Algorithmic Rectification of Visual Illegibility under Extreme Lighting

ALGORITHMIC RECTIFICATION OF VISUAL ILLEGIBILITY UNDER EXTREME LIGHTING

BY

ZHENHAO LI, B.Eng.

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2018) McMaster University

(Electrical & Computer Engineering) Hamilton, Ontario, Canada

TITLE:	Algorithmic Rectification of Visual Illegibility under Extreme Lighting
AUTHOR:	Zhenhao Li B.Eng., (Electrical Engineering) University of Science and Technology of China, Hefei, China
SUPERVISOR:	Dr. Xiaolin Wu
NUMBER OF PAGES:	xvii, 128

Abstract

Image and video enhancement, a classical problem of signal processing, has remained a very active research topic for past decades. This technical subject will not become obsolete even as the sensitivity and quality of modern image sensors steadily improve. No matter what level of sophistication cameras reach, there will always be more extreme and complex lighting conditions, in which the acquired images are improperly exposed and thus need to be enhanced.

The central theme of enhancement is to algorithmically compensate for sensor limitations under ill lighting and make illegible details conspicuous, while maintaining a degree of naturalness. In retrospect, all existing contrast enhancement methods focus on heightening of spatial details in the luminance channel to fulfil the goal, with no or little consideration of the colour fidelity of the processed images; as a result they can introduce highly noticeable distortions in chrominance. This long-time much overlooked problem is addressed and systematically investigated by the thesis. We then propose a novel optimization-based enhancement algorithm, generating optimal tone mapping that not only makes maximal gain of contrast but also constrains tone and chrominance distortion, achieving superior output perceptual quality against severe underexposure and/or overexposure.

Besides, we present a novel solution to restore images captured under more challenging backlit scenes, by combining the above enhancement method and feature-driven, machine learning based segmentation. We demonstrate the superior performance of the proposed method in terms of segmentation accuracy and restoration results over state-of-the-art methods.

We also shed light on a common yet largely untreated video restoration problem called Yin-Yang Phasing (YYP), featured by involuntary, intense fluctuation in intensity and chrominance of an object as the video plays. We propose a novel video restoration technique to suppress YYP artifacts while retaining temporal consistency of objects appearance via inter-frame, spatially-adaptive optimal tone mapping. Experimental results are encouraging, pointing to an effective and practical solution to the problem.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Xiaolin Wu for his guidance, encouragement and great patience. His wide knowledge in image and video processing technologies and sharp vision of future trends formed the basis of this thesis.

I would like to thank Dr. Xun Li and Dr. Jun Chen for their help and time that they put on annual committee meetings and this thesis. I would also like to thank Dr. Xiao-Ping Zhang for being my external examiner.

My gratitude also goes to my best friends and fellow collaborators Junfeng, Xuesong, Yuheng, Feng Zhou, Xiao Shu, Xiaowei, Kaixuan, Xi Zhang, Reza, Bolin, Fangzhou, Zhixiang, Qianli, Zhouyang, Kuan, Huihui and Anna Wong for generously sharing their advice and knowledge. Their help and friendship have made my graduate study an unforgettable experience. I would also like to thank Kai Cheng, Jun Du and Shenglei Li from Shanghai Jiao Tong University who kindly provided helpful experimental resources for the research.

Last but by no means least, I would like to thank my parents for their unconditional love and support. To them I dedicate this thesis.

Notation and abbreviations

2-GMM Mixture of Two Gaussian Components

2D Two Dimensional

AHD Average Hue Difference

AHE Adaptive Histogram Equalization

ASD Average Saturation Difference

CFA Colour Filter Array

CLAHE Contrast Limited Adaptive Histogram Equalization

CRF Conditional Random Field

DP Dynamic Programming

DSP Digital Signal Processing

EM Expectation-Maximization

EPD Edge-Preserving Decomposition

HDR High Dynamic Range

HE Histogram Equalization

HPF High Pass Filtering

HVS Human Visual System

LOD Luminance Ordinal Distortion

LP Linear Programming

MAP Maximum A Posteriori

MSR Multi-Scale Retinex

OCTM Optimal Contrast-Tone Mapping

PDF Probability Density Function

RBF Radial Basis Function

ReLU Rectified Linear Unit

RMSE Root-Mean-Square Error

ROC Receiver Operating Characteristic

SLIC Simple Linear Iterative Clustering

SoC System on Chip

SVM Support Vector Machine

YYP Yin-Yang Phasing

Contents

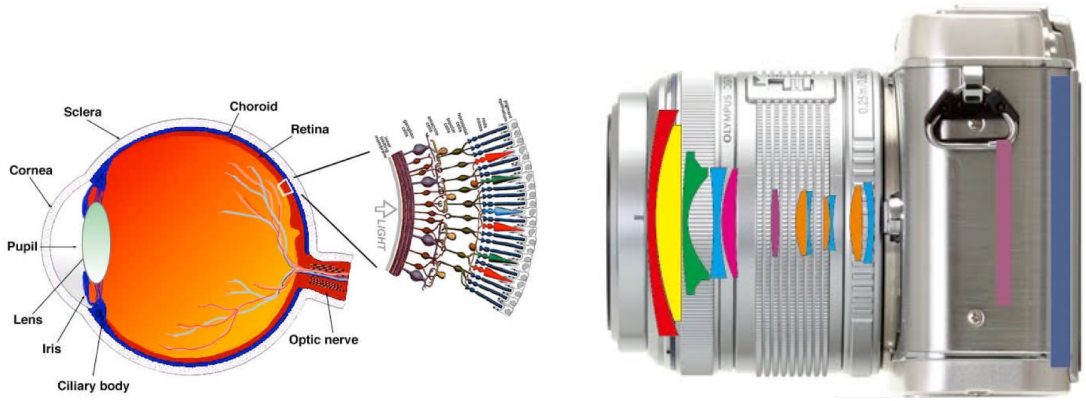
List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Photography is an invention of great significance in human history that enables one to record his or her vision. With the advances of imaging science and technology, cameras become ever more sophisticated, smaller in size and easier to use, making digital photography a ubiquitous and effective tool of communication, work and entertainment in daily lives. Bearing a strong resemblance to the human visual system (HVS) is the Daguerreotype camera, as depicted in Fig. 1.1. This classical design of cameras is the first and only one ever developed commercially with great success. Both Daguerreotype camera and HVS rely on a lens to converge the incoming lights from the physical world onto a sensor array and then transform the lights into two-dimensional images. Thanks to years of research and development, digital cameras nowadays are able to produce photos with almost no discernible differences from human vision. They can even surpass HVS in some aspects of image quality, including higher resolutions in spatial, spectral and value domains.



(a) (b)

Figure 1.1: Structure of human visual system (HVS) (a, [40]) and a digital camera (b).

However, in one perception attribute, modern image sensors still cannot match the power and versatility of HVS, that is, high dynamic range vision that can accommodate and adjust to extreme and/or complex lighting conditions. Fig. 1.2 exhibits some frequently encountered typical problems of poor quality images captured by modern cameras operating under ill illuminations, including underexposure, overexposure, uneven exposure, low contrast, etc. These problems may aggravate when producing videos under poor lighting. If the objects and/or cameras move in an environment of uneven illuminations, the acquired images may drift between underexposure and overexposure, back and forth over frames, causing severe temporal artifacts as well. Unlike cameras, human eyes can adapt to drastic changes of light strength in space and time, and function properly in a very high dynamic range of energy level of visual stimuli. This is why we can perceive objects and understand the scenes much better than cameras in poor and fast changing illumination conditions.

Technically there are several causes for failures of cameras to produce good, legible images, including:

(a) (b) (b)

Figure 1.2: Typical scenes of poor lighting conditions such as (a) dark scenes leading to underexposure, (b) extremely bright scenes leading to overexposure and (c) backlighting.

- Improper camera settings. In photography, cameras need to adapt to different lighting conditions by adjusting photographic parameters (such as exposure time, aperture, ISO values *etc*). The parameter setting is done either manually or automatically. However the quality of acquired images may still suffer from poor lighting conditions, no matter in manual automatic automatic mode. Backlighting is such an example with extremely varied illumination in different regions of an image; the backlit foreground surfaces are severely underlit while the frontlit background surfaces are comparatively overlit. Hence backlit and frontlit surfaces require entirely different exposures from each other. This is beyond photographers' control and defies cameras' automatic exposure settings. As a result, no photographic settings can prevent serious degradation quality in some parts, with details lost and colours shifted or washed out.

- Limitations of hardware. Modern digital cameras are known for their high spatial resolutions. However, for compact and inexpensive mass-marketed cameras high spatial resolutions necessarily shrinks the pixel size. As a result, the optical energy impinging on the physical pixel area diminishes. In addition, the use

of colour filter array (CFA) for colour reproduction further reduces the light influx. These mean, in the presence of inevitable sensor noises, low signal-to-noise ratio. The problem will be exacerbated by insufficient illumination, causing low contrast and poor legibility of captured images.

- Limitations of in-camera digital signal processing (DSP) chip. Digital cameras employ a DSP pipeline that converts raw sensor data into images by a series of algorithms, including denoising, gamma correction, colour demosaicking, white balancing, compression, *etc*[13, 58, 63]. But due to very tight constraints of the architecture of system on chip (SoC) [21] and the real-time requirements, the hardware implemented algorithms have to be streamlined and their performances are often subpoptimal.

Granted, the above problems can be alleviated by using a more sophisticated professional-grade camera, by more experienced photographers, or by using auxiliary equipment such as flash, tripods and light reflectors. But these solutions are expensive, user unfriendly, or unportable. A far more effective and flexible alternative is the off-line use of advanced image postprocessing algorithms to repair and enhance undesired poor quality images. The restoration and enhancement algorithms are typically implemented in software and executed on smartphones, computers, over even in cloud, if needed.

Image enhancement and restoration have been active research topics since the birth of image processing and computer vision fields. A large number of algorithms have been published to tackle the problem of image quality degradation due to poor lighting (see Sec. 1.2). Although the existing methods are highly effective in maintaining good image quality against less than ideal lighting conditions, their performances

are still unsatisfactory on images acquired under extremely pathological illuminations.

Pushing the envelope to reproduce legible and visually pleasing images in seemingly impossible lighting conditions is more than just an academic interest. Not only professional purists, ordinary users have also developed an insatiable appetite for good images at anytime and anywhere. In fact, good image quality defying adversary environments, for instances in the night or a dark room, is one of the main technical merits for which camera and smartphone manufacturers are competing with each other. In this Ph.D. thesis, we strive to explore and develop new image restoration and enhancement methods that can withstand ill, extreme and complex lighting conditions and achieve an image quality surpassing the current state of the art.

1.2 Related Work

Existing image enhancement methods can be classified into two categories: 1) local, or pixel context-sensitive; and 2) global or pixel context-free.

1.2.1 Context-sensitive Enhancement

In the class of context-sensitive enhancement approach, the contrast is viewed as the rate of change in intensity between neighbouring pixels. The contrast is increased by exaggerating the local waveform changes on a pixel by pixel basis. High-pass filtering (HPF) methods, such as unsharp masking and highboost filtering [24], fall into this class. Since edges and high textures are high-frequency features of an image, they will be made more prominent by the HPF methods. A common problem with HPF is that boosting frequency bands is likely to introduce ringing and halo artifacts around

edges, because HPF methods may alter the relative ranking of pixel values in a neighborhood [60]. To attenuate such artifacts, improved HPF-type methods are proposed, including edge-preserving filtering (bilateral filtering) [19, 20] and multiscale wavelet transform [12, 31].

Homomorphic filtering is another example of HPF enhancement. It is based on the illumination-reflectance image formation model $I(x, y) = l(x, y)r(x, y)$, where each pixel $I(x, y)$ is expressed as the product of the illumination $l(x, y)$ and reflectance $r(x, y)$ components [24]. As illumination signal $l(x, y)$ is predominantly low-pass and reflectance signal $r(x, y)$ is high-pass, one can enhance details by high-pass filtering of the logarithm image $\log I(x, y) = \log l(x, y) + \log r(x, y)$. Retinex algorithm and its variants [22, 23, 35, 36, 70] adopt a more precise decomposition of illumination and reflectance than homomorphic filtering; they can prevent many artifacts that plague earlier HPF methods. However, illumination-reflectance decomposition problem is a severely ill-conditioned; it cannot be solved satisfactorily in many cases. The enhancement techniques based on the illumination-reflectance model are prone to introduce a degree of unnaturalness in output images.

1.2.2 Context-free Enhancement

The context-free contrast enhancement approach, on the other hand, does not adjust the local waveform on a pixel by pixel basis. Instead, the class of context-free contrast enhancement techniques adopt a statistical approach. They manipulate the histogram of the input image to separate the gray levels of higher probability further apart from the neighbouring gray levels. In other words, the context-free techniques aim to increase the average difference between any two altered input gray levels. Compared

with its context-sensitive counterpart, the context-free approach does not suffer from the ringing artifacts because it can preserve the relative ordering of altered gray levels.

Among existing methods lying in this category, histogram equalization (HE) is one of the earliest and most commonly used. It seeks for a tone mapping that best flattens the luminance histogram. Despite its popularity, HE does not always give satisfying results, mainly characterized by overexposures and overemphasized textures. Many authors published works to improve HE, mainly by alleviating its side effects. For instance, the technique of adaptive histogram equalization (AHE) [56] performs HE on subimages and then fuses the local HE results. Contrast limited AHE (CLAHE) [78] is a famous variant of AHE which effectively relieves overexposures. More HE-based algorithms have been introduced since then for both grayscale and colour images, including 3-dimensional HE for RGB images [26, 68], optimization based HE [3] and HE in logarithmic transform domain [2], *etc*[14, 64]. Note that all the HE variants mentioned above inherit the spirit of HE that the output histogram should be as uniformly distributed as possible so as to reach maximal global contrast. This view is criticized in [72], in which the author reexamines HE-based algorithms and points out that histogram uniformity is a poor proxy for high contrast.

Another family of methods adopt new design criteria rather than histogram equalization. They try to optimally trade off between high contrast and image naturalness. For instance, Arici *et al.* propose a Langragian optimization scheme to balance histogram uniformity and histogram faithfulness [3]. The automatic exposure correction algorithm in [75] tries to find the optimal S-curve mapping according to the distribution of luminance. Wu proposes an optimization framework named optimal

contrast-tone mapping (OCTM) that maximizes global contrast under a tone distortion constraint. The methods in [6] and [59] use contrast metrics of second-order statistics.

