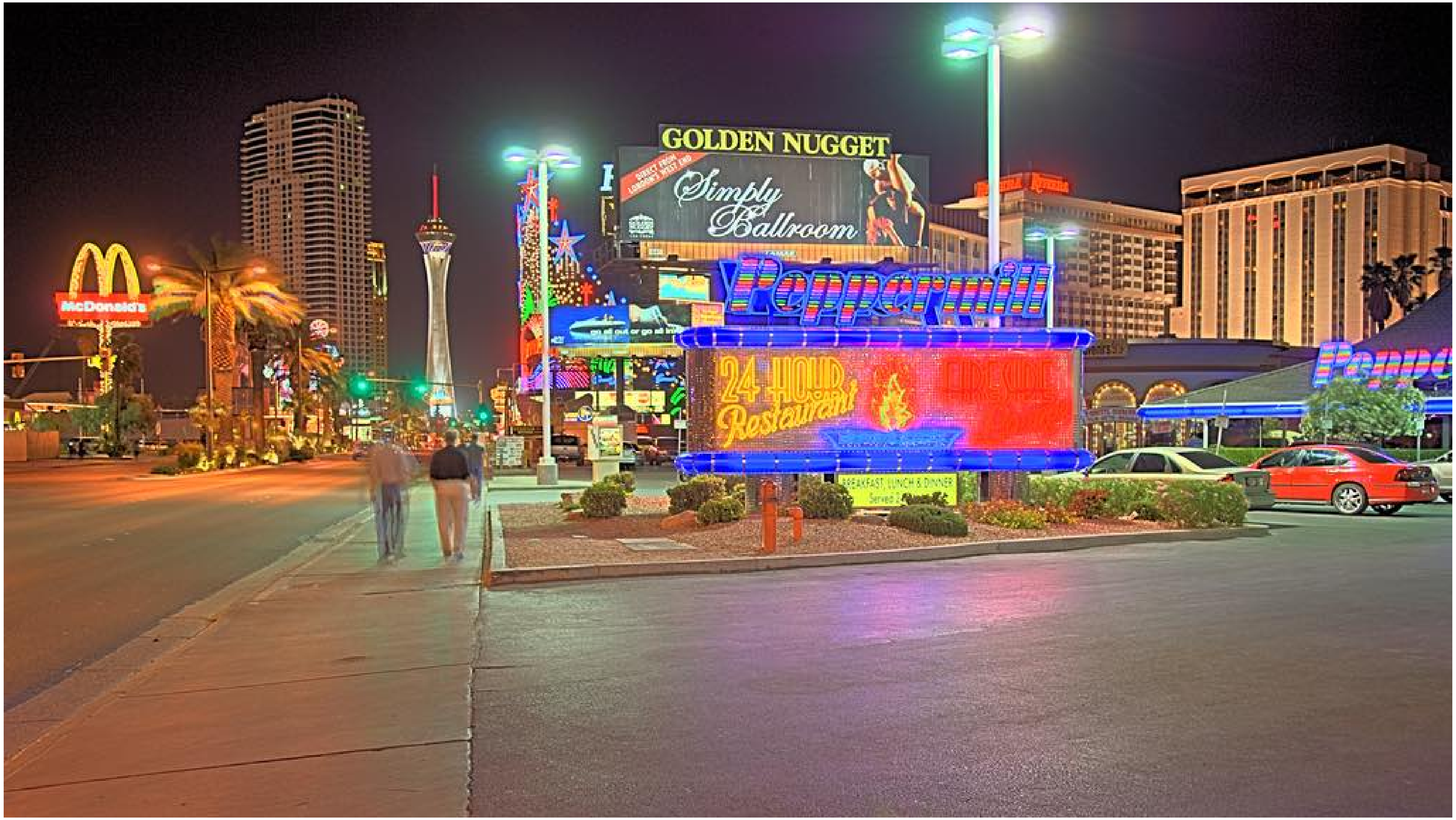


Figure C.8: High resolution version of Figure 5.8(h).

CURRICULUM VITAE

EDUCATION

Degree	Institution	Year of Graduation
M.Sc.	Dept. of Computer Engineering, METU	2011
B.S.	Dept. of Computer Engineering, METU	2008

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
2018 - Present	New Work SE, Hamburg	Data Scientist
2016 - 2018	Bytro Labs GmbH, Hamburg	Backend Developer
2010 - 2016	Dept. of Computer Engineering, METU	System Administrator

PUBLICATIONS

1. Ahmet Oguz Akyüz, Kerem Hadimli, Merve Aydinlilar, and Christian Bloch. Style-based tone mapping for HDR images. SIGGRAPH Asia 2013 Technical Briefs, SA 2013.

2. Mehmet Akif Akkus, Merve Aydinlilar, and Sinan Kalkan. Range analysis of junctions. Signal Processing and Communications Applications Conference, SIU 2013.

3. Merve Aydinlilar, Adnan Yazici. Semi-automatic semantic video annotation tool. Computer and Information Sciences III - 27th International Symposium on Computer and Information Sciences, ISCIS 2012.