1.3 Image Contrast and Naturalness

All image enhancement algorithms are designed to serve a common purpose: to boost the visibility of object surfaces and edges. This design goal agrees with the fact that human cognition of the physical world is rooted in visual interpretation of object surface and edge properties, such as the geometric, textural and spectral characteristics. This visibility is largely a synonym to image contrast; therefore, sharp contrast of edges and local patterns on the object surfaces correlates to high visual quality. The problematic images in Fig. 1.2 suffer from low contrast (either globally or locally); details and textures get severely degraded and become illegible. Situation can be even worse: as shown in Fig. 1.3(a), due to extremely low contrast of some patches and the image in general, observers can hardly recognize any objects in the scene. Fortunately, despite the seemingly total illegibility to human observers, the degraded information is not completely lost in most cases; therefore it is possible for image processing techniques to repair the damaged details by increasing the contrast in regions of low dynamic range (Fig. 1.3(b)). For the same reason, contrast has been used as a key factor of perceptual image quality [2, 34, 54, 71].

However, perceptual image quality is more than simply high contrast. The natural appearance of the image is also important and should not be compromised merely for the sake of high contrast. Many existing methods can indeed achieve high contrast with heightened textures and details, but at the same time they also introduce various

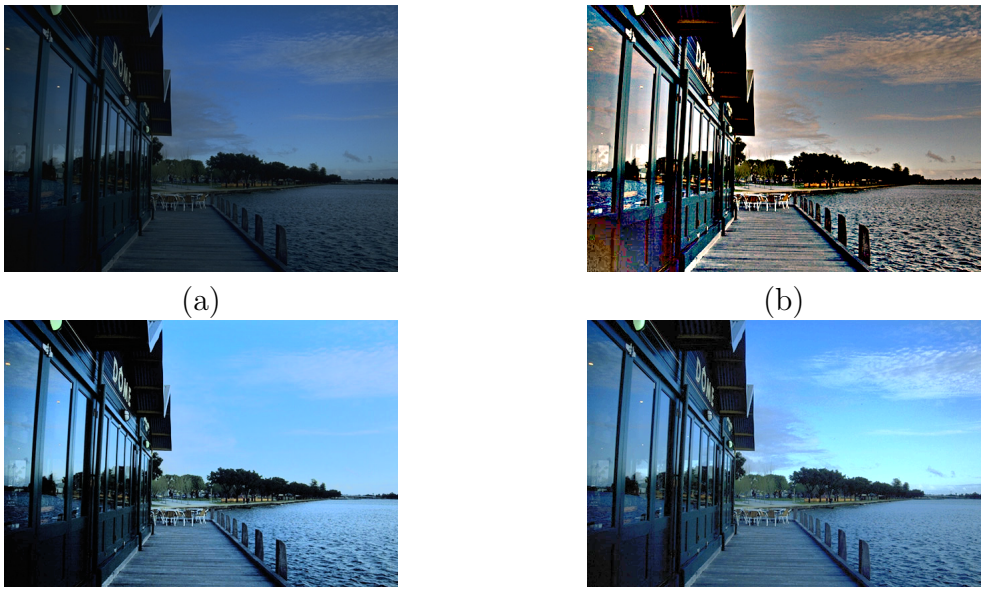
(a) (b)

Figure 1.3: An image with extremely dark appearance (a) still contains information that can be repaired by the proposed enhancement technique, resulting in (b).

types of visual artifacts, making images appear unnatural. Below we summarize the most frequently encountered artifacts with existing image enhancement methods.

- Halo artifact. Halo artifact commonly occurs when applying transform domain (*e.g.* frequency or wavelet domain) based or spatially adaptive high-pass filters. Fig. 1.4(b) shows a typical halo effect: a glowing layer appears between the sky and the roof of the house. The main cause of halo artifact is that high pass filtering tends to alter the local order statistics of pixel values.

- Tone distortion. Camera sensors advance rapidly in dynamic range in recent years. Nowadays, 10-bit (*i.e.* supporting 1024 intensity levels) image sensors are commonplace and have been widely adopted in smartphone cameras; sensors in higher-end digital cameras can reach even higher precision up to 16-bit. However, the raw sensor data has to be quantized into 8-bit precision for display and transmission. The quantization procedure removes the subtle differences between adjacent intensity levels as multiple intensity levels are quantized into



(a) (b)

(c) (d)

Figure 1.4: Frequently encountered artifacts with existing image enhancement methods. (a) Input image; (b) enhancement result by [35] with halo artifact; (c) enhancement result by [3] with tone distortion in the clouds, almost washed out; (d) enhancement result by HE with chromaticity distortion where the colour of the sky is turned into cyan.

one, resulting in distortion in tone.

Such many-to-one mappings may also occur in image enhancement methods involving luminance mapping, causing tone distortion. Fig. 1.4(c) shows an example: the enhancement algorithm improves the visibility of most objects in the scene, but it also erases the clouds from the sky.

- Chromaticity distortion. By conventional definition, the term contrast is concerning the luminance channel solely, and so are the aforementioned artifacts. But, observable distortions in chromaticity can also be found as in colour images generated by many existing enhancement methods. For example, in Fig. 1.4(d), the colour of the sky is turned from blue to cyan by an enhancement method.

Such effect also exists in Fig. 1.4(b)(c).

- Noise. As mentioned before, images inevitably have background noises due to imaging sensor imperfection. The noises are typically white, temporally variant and independent at each pixel, with considerable contribution in high frequency range [52]. Since image enhancement algorithms try to amplify the high frequency components of images [24], they cause the side effect of exaggerating the noises.

Several previous works are dedicated to modeling enhancement artifacts as well as devising naturalness-aware enhancement methods. [31] propose an empirical image naturalness metric and utilize it to evaluate enhancement methods. Edge-aware filters such as bilateral filtering [67] and guided filtering [30] can smooth pixels in an object surface while preserving edges, thus effective in neutralizing sensor background noises and suppressing halo artifacts [19]. In [72], the author proposes a mathematical model of tone distortion, and develops a tone distortion constrained enhancement algorithm based on the model.

Past research on image enhancement is mostly devoted to the manipulation of the luminance channel. Chromaticity distortion has been largely overlooked till the writing of the thesis. Ironically, research on colour science [37] reveals that although the HVS can only distinguish roughly 450 gray levels, it can detect a significantly larger number, around one million, of different colours. In other words, the chrominance channels carry far more information than the luminance channel alone. As such, colour fidelity plays a vital role in perceptual image quality, in addition to contrast. This motivates us to study chromaticity distortions and investigate new methods that can enhance low-lighting images while preserving colour fidelity.

In summary, increasing contrast is capable of improving the sharpness of edges and textures, but it can also harm the naturalness of images if done without care. Balancing between contrast and naturalness is therefore critical in image enhancement. We address the balancing problem by a novel approach of constrained optimization; specifically, we formulate and carry out image enhancement as an optimization problem with contrast being objective function and the aforementioned distortions being constraints. This methodology of balancing between contrast and naturalness is used in the development of all the proposed methods in the thesis.

Recall that in Sec. 1.2 we reviewed several optimization-based enhancement methods; some of them also advocate the above said balancing. But many of the objective functions and constraints are empirical and quite ad hoc; for examples, assuming that uniform histogram is optimal, or an S-curve is optimal [75]. In contrast, our optimization criteria are based on a physical image formation model and the knowledge of colour science. The validity and effectiveness of our optimization methodology are corroborated by the results of extensive experiments on severely underexposed and overexposed images.

Moreover, the proposed optimization-based framework can be easily extended (to gain desired properties, so as to) deal with more complex lighting conditions that overwhelm most existing image/video processing approaches. A typical example is to restore images suffering from backlighting (Fig. 1.2(c)), where an image consists of two types of regions that require drastically different adjustment as mentioned in Sec. 2.1. Unlike existing context-free enhancement methods that operate on both regions indiscriminately, we combine our optimization-based method with learning-based segmentation to employ two optimal tone mappings, one on each region. Such

combination adds spatial adaptivity to the original enhancement method, without introducing spatial artifacts. We also explore the means to ensure temporal consistency, apart from spatial adaptivity, when processing videos. This is achieved by an inter-frame harmonization technique between frames on top of the original enhancement method.

1.4 Research Contributions

The thesis presents a series of algorithms that restore images and videos of undesirably illuminated scenes. The major contributions are listed as follows.

- We address the long-time neglected chromaticity distortions with existing enhancement methods, which may introduce highly noticeable artifacts degrading the naturalness of resulting images. We conduct quantitative analysis on the chromaticity error and devise a novel scheme to suppress it. Based on this, we propose a novel optimization-based contrast enhancement framework for colour images that maximizes global contrast with zero halo artifact as well as constrained tone and chromaticity distortions [44]. We further develop a fast algorithm that solves the optimization problem by dynamic programming.

- We propose a learning-based spatially adaptive technique of optimal tone mapping to restore backlit images. Object surfaces illuminated from behind in a scene are detected by a soft, binary classifier that is constructed by supervised learning. Two optimal tone mapping functions, one for backlit regions and the other for the remainder of the image, are used and their outputs are fused to restore illegible surface details in backlit regions and at the same time improve

contrast in overexposed regions, if any [45, 46].

- We extend our work to videos, addressing the problem of Yin-Yang Phasing (YYP), which is characterized by involuntary and abrupt flip-flop in luminance and possibly chrominance due to incorrect auto-exposure of cameras. We then propose a video restoration technique that neutralizes YYP artifacts while retaining temporal consistency in terms of exposure and chrominance of object surfaces in the scene via inter-frame, spatially-adaptive optimal tone mapping [73].

1.5 Thesis Organization

The rest of the thesis is organized as follows. In Chapter 2 we systematically analyze the causes of artifacts accompanying current image enhancement methods, and investigate how to minimize them and preserve image naturalness. Having exposed the weaknesses of existing methods, we propose a novel optimization-based framework for image enhancement and develop a fast dynamic programming algorithm to solve the optimization problem in Chapter 3. In Chapter 4, we present a novel solution to restore backlit images, by combining the above enhancement method and a supervised learning based segmentation method. In Chapter 5, we extend the work in Chapter 3 from images to videos; we address YYP, a common yet largely untreated degradation problem with videos, and propose a novel inter-frame, spatially-adaptive optimal tone mapping algorithm to eliminate YYP artifacts and maintain temporal tonal consistency between frames at the same time. Finally Chapter 6 summarizes the thesis.

Chapter 2

Chrominance Preservation in Image Enhancement

2.1 Introduction

As pointed out in Sec. 1.3, a common problem with existing image enhancement methods is that they only focus on heightening spatial details in the luminance channel with no or little consideration of potential loss of colour fidelity. For certain scenes this deficiency can be a serious perceptual quality issue, as it produces objectionable colour distortions in enhanced images, such as hue shifting and low saturation. In pursue of higher chrominance precision in image enhancement, we launch, in this section, an in-depth study into the root cause of chrominance distortions of existing contrast enhancement methods and strive to prevent or at least alleviate such distortions by improving the existing methods. We discover that the limited range of the 3D colour spaces used in practice is the main cause of chrominance distortions. This is because increasing contrast often demands reproduction colours to fall outside of the device

gamut; the necessary remapping back within the RGB cube of the display hardware easily shifts the original hues, if done without due care. Accordingly, to prevent shifted hues, we propose an optimal clipping strategy that keeps the hue intact, while minimizing the distortion in saturation.

The remainder of this chapter is organized as follows. In Sec. 2.2 we review the general work flow of enhancement methods in relation to the physical image formation process. And then in subsequent sections, we explore various means of preserving colour fidelity when manipulating contrast. In Sec. 2.3, we propose a novel LHS colour space that is designed to facilitate psychovisually based image enhancement; in Sec. 2.4 and 2.5 we expose the out-of-gamut problems, the main cause of chrominance distortions in enhancement methods, and develop an optimal clipping scheme that aims to minimize such chrominance distortions in Sec. 2.6 and 2.7.

2.2 Image Formation and Enhancement Model

We use a physical image formation model to shed some light on the relationship between luminance and chrominance. A camera of spectral response $s(\lambda)$, λ being the wavelength, generates the colour (spectral) signal at pixel (x, y):

$$I(x, y, \lambda) = s(\lambda)\ell(x, y, \lambda)r(x, y, \lambda), \tag{2.1}$$

where $\ell(x, y, \lambda)$ is the illumination light, $r(x, y, \lambda)$ is the spectral reflection function of the small surface area corresponding to pixel (x, y). In natural outdoor scenes, the

illumination source $\ell(\cdot)$ is typically white (constant energy level for all λ), thus

$$I(x, y, \lambda) = s(\lambda)\ell(x, y)r(x, y, \lambda). \tag{2.2}$$

More specifically, the spectral reflection $r(\cdot)$ can be modeled by the bidirectional reflectance distribution function (BRDF) [66]. BRDF splits the spectral reflection $r(\cdot)$ into two components, said diffuse and specular reflections:

$$r(x, y, \lambda) = k_d(\lambda)\cos\alpha(x, y) + k_s(\lambda)\cos^s\beta(x, y),^1 \tag{2.3}$$

where the first term stands for diffuse reflection, with $\alpha(x, y)$ being the angle between incident light and the surface normal located at pixel (x, y); the second term approximates specular reflection, with s being the degree of surface glossiness and $\beta(x, y)$ being the angle between the view direction and specular direction (see Fig. 2.1). Both $\alpha(\cdot)$ and $\beta(\cdot)$ depend on the geometric properties of the surface. On the other hand, $k_d(\lambda)$ and $k_s(\lambda)$ are both determined by the spectral features of the surface material but independent of the geometric properties.

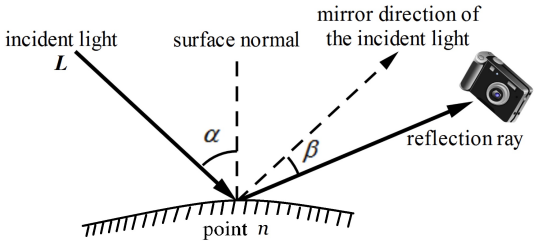


Figure 2.1: An illustration of $\alpha(x, y)$ and $\beta(x, y)$ in BRDF(2.3).

Given the scene to be imaged and the camera, $s(\lambda)$ is fixed and so is $r(\cdot)$. These

[1]The formula holds for surfaces made of isotropical material. Most surfaces in natural scenes are isotropic.

two terms $s(\lambda)$ and $r(\cdot)$ are completely determined by the camera and the scene being imaged. The quality of image I can be interpreted as the amount and fidelity of details regarding the scene (conveyed by r) that an observer can perceive from the image. To improve the quality, one can only adjust the illumination term $\ell(\cdot)$, either physically (e.g., adding auxiliary lighting, adjusting exposure time and/or lens aperture), or algorithmically through digital image processing techniques. Indeed, improper illuminations and/or operator's errors on exposure time setting are main causes of poor image quality, in the form of either underexposure (too low an ℓ to reveal subtle details) or overexposure (so high an ℓ that the sensors saturate). Rectifying the problem through physical means like in a professional photo studio is cumbersome, unportable, and expensive; furthermore, this requires the knowledge of expert photographers. A much more attractive alternative is to compensate for the imperfection of illumination conditions or operator's errors by image processing methods.

Under white illumination conditions, if one could vary ℓ to improve image quality, the changed light source should remain white as well; otherwise, noticeable unnatural chrominance distortions may occur in the enhanced image. For the same visual effect the enhancement method should increase contrast with no or little changes in the spectral distribution of the object surfaces, which is essentially determined by the spectral response patterns of 1) the camera $s(\lambda)$ and 2) the surface material at the pixel $k_d(\lambda), k_s(\lambda)$, but not ℓ. However, none of existing contrast enhancement techniques are designed to preserve chrominance. When being applied to colour images, almost all image enhancement methods adopt a naïve separation approach as illustrated in Fig. 2.2. An RGB input image I is first transformed into another colour

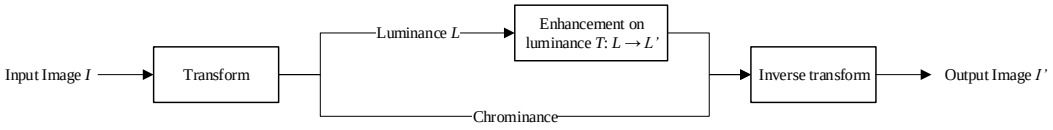


Figure 2.2: A widely used pipeline for existing colour image enhancement techniques.

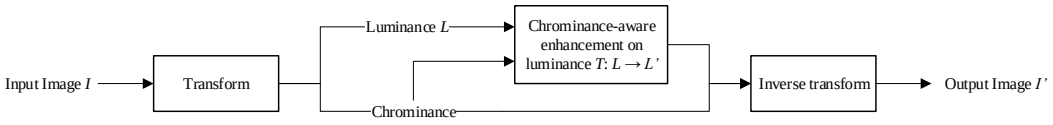


Figure 2.3: Pipeline of the proposed enhancement method.

space that decouples luminance and chrominance (a two-dimensional subspace that specifies spectral distribution independent of luminance). I is then enhanced on the luminance channel (denoted by $\mathcal{L}$) by a mapping function T, and finally transformed back to RGB space to form the output image I'. The last step is carried out by combining the enhanced luminance (denoted by $\mathcal{L}'$) and the original chrominance information.

Independent adjustment of luminance, as in current practice, is prone to chrominance distortions in the resulting image, even without any explicit modification of chrominance components. An image enhancement method should not manipulate the luminance channel in isolation; instead it should jointly consider the luminance and chrominance channels to prevent chrominance distortions. This new design principle is schematically described by Fig. 2.3.

2.3 Colour Space

Under most lighting conditions, human perception of colours originates from the excitations of the three types of cone cells by lights in visible spectrum. The excitation of a cone cell depends on the intensity and spectrum of the incident light; each type of cone cells has a specific response function to wavelength of light. As such the colour sensation is intrinsically a trivariate function, physically corresponding to the trichromatic cone cell reactions. A colour space is a commonly used tool to organize (and distinguish) perceivable colours, which is typically three-dimensional to encode trichromatic colour vision. Existing colour spaces are designed for various purposes with different meanings of the three components. For example, the RGB space is hardware oriented, as most cameras and displays work in red, green, and blue primary colours. The YUV space is designed for reducing redundancy among colour components for image/video compression applications.

However, most of these colour spaces are not designed specifically for image enhancement tasks. In this work, we investigate the colour space whose dimensions are most favourable to the goal of enhancement, that is 1) increasing contrast, while 2) keeping chrominance intact. In the subsequent subsections below, we justify each selected component of our novel, enhancement-oriented colour model, called LHS.

2.3.1 Luminance Component

As mentioned in Sec. 2.2, most enhancement methods in literature associate contrast with variance in luminance. To increase contrast, they adopt colour spaces with a luminance channel along with another two chrominance channels, such as HSI [14, 76], HSV [27], YUV [72, 74], *etc.* But these colour models have a drawback in

common: there is no explicit correlation between perceptual contrast and luminance. This leads to inconsistent adjustment on objects with different luminance values, where some objects are greatly enhanced whereas others are still of low contrast.

Instead, we adopt the L^* component from the CIELab colour space because of its property of perceptual uniformity: the distance between two colors in the CIELab space is approximately linear to the perceptual difference of the two colors. Thanks to this property, the distance in the L^* axis can be interpreted as contrast.

For an RGB colour vector $\mathbf{c} = (r, g, b)$, the corresponding L^*, ranging from 0 to 100, is defined as

$$\begin{aligned} L^* &= f(Y), \\ Y &= 0.2126r + 0.7152g + 0.0722b, \end{aligned} \tag{2.4}$$

where Y is weighted average of R, G and B that corresponds to the measured brightness, *i.e.*, perceived luminance by human (on devices under Rec. 709 standards [33]); human eyes are, as well known, most sensitive to green and least sensitive to bluish light. Y is subsequently nonlinearly transformed by $f(\cdot)$ to gain perceptual uniformity. Details of $f(\cdot)$ can be found in [17].

For convenience, we adopt the normalized L^*, *i.e.*, $L \triangleq 0.01L^*$, which ranges from 0 to 1 in the proposed colour space.

2.3.2 Chrominance Components (Hue and Saturation)

Since we adopt the luminance component L^* from the CIELab space, a straightforward method is to use the CIELab space directly in image enhancement. But its two chrominance channels a^* and b^* cannot determine a specific spectrum as their ranges

are correlated with the luminance; therefore preserving chrominance dimensions a^* and b^* cannot ensure the constancy of colour spectrum. In order to specify colour spectrums independent of luminance, the proposed colour space adopts hue and saturation as its chrominance components. Hue corresponds to dominant wavelength of the spectrum of a colour, specifying the overall colour impression (*e.g.* red, yellow and green); saturation corresponds the relative strength of the dominant wavelength of a colour over other wavelengths, specifying how much the colour is diluted by white light.

We adopt the mathematical definitions of hue H and saturation S from the HSI colour space [24]:

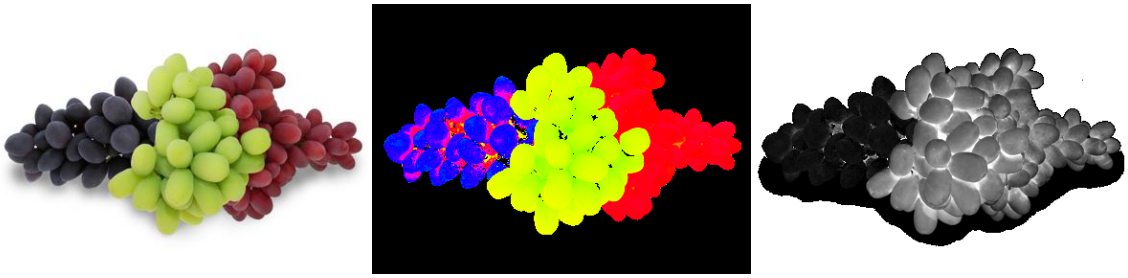
$$H = 60^\circ \times H',$$

$$H' = \begin{cases} \text{undefined} & \text{if } M = m \\ \frac{g-b}{M-m} \mod 6 & \text{if } M = r \\ \frac{b-r}{M-m} + 2 & \text{if } M = g \\ \frac{r-g}{M-m} + 4 & \text{if } M = b \end{cases} \tag{2.5}$$

$$S = 1 - \frac{m}{I} \tag{2.6}$$

where $M = \max\{r, g, b\}, m = \min\{r, g, b\}$, and $I = (r + g + b)/3$. Hue H is a circular value in the polar angle range $[0, 360^\circ)$; an angle corresponds to a particular wavelength of the light. Saturation S ranges from 0 to 1, representing the relative purity of the dominant wavelength in the light; a larger S value corresponds to a more vivid, "purer" colour.

As the hue and saturation of a pixel is completely specified by the spectral distribution, they are approximately constant if the object surface area pertaining to the pixel consists of a consistent material of a given spectral signature. Demonstrated in Fig. 2.4 is such an example, where pixels of the same object surface material have similar hue and saturation values regardless of luminance. Thus, an enhancement approach that adjusts the luminance channel should keep hue and saturation values intact in order to preserve the naturalness of the resulting image.



(a) (b) (c)

Figure 2.4: Hue and saturation of a set of coloured wool balls. (a) A colour image; (b) hue map of (a) given by (2.5); (c) saturation map of (a) given by (2.6), where white corresponds to 1 and black corresponds to 0.

2.4 Basic Colour Constructs

As discussed in Sec. 2.2, adjusting luminance L while keeping hue H and saturation S intact is a way of increasing contrast without altering chrominance. But working in the LHS colour space alone cannot guarantee zero error in chrominance. Like other colour spaces, the LHS space is finite, bounded by surfaces called *gamut* due to the limited number of colours supported by display devices and/or perceivable by HVS, and problem will arise once the enhancement operator $T : \mathcal{L} \rightarrow \mathcal{L}'$ yields a colour vector outside the gamut. We expose such so-called out-of-gamut problem in

the following sections, and investigate how to maximally preserve colour fidelity when contrast enhancement pushes the input colour outside the device gamut.

For convenience and clarity of presentations, we introduce the terms as follows to be used later.

RGB Cube

As all the three dimensions are orthogonal with each other and range in $[0, 1]$, RGB colour space is cube-shaped as depicted in Fig. 2.5(a).

Isospectral Line

Looking from (2.5) and (2.6), colours with identical hue and saturation have the same ratio among r, g and b values, thus located along a ray that emits from the black point O. Preserving hue and saturation is then equivalent to keeping the colour staying on the corresponding ray during enhancement process. We call the ray on which colours have consistent hue and saturation values an isospectral line.

It should be noted that, due to the limitation of gamut, isospectral lines practically end at a gamut surface.

Saturated Colour

Saturated colours are those without being diluted by white light, in other words with maximal saturation ($S = 1$), or $\min\{r, g, b\} = 0$. The set of pure colours forms the three surfaces of the RGB cube that include black point O (the shaded surfaces in Fig. 2.5(a)).

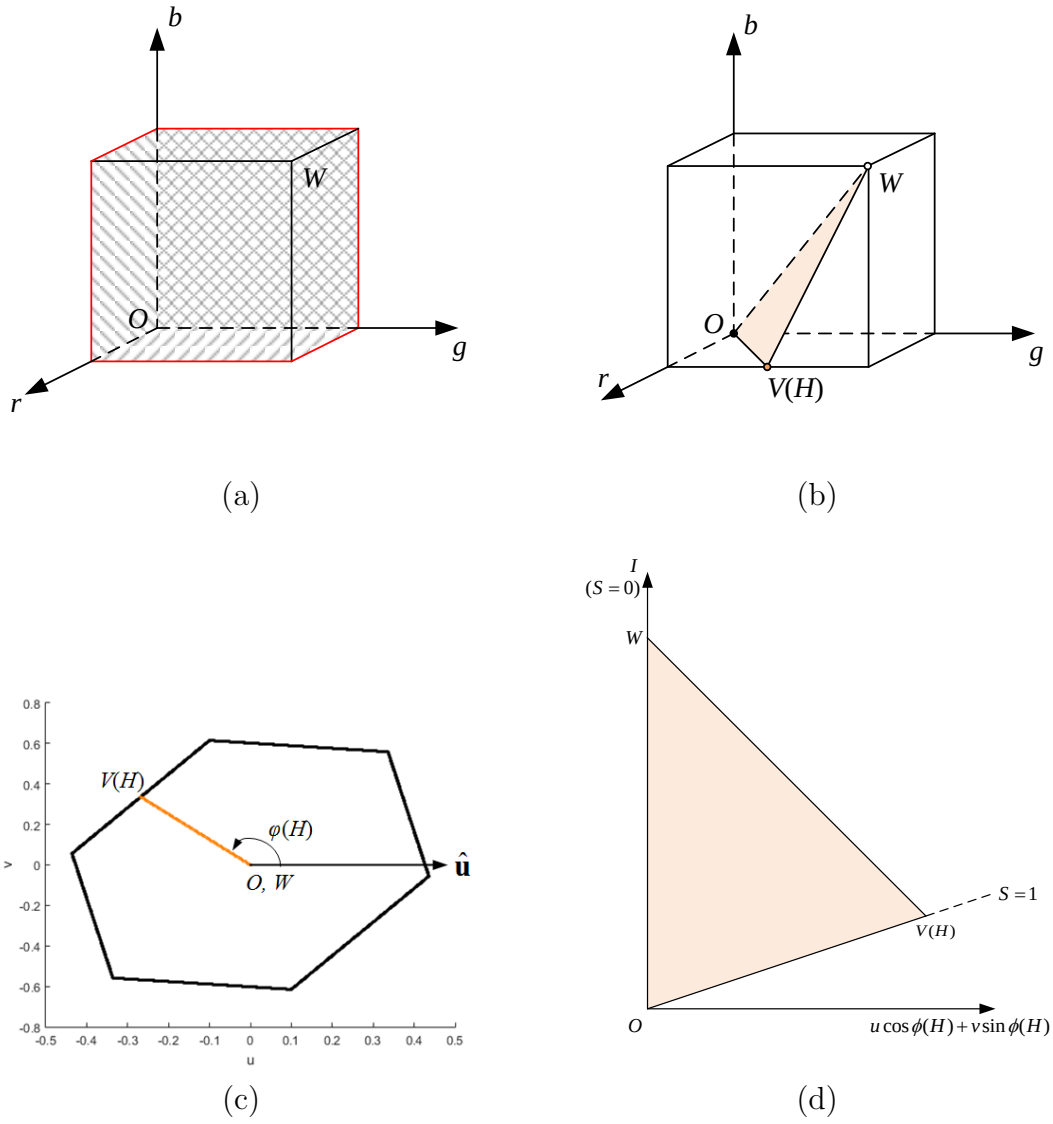


Figure 2.5: (a) The RGB cube, where pure colours are those on the three shaded surfaces and vertex colours are those on the red edges; (b) an equihue triangle in the RGB cube; (c) projection of RGB cube and the equihue triangle in (b) on U-V plane, marked in black and orange respectively; (d) plane view of the equihue traingle in I-ϕ plane.

Extreme Colour

Extreme colours are saturated colours satisfying $\max\{r, g, b\} = 1$. They are located at the edges marked in red in Fig. 2.5(a). Each hue value H corresponds to a unique extreme colour, denoted by $V(H)$.

Gray Axis

The gray axis **I** corresponds to the axis along the diagonal vector from black point $O(0, 0, 0)$ to white point $W(1, 1, 1)$. Colours on the gray axis have zero saturation.

Equihue Triangle

The equihue triangle of hue value H is a triangular planar region enclosed by O, W, and a vertex colour point $V(H)$. According to (2.5), all colours lying in the region have the same hue value H. Fig. 2.5(b) is an example of equihue triangle in the RGB cube.

U-V Plane

U-V plane is a plane orthogonal to the gray axis and thus orthogonal to every equihue triangle. Here $u = -0.0999r - 0.3261g + 0.4360b$ and $v = 0.6150r - 0.5586g - 0.0564b$ are from the YUV colour space [9][2]. In the polar coordinate system with the pole located at $(u, v) = (0, 0)$, colours along any polar axis have the same hue due to the fixed ratio between r, g and b, as illustrated in Fig. 2.5(c). We denote the polar angle corresponding to the equihue triangle of hue H by $\phi(H)$; the polar axis in vector

[2]Both vectors $(-0.0999, -0.3261, 0.4360)$, $(0.6150, -0.5586, -0.0564)$ are orthogonal to the gray axis $(1, 1, 1)$.

form is then $\mathbf{p}(H) \triangleq \hat{\mathbf{u}}\cos\phi(H) + \hat{\mathbf{v}}\sin\phi(H)$, $\hat{\mathbf{u}} = (-0.0999, -0.3261, 0.4360), \hat{\mathbf{v}} = (0.6150, -0.5586, -0.0564)$ in RGB space.

Plane View of Equihue Triangles (on I-p Plane)

We can obtain the plane view of an equihue triangle by taking the plane determined by its corresponding polar axis $\mathbf{p}(H)$ and the gray axis $\mathbf{I}$, as depicted in Fig. 2.5(d). This plane view offers us a more intuitive understanding on equihue triangles than the RGB cube view (Fig. 2.5(b)).

Isospectral lines correspond to the rays emitting from the origin O. As depicted in Fig. 2.5(d), saturation is 0 on the gray axis (OW), and 1 on (OV). Saturation value decreases as the slope of the ray goes up.

Maximal Adjustable Luminance

Recall that ideally our goal is to change the luminance of a colour without changing its hue or saturation. As the isospectral lines are bounded by the gamut, there is an upper limit of the output luminance, which is determined by the end of the corresponding isospectral line segment. For any colour vector $\mathbf{c} = (r, g, b)$, the corresponding in-gamut end color point is, as illustrated in Fig. 2.6,

$$\mathbf{c}_V = \mathbf{c} \cdot \frac{1}{\max\{r, g, b\}} \tag{2.7}$$

The luminance of $\mathbf{c}_V$, $L(\mathbf{c}_V)$ or the *maximal adjustable luminance* of $\mathbf{c}$ is

$$\eta(\mathbf{c}) = L\left(\mathbf{c} \cdot \frac{1}{\max\{r, g, b\}}\right) \tag{2.8}$$

2.5 Necessity of Colour Clipping

Unless $\mathbf{c}$ is colourless (*i.e.* with zero saturation), the maximal adjustment luminance $\eta(\mathbf{c})$ is always strictly less than 1. The out-of-gamut problem will arise if we increase the luminance of $\mathbf{c}$ to a value greater than $\eta(\mathbf{c})$, in which case the colour point falls outside the RGB gamut. This is quite common in image enhancement, especially when coping with underexposed images.

The out-of-gamut problem has been investigated in a number of research projects, but most of them are on the topic of device gamut mapping [38, 65]. Device gamut mapping techniques aim to harmonize the different gamuts of display devices and printers, in order to optimize the visual quality of colour printing and display.

More relevant to the theme of this thesis are a couple of papers that discuss how to move out-of-gamut colours created by image processing algorithms back into a realizable gamut. A straightforward method is to truncate all RGB values greater than 1 after the inverse transform in Fig. 2.2. This typically introduces errors in all three colour attributes (*i.e.* luminance, hue and saturation) as the truncation is likely to push a colour vector away from the isospectral line, and even out of the equihue triangle. Moreover, such errors in the three dimensions may be inconsistent among different pixels, further aggravating the colour artifacts. An alternative solution is to normalize the resulting R, G, B values after the inverse transform such that $\max\{r, g, b\} = 1$, as done in [14, 74]. Although causing no chrominance error, the above method is too conservative because it greatly reduces the extent of luminance adjustment and thus cancels out the effects of contrast enhancement. Other schemes in the literature include [74] that clips saturation in YIQ colour space, and [51] that makes the tone mapped results fall within the RGB gamut. But we find the above

two algorithms to have large distortions in chrominance or/and luminance.

Fig. 2.6 illustrates a typical out-of-gamut problem where a luminance mapping function attempts to brighten $\mathbf{c}$ to a luminance value l' along the isospectral line ($S = s$, s being the saturation of $\mathbf{c}$), resulting in an out-of-gamut colour vector $\mathbf{c}'$, which must be clipped back into the gamut so as to be rendered on RGB display devices. In this case, there is no perfect clipping scheme that preserves both hue and saturation, unless we alter the luminance from the value $\mathcal{L}'$ determined by the enhancement operator $T : \mathcal{L} \rightarrow \mathcal{L}'$. But $\mathcal{L}'$ should not change in order not to negate the effects of contrast enhancement operator. Therefore, the performance of a colour clipping scheme is determined by the level of chrominance distortions caused by altering hue and/or saturation.

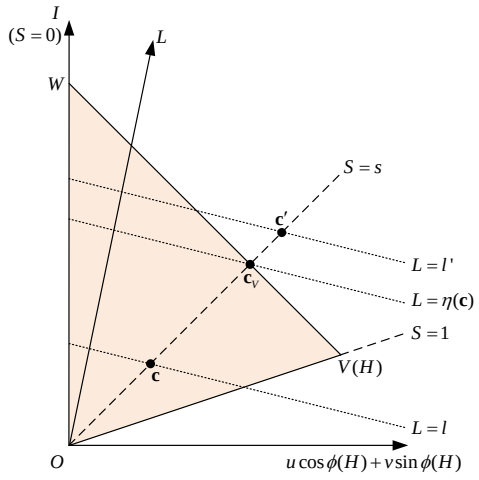


Figure 2.6: An example of out-of-gamut problem. Given a colour vector $\mathbf{c}$, out-of-gamut problem arises if an enhancement algorithm changes the luminance of $\mathbf{c}$ from l to l'

2.6 Perceptual Sensitivity of Hue and Saturation

In order to discover ways to best preserve colour fidelity in image enhancement, we study the sensitivity of the human visual system to changes in hue and saturation. The perceptual difference between two colours is very difficult to quantify linearly. This requires a colour space that is perceptually uniform; namely, in a uniform colour space the Euclidean distance between any two colour points accurately measures their perceptual distance. The CIELCh colour space (the polar coordinate variant of the CIELab space) [17] is the one that achieves the best approximate perceptual uniformity up to now, and hence we adopt it in the following analysis.

The CIELCh space consists of three dimensions: luminance (L), chroma (C) and hue (h). Chroma, defined by $C = \sqrt{a^2 + b^2}$ (for simplicity, we drop the star superscript of a^*, b^* in the CIELab notation), describes the colourfulness of a colour; thus chroma has similar physical meaning to saturation[3]. The CIELCh hue is highly correlated with the hue in LHS colour space, defined by $H = \arctan(b^*/a^*)$. To avoid ambiguity with the hue defined by (2.5), we denote the CIELCh hue by H_U. It can be inferred from the definitions of C and H_U that, the a-b plane and the C-H_U plane corresponding to the same luminance are identical.

In the CIELCh colour space, let ΔL, ΔC and ΔH_U be the difference between two colour points in luminance, chroma and hue, respectively. The CIEDE2000 perceptual difference ΔE between the two colours has the form [47]

$$\Delta E = \sqrt{\left(\frac{\Delta L}{k_L S_L}\right)^2 + \left(\frac{\Delta C}{k_C S_C}\right)^2 + \left(\frac{\Delta H_U}{k_H S_H}\right)^2 + R_T f(\Delta C \Delta H_U)} \qquad (2.9)$$

[3]The main difference between chroma and saturation is that the range of chroma depends on luminance and hue, but that of saturation is constant to be $[0, 1]$. More saturated colours take larger values in both chroma and saturation.

where $k_{\{L,C,H\}}, S_{\{L,C,H\}}, R_T$ are approximately constant here and $f(0) = 0$. The formula (2.9) originates from psychophysical experiments revealing that the set of just-noticeable different colours with respect to a reference colour form an elliptical pattern, called colour discrimination ellipse, in the C-H_U plane (*i.e.* a-b plane), as shown in Fig. 2.7. For our design purposes, the colour discrimination ellipse manifests an important property: human vision has different sensitivities to hue and chroma.

In order to analyze LHS colours using the CIEDE2000 formula (2.9), we need to associate the hue H and saturation S in the LHS space with CIELCh hue H_U and chroma C. As demonstrated in Fig. 2.8, for a fixed luminance, there is an approximately linear relationship between H and H_U, as well as between C and S. As a result, colour discrimination ellipses still roughly hold the elliptical shapes in H-S planes; and H_U and C can be treated as proxies for H and S when measuring the perceptual difference between two LHS colours.

Given a reference colour point $\mathbf{c}$, we slightly shift it along the tangential and radial directions by the same small distance r in the H-S plane. In the above two cases, the resulting increments in the H and S axes are:

- in the case of colour shifting in tangential (hue) direction: hue increment $\Delta_t H = \frac{r}{S} \times \frac{180^\circ}{\pi}$, saturation increment $\Delta_t S = 0$, S being the saturation of $\mathbf{c}$;
- in the case of colour shifting in radial (saturation) direction: hue increment $\Delta_r H = 0$, saturation increment $\Delta_r S = r$.

Due to the approximate linearity mentioned in the previous paragraph, we have $\Delta_t H \simeq \Delta_t H_U, \Delta_r C \simeq \alpha \Delta_r S$, $\alpha = C/S$ being a constant[4]. According to (2.9),

[4] α can be deemed as a constant given fixed hue and luminance, as verified in Fig. 2.8(b).

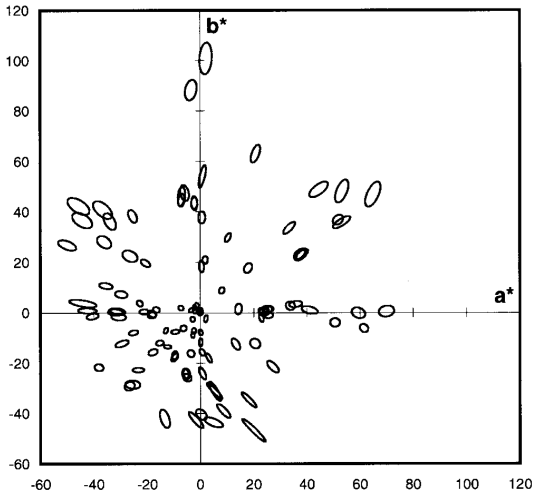
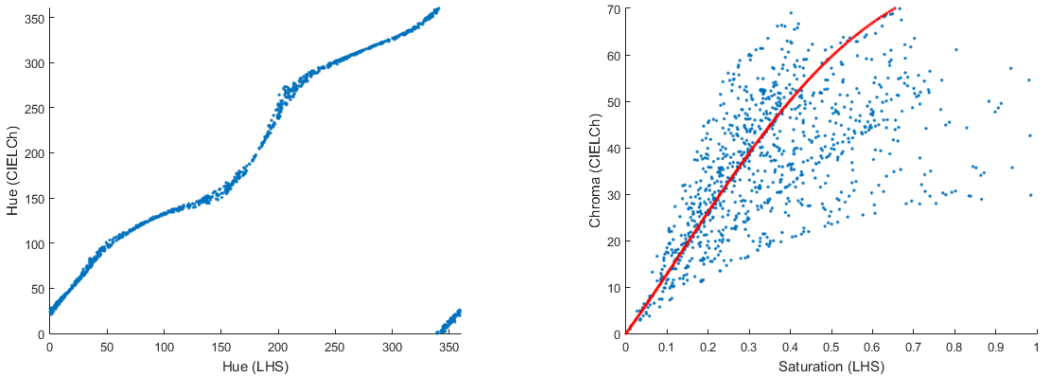


Figure 2.7: Colour discrimination ellipses in CIELab (CIELCh) colour space [47]. The contour of each ellipse is the just-noticeable different colour against the colour at centroid.



(a) (b)

Figure 2.8: Relationship between (a) H and H_U, and (b) between chroma C and saturation S for colours with $L = 0.5$. The red line in (b) is the set of colours with $L = 0.5, H_U = 0$, indicating the approximate linearity between chroma and saturation.

the perceptual differences caused by the shifting in the two cases are

$$\begin{aligned} \Delta_t E &= \frac{1}{k_H S_H} \Delta_t H_U \\ \Delta_r E &= \frac{1}{k_C S_C} \Delta_r C. \end{aligned} \tag{2.10}$$

We then define the hue-saturation sensitivity ratio with respect to a colour point $\mathbf{c}$ by $\Delta_t E / \Delta_r E$:

$$\frac{\Delta_t E}{\Delta_r E} = \frac{k_C S_C \Delta_t H_U}{k_H S_H \Delta_r C} \simeq \frac{k_C S_C \Delta_t H}{k_H S_H \alpha \Delta_r S} = \frac{k_C S_C}{k_H S_H} \cdot \frac{180}{\pi C} \tag{2.11}$$

The ratio reveals the relationship between the sensitivity of HVS to hue and to saturation. If it is greater than 1, then HVS is more sensitive to tangential (hue) changes than radial (saturation) changes given the reference colour $\mathbf{c}$. Conversely, if the ratio is less than 1, then HVS is more sensitive to alterations in saturation than in hue.

(2.11) shows that the ratio is independent of r and the luminance of $\mathbf{c}$; instead it only depends on the hue H_U and the chroma C of $\mathbf{c}$. Specifically, according to [47],

$$\begin{aligned} K_H = K_C &= 1 \\ S_H &= 1 + 0.015CT \\ S_C &= 1 + 0.045C, \end{aligned} \tag{2.12}$$

where

$$T = 1 - 0.17\cos(H - 30^\circ) + 0.24\cos 2H + 0.32\cos(3H + 6^\circ) - 0.2\cos(4H - 63^\circ) \tag{2.13}$$

We plot the ratio $\Delta_t E / \Delta_r E$ in the a-b plane, where $a = C\cos H_U, b = C\sin H_U$,

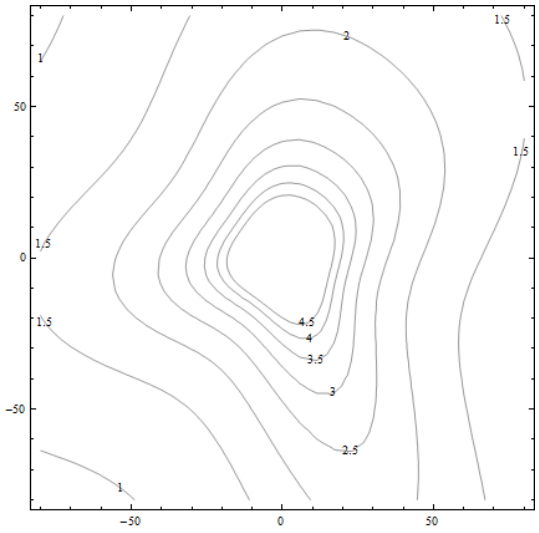


Figure 2.9: Contour map of $\Delta E_1/\Delta E_2$ over a-b (C-H) plane.

both ranging from -80 to 80 for 98% types of RGB colours[5], as shown in Fig. 2.9. It can be seen that $\Delta_t E/\Delta_r E > 1$ holds for overwhelmingly majority of colours, which implies that HVS is more sensitive to changes in hue than in saturation. This (Such observation) also agrees with the colour discrimination ellipses in Fig. 2.7, in which the major axes of the ellipses are mostly aligned with radial lines (in other words, a larger variation in saturation than in hue is required to make just noticeable difference).

Therefore we draw the conclusion that it is more important to preserve hue than to preserve saturation.

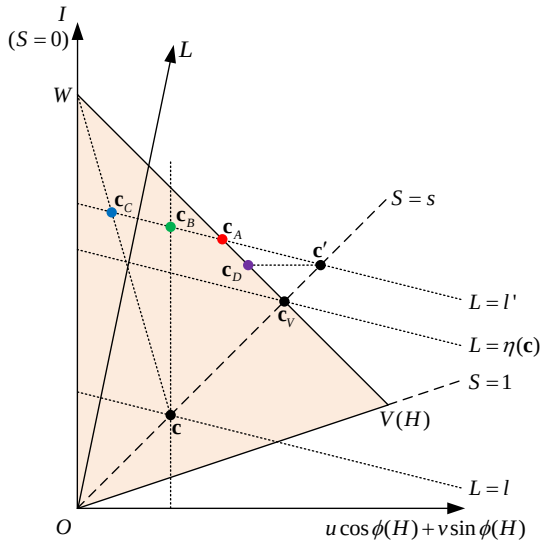


Figure 2.10: Various hue-preserving clipping methods that puts an out-of-gamut colour point $\mathbf{c}'$ back into the gamut, resulting in: $\mathbf{c}_A$ by optimal mapping, $\mathbf{c}_B$ by the method [72], $\mathbf{c}_C$ by the method [51] and $\mathbf{c}_D$ by the methods mentioned in [57] and [74]. Note that $\mathbf{c}_A$, $\mathbf{c}_B$ and $\mathbf{c}_C$ have the same hue and luminance as $\mathbf{c}'$ but $\mathbf{c}_C$ has the largest saturation error while $\mathbf{c}_A$ the smallest. $\mathbf{c}_D$ has distortions in both saturation and luminance.

2.7 Optimal Clipping Scheme

As discussed previously, an image enhancement algorithm should preserve hue, that is, confine the colour clipping within each equihue triangle. The optimal clipping is then the one that minimizes saturation error within the corresponding equihue triangle.

Consider an out-of-gamut problem where the luminance of a colour point $\mathbf{c}$ is changed from l to $l' > \eta(\mathbf{c})$ resulting in $\mathbf{c}'$. As illustrated in Fig. 2.10, the optimal solution to the out-of-gamut problem is to move the problematic colour $\mathbf{c}'$ along the

[5]We investigate the corresponding coordinates in the CIELab space of 1 million random RGB colours. Merely 14136 colours have a or b exceeding $[-80, +80]$.

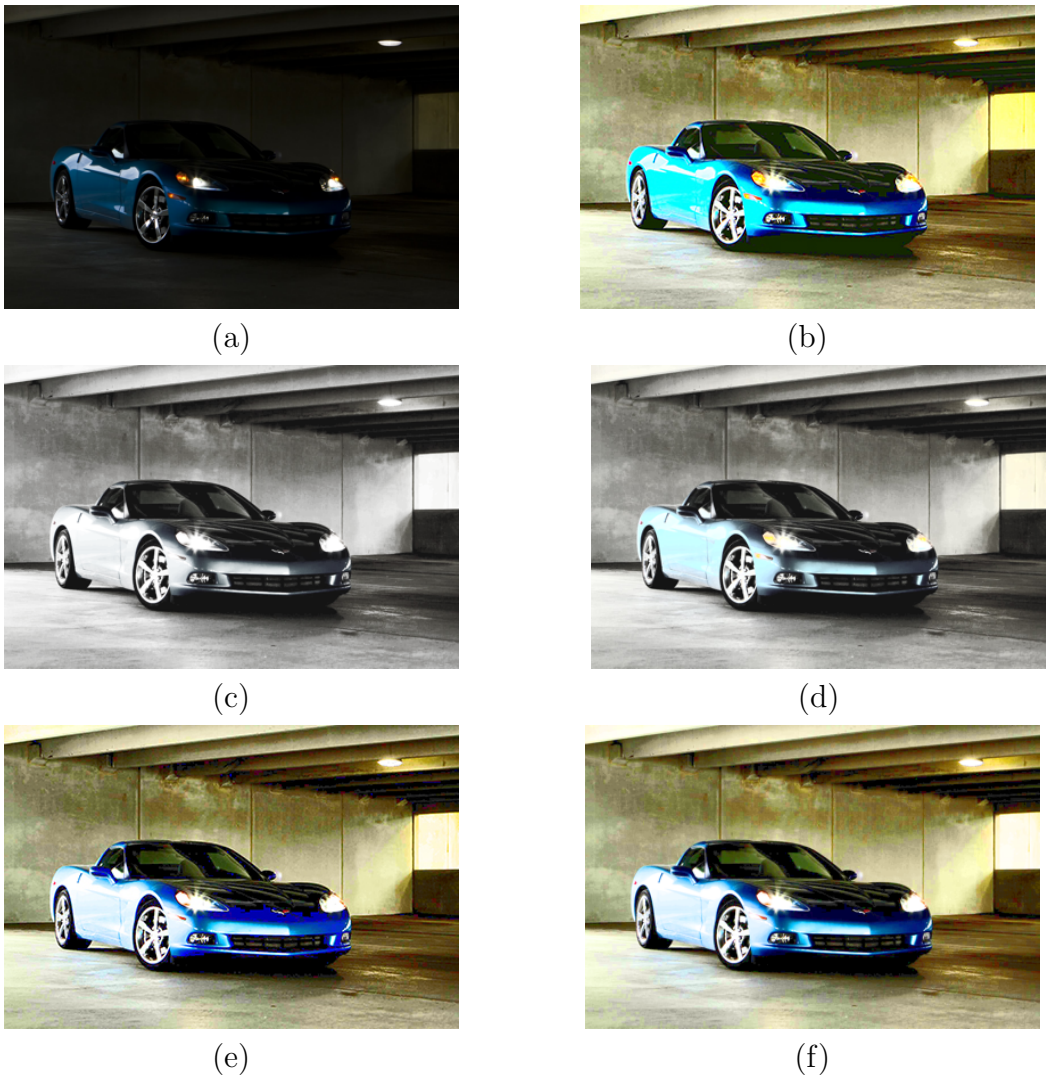


Figure 2.11: Results of HE using different clipping schemes: (a)original image; (b)traditional HE with simple clipping; (c)HE with Naik's method [51]; (d)HE in YUV colour space; (e)HE with clipping method in [57] and [74]; (f) HE with the optimal clipping in LHS colour space.

equi-luminance line $L = l'$ to the gamut surface, resulting in $\mathbf{c}_A$. Also exhibited in the figure are the clipping results from several other hue-preserving methods. A common practice of shifting luminance in Descartes coordinate system-based colour spaces (such as YUV, YCbCr and CIELab) maps the original colour vector $\mathbf{c}$ to $\mathbf{c}_B$, leading to greater saturation error. The clipping technique proposed in [51] tries to keep colours staying within the device gamut either in RGB or CMY (the inverted RGB) space by moving input colour $\mathbf{c}$ towards the white point W, resulting in $\mathbf{c}_C$, which suffers from even greater distortion comparatively. The clipping methods in [57] and [74] have the same spirit with the optimal clipping, but moves colours along incorrect equi-luminance lines[6] resulting in $\mathbf{c}_D$, causing distortions not only in saturation but also in perceptual luminance.

Fig. 2.11 demonstrates the impact of clipping, where (b)(c)(d)(e)(f) are all enhanced by HE on the luminance channel, but with different colour clipping schemes on the chrominance channels. Simple clipping severely distorts the hue, turning the blue car paint into cyan; the conventional method of luminance shifting in the YUV space and the method of forcing colours within in gamut [51] tend to desaturate the image, making the images look dull. The clipping methods in [57] and [74] go to the other extreme making colours over-saturated in some cases (see Fig. 2.11(e)). In contrast, the proposed optimal clipping method best preserves the chrominance with the same luminance mapping as HE. But it should be noted that even the above optimal clipping is not free of chrominance distortions , for instance the pale looking car doors in Fig. 2.11(f). This motivates us to find better enhancement algorithms to control chrominance distortions more tightly.

[6] [57] and [74] adopt $I = (r+g+b)/3$ and $Y = 0.299r + 0.587g + 0.114b$ as luminance respectively. Neither of them corresponds to the perceptual luminance of HVS.

Chapter 3

Algorithm for Image Enhancement with Bounded Perceptual Chrominance Errors

3.1 Introduction

As pointed out in the proceeding chapter, adjusting luminance to increase contrast as in existing image enhancement methods, may carry some colour points out of the device gamut. These points have to be clipped back into the gamut to be visually presentable. No matter what clipping methods are used, including our optimal hue-preserving clipping method, loss of chrominance fidelity is inevitable. In this chapter we address the root of the problem, and develop a more principled algorithmic approach for chrominance-preserving image enhancement. The idea is to reduce the chances of out-gamut occurrences by adaptively setting upper bounds of luminance adjustment range in the process of contrast stretching. The proposed adaptive

gamut-aware luminance adjustment and the hue-preserving clipping are integrated into an algorithmic framework of constrained optimization for image enhancement with bounded chrominance errors, formulated as

$$\max_{T} G(T) - \lambda_t D_t(T) - \lambda_c D_c(T) \tag{3.1}$$

The above objective function consists of three reward/penalty terms. The first term $G(T)$ is the luminance contrast gain made by the colour mapping function T; the second term $D_t(T)$ is the distortion in tonal reproduction by T; the third term $D_c(T)$ is the distortion in chrominance caused by T. This optimization framework allows us to boost the conspicuity of image details while maintaining naturalness of enhanced images. We develop, in this chapter, an efficient, dynamic programming algorithm to solve (3.1), which can be proven to be even faster than existing OCTM-type algorithms. Experiment results validate the effectiveness of the proposed algorithm as to meet its design goal. It can restore severely underexposed/overexposed images without noticeable chrominance distortions and other artifacts that plague the existing methods.

The rest of the chapter is organized as follows. We begin with elaborating the formulation of the optimization problem for enhancement in Sec. 3.2. In Sec. 3.3, we present a dynamic programming based algorithm that solves the optimization problem efficiently. In Sec. 3.4 we introduce a technique to perform fast transform between RGB and LHS colour spaces, which can further accelerate the algorithm. Next in Sec. 3.5 we discuss the impact of each user-specified parameter in (3.1) to output images. Finally in Sec. 3.6 we present the experimental results and make a comparison between the proposed method and several mainstream enhancement

methods in terms of both objective and subjective image quality.

3.2 Algorithmic Framework for Enhancement

In [72], Wu argues that tone continuity is a key element for image enhancement approaches other than high global contrast. Those methods with poor tone continuity are likely to suffer from tone distortion artifacts that wash out the gradients and textures on object surfaces. The author proposes the OCTM algorithm that views contrast enhancement as an optimization problem that maximizes output global contrast while confining tone distortion, which is formulated as

$$\max_{T} G(\mathbf{p}, T) - \lambda_t D_t(\mathbf{p}, T) \tag{3.2}$$

where $G(\mathbf{p}, T)$ is a metric of the contrast gain made by tone mapping function T over the initial histogram $\mathbf{p} = (p_0, \cdots, p_{K-1})$ pertaining to K existing intensity levels $\{0, \cdots, K-1\}$, and $D_t(\mathbf{p}, T)$ is a measure of tone distortion caused by T weighted by Lagrangian coefficient λ_t.

Operating only on the luminance component, the algorithm suffers the same problem of loss of colour fidelity as other enhancement techniques when dealing with colour images. But, based on the discussions in Sec. 2.4, we can refine the framework by adding a chrominance distortion constraint term:

$$\max_{T} G(\mathbf{p}, T) - \lambda_t D_t(\mathbf{p}, T) - \lambda_c D_c(\mathbf{p}, T) \tag{3.3}$$

where $D_c(\cdot)$ denotes the chrominance distortions caused by T.

Note that, because image pixel values are represented using integer numbers in practice, we assume the mapping function $T(\cdot)$ to be:

$$T : \{0, 1, ..., M-1\} \rightarrow \{0, 1, ..., N-1\}, \tag{3.4}$$

given an input image with M intensity levels as well as an output image supporting N intensity levels. Either set of intensity levels is uniformly quantized from luminance value L as defined in (2.4). In addition, T should be monotonically nondecreasing so as not to revert the order of input intensity levels.

Furthermore we write T in the following form:

$$T(i) = \sum_{0 \leq j \leq i} s_j, \ i = 0, 1, \ldots, M-1, \tag{3.5}$$

introducing

$$s_j \triangleq \begin{cases} T(j) - T(j-1) & j = 1, \ldots, M-1 \\ T(0) & j = 0 \end{cases} \tag{3.6}$$

which corresponds to the increment in output luminance versus one unit step up in input luminance. Step size s_j essentially represents the gradient of T at luminance j: when $s_j > 1$, the dynamic range around j is stretched leading to larger contrast, and oppositely $s_j < 1$ compresses neighbouring dynamic range and thus decrease the contrast.

It follows from (3.5) and (3.6) that there is a one-to-one mapping between T and $\{s_j\}$. For ease of algorithm development, we prefer to use $\mathbf{s} \triangleq [s_0, s_1, \cdots, s_{M-1}]$, called *tone mapping vector*, in sequel. In the following subsections we detail the formulation of the optimization problem (3.3) by elucidating the concrete definitions of $G(\cdot)$, $D_t(\cdot)$

and $D_c(\cdot)$, with respect to $\mathbf{s}$.

3.2.1 Global Contrast Gain Metric

We state in Sec. 1.3 that sufficient contrast in edges and local patterns of object surfaces plays an important role in human cognition. Edges and object surfaces with low contrast are hardly legible, hence perceived as poor image quality. Statistically, areas of low contrast in an image tend to have rather narrow dynamic ranges, where the differences between adjacent pixels are much too small. In order to expand the narrow dynamic ranges, we need to design the tone mapping function T, or the corresponding tone mapping vector $\mathbf{s}$, according to the scene. Recall that the luminance dimension in the LHS colour space is perceptually uniform, so intensity level increment s_j is linear in perceived contrast gain of luminance level j, *i.e.* how much the corresponding dynamic range gets amplified by T. For example, given a low contrast object in a scene whose luminance is around j, we can let $s_j > 1$ so as to boost the visibility of the object. The mapping also affects pixels that have luminance j yet do not belong to the object. But in natural images, edge pixels are much fewer than those of low contrast [59]; hence we neglect the side effects of T on high contrast pixels. In fact, no visual artifacts have ever been found on those high contrast pixels in our experiments.

It should be noted that the adjustment of $\mathbf{s}$ is not simply making every element as large as possible, as one has to respect the finite number of output intensity levels. To take full advantage of the output levels, a natural idea is to grant larger dynamic range to more populated luminance levels prior to less populated ones, as pixels of more populated luminance levels typically have larger impact to the image quality.

Therefore, we define the global contrast gain metric $G(\mathbf{s})$ by

$$G(\mathbf{p}, \mathbf{s}) = \sum_{j=0}^{M-1} p_j s_j \tag{3.7}$$

so that the optimization result tends to have larger s values on more populated luminance levels.

3.2.2 Tone Distortion Metric

Tone distortions occur when a mapping function compresses dynamic range by mapping multiple consecutive intensity levels into one, during which the tonal details in the compressed levels are totally lost. Such distortions are typically inevitable especially when $N \leq M$ due to limited number of output intensity levels available.

The OCTM framework defines the tone distortion, $D(\mathbf{s})$, by

$$D(\mathbf{s}) = \max_{0 \leq j < k < M} \{k - j | T(l_j) = T(l_k)\}, \tag{3.8}$$

which can be interpreted as the maximal number of adjacent gray levels that map into a single one. The larger $D(\cdot)$ is, the greater distortion there exists in tone continuity. OCTM then imposes an upper bound on $D(\mathbf{s})$ with a linear inequality constraint:

$$\begin{aligned} &\max_{T} \ G(\mathbf{p}, \mathbf{s}), \\ &s.t. \quad \sum_{n \leq j \leq n+d} s_j \geq 1, 0 \leq n < M - d, \end{aligned} \tag{3.9}$$

so that s_j cannot be zero for more than d consecutive input gray levels. In other words, d functions as an upper limit of $D(\mathbf{s})$.

However, the definition (3.8) has a drawback that it is simply a constraint on $\mathbf{s}$ without consideration of the statistics of the image. As a result it overlooks an important fact that compressing (the dynamic range of) less populated gray levels causes smaller tone distortion than compressing (that of) higher populated ones.

Taking the fact into consideration, we develop a new histogram-based tone distortion metric $D_t(\mathbf{p}, \mathbf{s})$:

$$D_t(\mathbf{p}, \mathbf{s}) = \sum_{j=0}^{M-1} p_j \delta(s_j), \tag{3.10}$$

where $\delta(\cdot)$ is discrete Dirac function taking 1 at $s_j = 0$ and 0 elsewhere. It essentially corresponds to the proportion of pixels that get distorted in tone by the mapping function.

3.2.3 Chrominance Distortion Metric

As discussed in Sec. 2.5, chrominance distortions are caused by out-of-gamut problems where a colour point $\mathbf{c}$ is brightened to a luminance beyond the corresponding maximal adjustable luminance $\eta(\mathbf{c})$. The chrominance distortion term $\lambda_c D_c(\cdot)$ in (3.3) aims to penalize such over-adjustment on luminance.

Note that maximal adjustable luminance is not consistent among the pixels of the same luminance. Accordingly, we define

$$\eta_{min}(l) = \min_{\mathbf{c}} \{\eta(\mathbf{c}) | \mathbf{c} \in I, L(\mathbf{c}) = l\}, \tag{3.11}$$

which is the maximal adjustable luminance for all the colours $\mathbf{c}$ in the image I of a given luminance value l. Ideally, the mapping function T satisfying $T(l) \leq \eta_{min}(l)$ should be free of chrominance error. However, the room for luminance adjustment

may be too small to achieve high contrast under such circumstance, especially for bluish colours. Take pure blue colour $(0, 0, 1)$ as example: its luminance is merely 0.323 that appears rather dark for human vision. But because $\mathbf{c}$ is on the edge of RGB gamut, its luminance cannot be increased any more. The blue colour also blocks other colours with luminance 0.323 from being increased if the constraint $T(l) \leq \eta_{min}(l)$ is strictly imposed on the image.

To overcome the limitation of the excessively strong constraint $T(l) \leq \eta_{min}(l)$, we relax the maximal adjustable luminance with respect to equi-luminance pixels, tolerating chrominance distortions to some extent. Consequently, we adopt the average maximal adjustable luminance of equi-luminance pixels instead of the minimum:

$$\bar{\eta}(l) = E_{\mathbf{c}|\mathbf{c}\in I, L(\mathbf{c})=l}\left[\eta(\mathbf{c})\right] \tag{3.12}$$

And then we define the chrominance distortion metric using $\bar{\eta}(l)$:

$$D_c(\mathbf{p}, T) = \sum_{j=0}^{M-1} p_j \frac{ReLU\left(T(j) - \bar{\eta}(j)\right)}{M} \tag{3.13}$$

penalizing any luminance level j violating $T(j) \leq \bar{\eta}(j)$. Here $ReLU(x) \triangleq \max\{0, x\}$ is rectified linear unit (ReLU) function.

3.2.4 Optimization Objective

Apart from constraining tone and chrominance distortions, we enforce an upper bound u on each element of $\mathbf{s}$ to avoid excessive boost on contrast as otherwise the tone mapping may result in a binary image with watershed at the most populated luminance

level.

To sum up, the optimization problem (3.3) is formulated as:

$$\begin{aligned} \max_{\mathbf{s}\in\mathbb{N}^M} \quad & \sum_{j=0}^{M-1} p_j \left[s_j - \lambda_t \delta(s_j) - \lambda_c \frac{ReLU\left(T(j) - \bar{\eta}(j)\right)}{M} \right] \\ s.t. \quad & \sum_{j=0}^{M-1} s_j = N - 1 \\ & \mathbf{s} \preceq u \end{aligned} \tag{3.14}$$

3.3 Dynamic Programming Solution to the Optimization Problem (3.14)

The optimization problems in OCTM [72] as well as its variants [44, 59] are linear programming (LP). They can be easily solved by LP solvers. On the other hand, in the proposed optimization problem (3.14), the objective function and constraints are nonlinear; solving it requires more sophisticated algorithm techniques. In this section we develop an efficient scheme to solve the optimization problem (3.14) based on dynamic programming (DP).

The problem (3.14) can be divided into a set of subproblems formulated as

$$\begin{aligned} \max_{\mathbf{s}\in\mathbb{N}^\mu} \quad & \sum_{j=0}^{\mu-1} p_j \left[s_j - \lambda_t \delta(s_j) - \lambda_c \frac{ReLU(T(j) - \bar{\eta}(j))}{M} \right] \\ s.t. \quad & \sum_{j=0}^{\mu-1} s_j = \nu - 1 \\ & \mathbf{s} \preceq u \end{aligned} \tag{3.15}$$

We call the optimization problem above subproblem (μ, ν), and denote the corresponding optimal solution and objective function value by $\mathbf{s}^{(\mu,\nu)}$ and $F(\mathbf{s}^{(\mu,\nu)})$, respectively:

$$F(\mathbf{s}^{(\mu,\nu)}) = \sum_{j=0}^{\mu-1} p_j \left[s_j^{(\mu,\nu)} - \lambda_t \delta(s_j^{(\mu,\nu)}) - \lambda_c \frac{ReLU(T^{(\mu,\nu)}(j) - \bar{\eta}(j))}{M} \right], \tag{3.16}$$

$T^{(\mu,\nu)}$ being the tone mapping function corresponding to $\mathbf{s}^{(\mu,\nu)}$ using (3.5). Furthermore, let $\mathbf{s}^*$ be the optimal solution to (3.14) with the objective function denoted by $F(\mathbf{s}^*)$. Obviously, subproblem (M, N) is equivalent to (3.14), *i.e.* $\mathbf{s}^* = \mathbf{s}^{(M,N)}$.

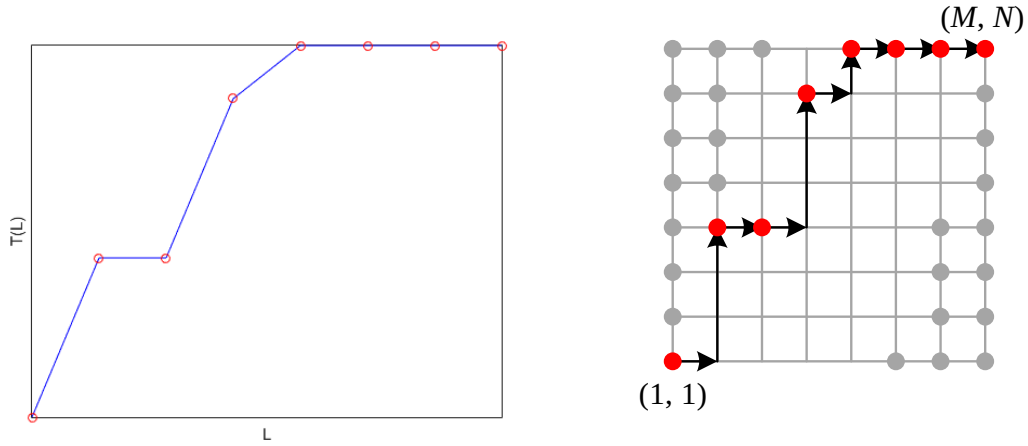
It can be shown that $F(\mathbf{s}^{(\mu,\nu)})$ has the following recursive form

$$\begin{aligned} F(\mathbf{s}^{(\mu,\nu)}) = & \max_{\substack{i=0,\dots,\mu-1 \\ j=0,\dots,u}} F(\mathbf{s}^{(\mu-i,\nu-j)}) + p_{\mu-1} j - \sum_{k=i}^{\mu-2} p_k \left[\lambda_t + \lambda_c \frac{ReLU(j - \bar{\eta}(k))}{M} \right] \\ & - \lambda_c p_{\mu-1} ReLU(\nu - \bar{\eta}(\mu-1)) \end{aligned} \tag{3.17}$$

Therefore, subproblem (μ, ν) can be solved via bottom-up recursion: given the optimal solutions of all subproblems $\mathbf{s}^{(\mu-i,\nu-j)}$, $i = 0, \cdots, \mu-1$, $j = 0, \cdots, u$, we can find the optimal i^* and j^* that maximize $F(\mathbf{s}^{(\mu,\nu)})$, and $\mathbf{s}(\mu, \nu)$ is formed by $\mathbf{s}^{(\mu-i,\nu-j)}$ appended by $(i^* - 1)$ zeros and one j^* at the end:

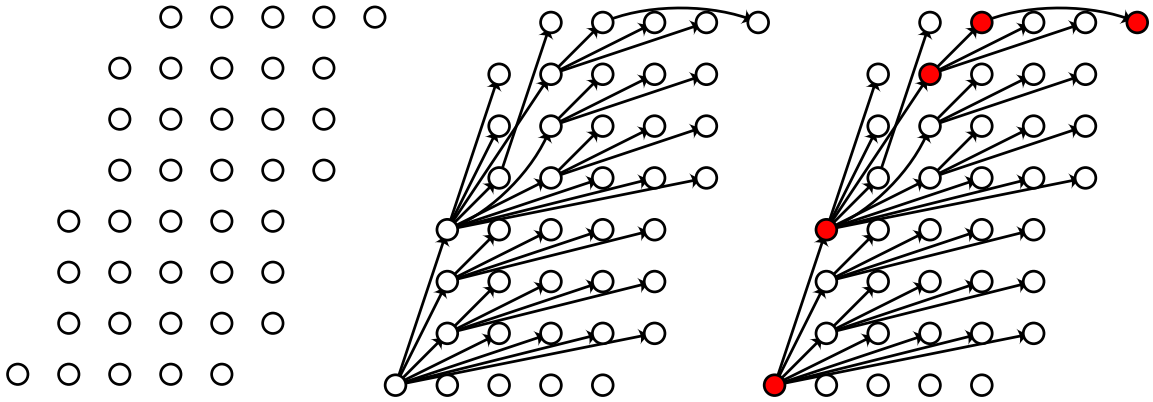
$$\mathbf{s}^{(\mu,\nu)} = \left[\mathbf{s}^{(\mu-i^*,\nu-j^*)}, \underbrace{0 \dots 0}_{i^*-1}, j^* \right] \tag{3.18}$$

Starting from $\mathbf{s}^{(1,1)} = [0]$, we can progressively obtain the final solution $\mathbf{s}(M, N)$ following the idea. As shown in Fig. 3.1, we consider the problem as seeking for an optimal path on an $M \times N$ chessboard from the bottom-left point $(1, 1)$ to the top-right point (M, N), along which $F(\mathbf{s})$ takes the maximum. We can know the



(a) (b)

Figure 3.1: A sample tone mapping function (a) together with its corresponding "path on the chessboard" (b), represented by black arrows.



(a) (b) (c)

Figure 3.2: The directed graph $g = \{V, E\}$ of Fig. 3.1. (a) The vertex set V; (b) the graph formed by the forward pass; (c) the key nodes (marked as red) given by the backward pass, finally leading to the optimal path.

following facts regarding the optimal path from the discussions above:

- The path must start from $(1,1)$. $F(\mathbf{s}^{(1,1)}) = 0$;
- Some points on the chessboard can never be passed by the path due to the constraint $\mathbf{s} \preceq u$, hence can be simply excluded (corresponding to the points marked by gray in Fig. 3.1(b));
- The path consists of $M-1$ steps: in each step i $(i = 1, \cdots, M-1)$ the piece goes one unit right and s_i units up. The points after each step (the red points in Fig. 3.1 (a)(b)) together with the starting point $(1,1)$ then determine the mapping function T.
- According to (3.17), any subproblem (μ, ν) cannot be solved until we have the solutions to subproblems $(x, y), x < \mu, y < \nu$ before hand; ultimately, subproblem (M, N) has to be solved after we have solutions to all the other subproblems.
- Also, (3.17) associates a point (μ, ν) with the solution of subproblem (μ, ν), corresponding to point $(\mu - i^*, \nu - j^*)$ on the chessboard. [1]

Following the rules above, we define a directed graph $g = \{V, E\}$ corresponding to the chessboard: the vertex set V is the set of candidate points of the path (namely those not marked by gray in Fig. 3.1(b)), and the edge set E comprises the connections that end at each candidate point (μ, ν) and start from the node corresponding to the optimal solution of subproblem (μ, ν), at $(\mu - i^*, \nu - j^*)$. Because each node of V has a unique parent node (except the starting point $(1,1)$ which is without parent), there exists a path solely connecting $(1,1)$ and (M, N).

[1] We assume that each subproblem has a unique solution $\{i = i^*, j = j^*\}$. If it is not the case in practice, we adopt the solution with the smallest i and j.

We operate a two-pass scheme to determine the optimal path, including one forward pass and a subsequent second backward pass. The forward pass aims to construct the edge set E in the directed graph by traversing each node and having each subproblem solved. Specifically, we let the input luminance μ iterate from 2 to M, and for each μ we solve all subproblems (μ, ν), ν iterating from the smallest to the largest. The backward pass then backtracks from (M, N) and keeps visiting the parent node until reaching $(1, 1)$. We call the nodes visited in the backward pass *key nodes* as they determined the entire optimal path, which can be obtained by filling the remaining nodes using (3.18). The two-pass scheme is graphically illustrated in Fig. 3.2.

3.3.1 Time and Memory Efficiency

We briefly discuss the time and memory cost of the DP-based solver. In the forward pass, we need to solve $O(M \times N)$ subproblems; for each subproblem (μ, ν), the solver has to find the maximum value of at most $O(\mu \times u)$ candidates. Thus the forward pass has $O(uM^2N)$ complexity. The backward pass takes $O(M)$ computations to determine the optimal path because the path is always of M length with no more than M key nodes. In total, the solver is of $O(uM^2N)$ time complexity. Note that it is independent of image size. The memory complexity is $O(MN)$, as the solver maintains a directed graph with $O(MN)$ nodes, each node having one ingoing edge and one outgoing edge respectively.

In the common case where $M = N = 256$, uM^2N is typically far greater than the number of image pixels. Based on the observation that the optimal i^* of (3.17) typically takes small values, we adopt a technique to reduce time complexity by

shrinking the searching range $i = 0, \cdots, \mu-1$ in (3.17) to $i = \max\{0, \mu-d\}, \cdots, \mu-1$. Such shrinkage essentially imposes another tone distortion constraint that is the same as the one in (3.9):

$$\sum_{n \le j \le n+d} s_j \ge 1, 0 \le n < M - d, \tag{3.19}$$

i.e., s_j cannot be zero for d contiguous intensity levels. As discussed in Sec. 3.2.2, (3.19) is a weaker constraint than $D_t(\cdot)$ defined by (3.10) (unless d is rather small). Applying (3.19) can significantly speed up the algorithm by reducing the time complexity from $O(uM^2N)$ to $O(udMN)$, without affecting the optimization result.

We time the OCTM method and the proposed method on a set of 10 underexposed colour[2] images. The tested algorithms include: 1) OCTM with LP solver; 2) OCTM with the DP solver presented in the thesis, which is of $O(udMN)$ time complexity; 3) the proposed method with the DP solver without the constraint (3.19); 4) the proposed method with the DP solver with the constraint (3.19), d being 10. All the algorithms are implemented in C++[3] and run on Windows® 7 with a single core of Intel® Core-i7 CPU. The running times of the above algorithms are listed in Table 3.3.1.

Table 3.1: Consumed time of solving the optimization problem of OCTM and the proposed method (in milliseconds).

OCTM (LP)	OCTM (DP)	Proposed method (DP, without d)	Proposed method (DP, $d = 10$)
31.68	25.18	78.98	**6.78**

The timing results demonstrate the higher efficiency of the proposed DP algorithm.

[2]All test images are 8-bit *RGB* images where each colour channel has 256 intensity levels, *i.e.* $M = N = 256$.

[3]The LP solver is from the open source COIN-OR CLP library.

Remarkably, by introducing a weak constraint (3.19), the proposed algorithm can achieve an even shorter running time than OCTM, although dealing with a more complicated optimization problem.

3.4 Fast Colour Space Transform

As depicted in Fig. 2.2, the proposed method needs to perform colour space transforms between RGB and the proposed LHS. Here we introduce a technique that realizes more efficient colour space conversion than applying the formulae (such as (2.4)(2.5)(2.6)) pixelwise.

Suppose the luminance mapping function maps the luminance of a given colour vector $\mathbf{c} = (r, g, b)$, denoted by l, to l'. We can then derive the corresponding output colour vector $\mathbf{c}_A$ by optimal clipping.

If $l' \leq \eta(\mathbf{c})$, there is no out-of-gamut issue. We have

$$\mathbf{c}_A = \frac{y'}{y}\mathbf{c}. \tag{3.20}$$

where

$$y = f^{-1}(l), y' = f^{-1}(l'). \tag{3.21}$$

If $l' > \eta(\mathbf{c})$,we have to perform clipping, as illustrated in Fig. 2.10. Let $y_\eta = f^{-1}(\eta(\mathbf{c}))$. Note that $\mathbf{I} = (1, 1, 1)$, $\mathbf{c}_A$ and $\mathbf{c}_V = \frac{y_\eta}{y}\mathbf{c}$ are collinear, satisfying

$$\frac{\mathbf{c}_A - \mathbf{c}_V}{\mathbf{I} - \mathbf{c}_V} = \frac{y' - y_\eta}{1 - y_\eta} \tag{3.22}$$

$$\Leftrightarrow \mathbf{c}_A = \frac{y' - y_\eta}{1 - y_\eta}\mathbf{I} + \frac{1 - y'}{1 - y_\eta}\mathbf{c}_V. \tag{3.23}$$

Therefore the output colour vector $\mathbf{c}_A$ can be computed without performing explicit colour space transform from LHS to RGB, as long as l and l' are given:

$$\mathbf{c}_A = \begin{cases} \dfrac{y'}{y}\mathbf{c} & \dfrac{y'}{y}\max\{r, g, b\} \leq 1 \\ \dfrac{y' - y_\eta}{1 - y_\eta}\mathbf{I} + \dfrac{(1 - y')y_\eta}{(1 - y_\eta)y}\mathbf{c} & \text{otherwise} \end{cases} \tag{3.24}$$

Finally, the proposed enhancement algorithm in pseudocode is presented in Algorithm 1.

3.5 Parameter Tuning

The proposed algorithm takes several user-specified arguments including M, N, u, λ_t and λ_c. In this section we discuss their impact on the output, and the optimal values assigned to them.

3.5.1 u, Upper Bound of s

As mentioned in (3.6), s_j reflects the contrast gain at input luminance level j. Hence its upper bound u positively correlates with output global contrast gain (see the examples in Fig. 3.3). Fig. 3.4 reveals that the resulting tone mapping function gets sharper especially around the highly populated levels as u increases. A much too large (overexaggerated) u causes severe overexposure patterns as exampled in the $u = 10$ case; therefore u should be carefully adjusted to balance between contrast gain and the distortions.

input : Input image I, λ_t, λ_c, u, M, N, (d)
output: Output image I'
begin
  Compute the luminance map $\mathcal{L}$ pixelwise using (2.4) and the luminance histogram $\mathbf{p}$;
  Compute $\bar{\eta}(l)$;
  Initialize T;
  // Initialize directed graph $g = \{V, E\}$
  $\mathrm{V} \leftarrow \{(x, y) | x = 1, \cdots, M, y = 1, \cdots, N\}$;
  // Forward pass
  for $x \leftarrow 2$ **to** M **do**
    $\mathsf{y\_min} \leftarrow \max\{1, N - u(M - x) - 1\}$;
    $\mathsf{y\_max} \leftarrow \min\{\bar{\eta}(x - 1), u(x - 2) + 1\}$; // the two lines exclude gray nodes in Fig.3.1(b)
    for $y \leftarrow \mathsf{y\_min}$ **to** $\mathsf{y\_max}$ **do**
      // Solve subproblem (x, y) using (3.17)
      // Can be speeded up by (3.19), given d
      $(i^*, j^*) \leftarrow \arg\max_{i,j} F(\mathbf{s}^{(x,y)})$;
      $\mathrm{V}(x, y).\mathsf{parent} \leftarrow \mathrm{V}(x - i^*, y - j^*)$;
    end
  end
  // Backward Pass
  $\mathsf{P} \leftarrow (M, N)$;
  $T(M) \leftarrow N$;
  while $\mathsf{P} \neq (1, 1)$ **do**
    for $i \leftarrow \mathsf{P.parent.x}$ **to** $\mathsf{P.x} - 1$ **do**
      $T(i - 1) \leftarrow \mathsf{P.parent.y} - 1$;
    end
    $\mathsf{P} \leftarrow \mathsf{P.parent}$;
  end
  // Perform luminance mapping together with inverse colour transform
  foreach $(x, y) \in I$ **do**
    $\mathcal{L}'(x, y) \leftarrow T(\mathcal{L}(x, y))$;
    $I'(x, y) \leftarrow$ result of (3.24) on pixel (x, y);
  end
  return I';
end

Algorithm 1: Proposed enhancement method (in pseudocode).

It is common that low contrast images occupy merely a portion of available luminance levels like the histogram in Fig. 3.4. An ideal mapping function should then expand them to all output luminance levels so as to gain maximal dynamic range. In this sense, we suggest $\{s_j\}$ have an average value close to $N/N_D(I)$, with $N_D(I)$ being the number of luminance levels occupied by image I. In practice, we estimate $N_D(I)$ by the number of levels whose frequency is no smaller than $1/M$ (thus excluding outlier levels with rather small population as well as those unpopulated). Finally, we let

$$u = \left\lceil \frac{N}{N_D(I)} \right\rceil, \tag{3.25}$$

taking 3 for the example image in Fig. 3.3.

3.5.2 Tone Distortion Coefficient λ_t

The tone mapping functions in Fig. 3.4 are more or less with flat segments where multiple contiguous luminance levels are mapped into one, resulting in tone distortion. Hence we introduce the tone distortion term (3.10) weighed by a coefficient λ_t.

Fig. 3.5 and 3.6 reveal the impact of λ_t with other parameters kept fixed: as it increases, more local details can be conserved (such as the clouds and water waves) and the mapping function appears to be smoother while the global contrast is slightly sacrificed. Empirically we assign 0.5 to it in our application.

3.5.3 Chrominance Distortion Coefficient λ_c

The enhancement results in Fig. 3.3 and 3.5 all suffer from some extent of overexposure in the sky with unnatural cyan colour. We show in Fig. 3.7 that it can be effectively constrained by the chrominance distortion term that forces the tone mapping function

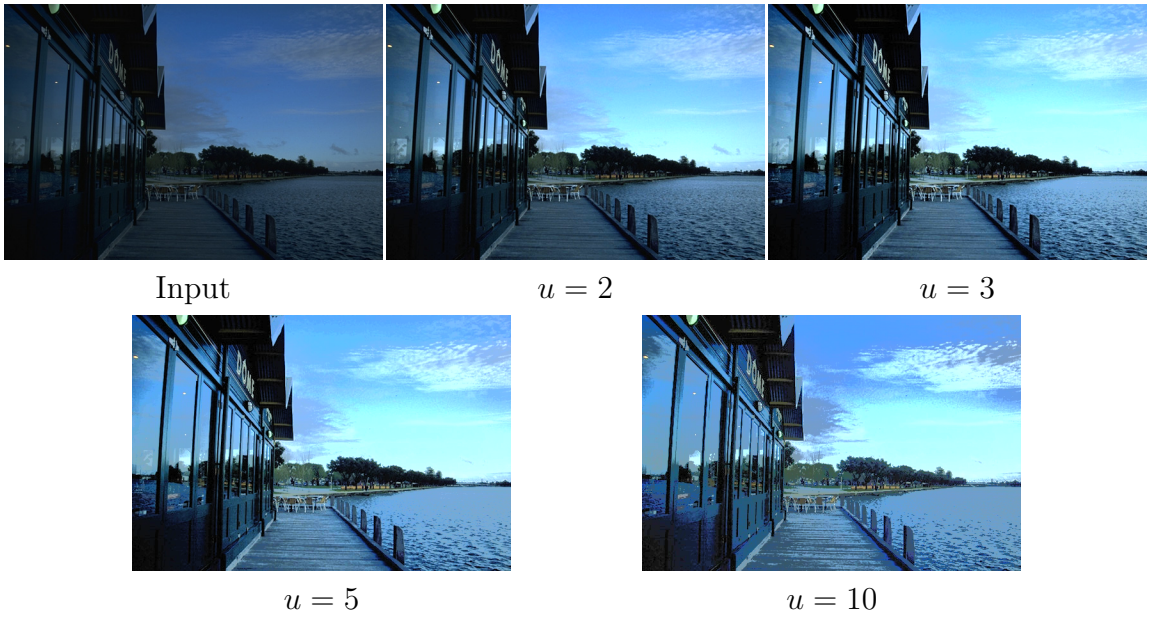


Figure 3.3: Results of the proposed method with different u values ($\lambda_t = \lambda_c = 0$).

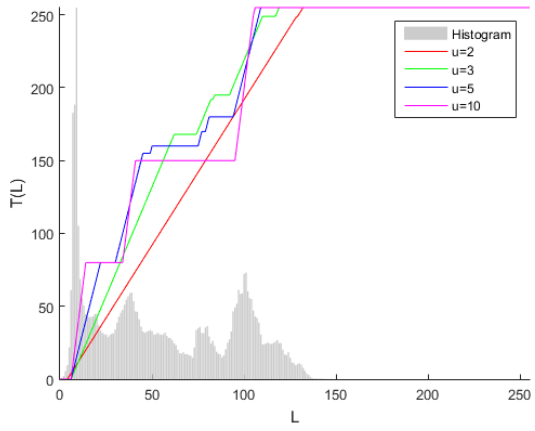


Figure 3.4: Tone mapping functions given by the proposed method under different u values.

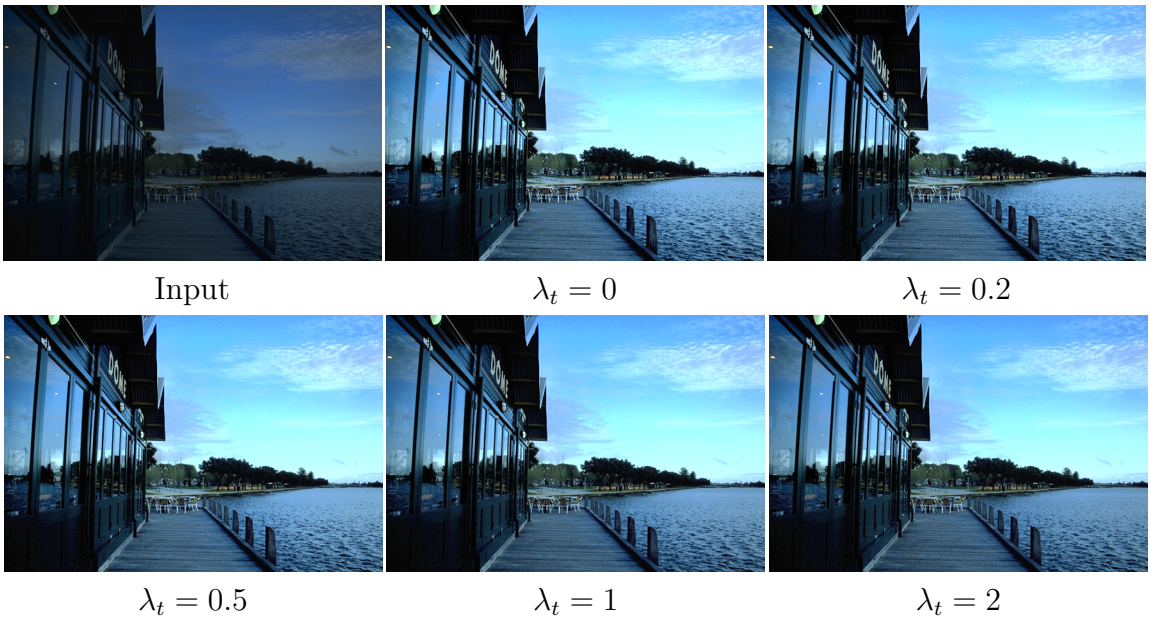


Figure 3.5: Results of the proposed method with different λ_t values ($u = 3, \lambda_c = 0$).

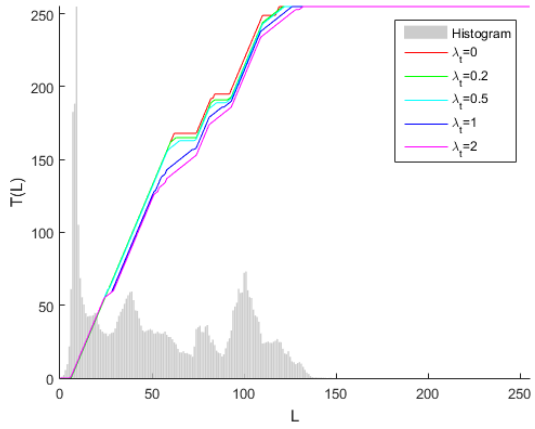


Figure 3.6: Tone mapping functions given by the proposed method under different λ_t values.

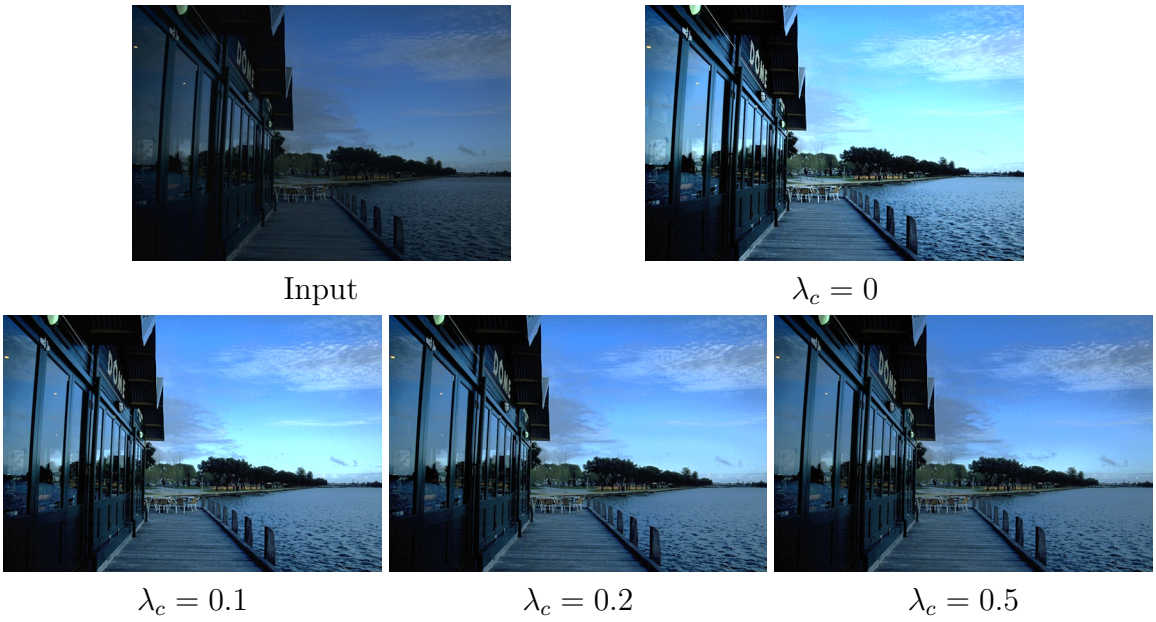


Figure 3.7: Results of the proposed method with different λ_c values ($u = 3, \lambda_t = 0.5$).

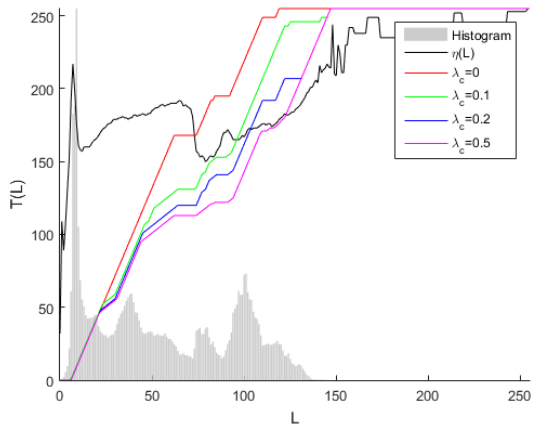


Figure 3.8: Tone mapping functions given by the proposed method under different λ_c values.

to be below $\bar{\eta}(L)$ (the black curve in Fig. 3.8) with a strength specified by coefficient λ_c. We adopt $\lambda_c = 0.2$ in our application.

3.5.4 Number of input luminance levels M and output luminance levels N

Recall Sec. 2.3.1 that luminance L of LHS space is a continuous value lying in $[0, 1]$. We must quantize it into M separate levels before running the proposed algorithm (as (3.5) does), and map the N-leveled algorithm output back to $[0, 1]$. We set M and N according to the number of intensity levels of the input image; as images are most commonly stored in a bitmap format where each channel of a pixel occupies 8 bits, we adopt $M = N = 256$ generally. Users can also apply greater M and N values for less quantization error or under particular conditions such as processing HDR images with even more intensity levels.

3.6 Experimental Results

In this section the proposed enhancement algorithm is compared with some mainstream context-free contrast enhancement approaches including classical HE algorithm, modified HE by Arici *et al.* [3], multiscale Retinex (MSR) algorithm [35] and OCTM algorithm [72]. The comparison study is carried out on a set of illy acquired colour images of poor contrast covering a diverse range of scenes. The performance of the algorithms is evaluated in both objective (in terms of image fidelity metrics) and subjective (in terms of perceptual quality) aspects. Note that, the latter five algorithms perform enhancement on the luminance channel with insufficient (or even

little) attention to the tone and chrominance distortion issue which is addressed in the thesis.

3.6.1 Objective Evaluation

We measure the enhanced images by the following four metrics:

- **Global Contrast Gain**. Most low contrast images cover merely a subset of the M available intensity levels. We denote the subset by $\mathbf{l} = \{l_0, \cdots, l_{K-1}\}, K \leq M$ and the corresponding histogram by $\tilde{\mathbf{p}} = \{\tilde{p}_0, \cdots, \tilde{p}_{K-1}\}$. $\tilde{\mathbf{p}}$ can de deemed as the original histogram $\mathbf{p} = \{p_0, \cdots, p_{M-1}\}$ with zero entries excluded. In [72], the author proposes a context-free contrast metric C with respect to $\tilde{\mathbf{p}}$:

$$C(\mathbf{p}) = p_0(l_1 - l_0) + \sum_{k=1}^{K-1} p_k(l_k - l_{k-1}) \tag{3.26}$$

  which takes maximum $K-1$ for binary black-and-white images and minimum 0 for constant images. Image enhancement is to amplify the difference between two adjacent intensity levels in $\mathbf{l}$ in order to expand dynamic range, increasing contrast $C(\tilde{\mathbf{p}})$. Then we define global contrast gain by $C(\tilde{\mathbf{p}}')/C(\tilde{\mathbf{p}})$, $\tilde{\mathbf{p}}'$ being the histogram $\tilde{\mathbf{p}}$ after tone mapping.

- **Histogram Entropy Gain**. Performing enhancement on low contrast images tends to stretch a histogram broadening luminance distribution, resulting in a rise in histogram entropy $H(\mathbf{p}) = -\sum_j p_j \log(p_j)$ [28].

  To measure gain in entropy, we uniformly quantize the luminance range $[0, 1]$ into 16 levels and then obtain the 16-quantized histograms for both input and

Table 3.2: Performance comparison in contrast and colour metrics.

Metric	HE	Arici *et al.*	MSR	OCTM	Proposed
Contrast gain	1.997	1.374	1.211	1.722	1.425
Histogram entropy	1.148	1.074	1.045	1.079	1.099
AHD (degree)	6.795	3.209	5.880	0.008	0.488
ASD	0.054	0.011	0.019	0.069	0.004

output image (denoted by $\mathbf{p}_n$ and $\mathbf{p}'_n$) and compute the entropy values $H(\mathbf{p}_n)$ and $H(\mathbf{p}'_n)$ accordingly. Histogram entropy gain is given by $H(\mathbf{p}'_n)/H(\mathbf{p}_n)$.

- **Average Hue Distortion (AHD)** and **Average Saturation Distortion (ASD)**, defined by the average difference in hue and saturation over each pixel. We exclude pixels with luminance under 0.15 or over 0.85 owing to their low precision[4].

These quality metrics for the five enhancement algorithms are reported in Table 3.6.1; the values in the table are averages over all test images. As shown HE makes the highest contrast gain and entropy than any other method, but it suffers from much larger errors both in hue and saturation meanwhile. The method [3] and MSR also enhance images with a price of chrominance error. As the original OCTM and the proposed algorithm keep colour vectors on the equihue triangle when manipulating the luminance (recall from Fig. 2.10), they should have zero error in hue; the very small errors in hue for these two methods are caused by quantizing the real luminance and chrominance values to integer R, G, B coordinates. Remarkable is that the average saturation error of the proposed algorithm is made orders of magnitude smaller than that of the original OCTM algorithm and other subject methods, without affecting

[4]Hue and saturation have relatively low precision for very dark or bright colours. For example, RGB colours $(0.01, 0, 0)$, $(0, 0, 0.01)$, $(0.01, 0.01, 0.01)$ look almost the same for human observers, but differ greatly in hue and saturation.

the performance on contrast enhancement.

3.6.2 Subjective Evaluation

To assess the perceptual image qualities of the tested enhancement algorithms, four test images and the corresponding enhancement results produced by the five algorithms are presented in Fig.3.9. The output images of HE appear overexposed and have large chrominance errors; for example, HE distorts hue severely in Case 2, 3, 4 (especially shift from blue to cyan), which is unacceptable to most users. The same problem exists in the method by Arici *et al.* and MSR as well: The former greatly desaturates the input in Case 3 and 5 and MSR on the other hand produces oversaturated colours together with halo artifacts. Although OCTM algorithm retains the original hue, which is expected based on the analysis of the thesis, it suffers from serious saturation distortions as well. The colours generated by OCTM algorithm look rather pale in Case 1, 2 and 4, due to much reduced saturation as the side effect of luminance stretching. In comparison, the output images of the proposed algorithm appear visually more pleasing, with both enhanced details and faithful colour reproduction at the same time. These observations are in agreement with the results in the objective colour metrics reported in Table 1. In addition, Case 5 demonstrates the superior performance of the proposed method over its competitors on overexposed images with the most proper adjustment in luminance channel as well as the most natural colour appearance.

CASE 1

Input

HE

Arici *et al.*

MSR

OCTM

Proposed

CASE 2

Input

HE

Arici *et al.*

MSR

OCTM

Proposed

CASE 3

Input

HE

Arici *et al.*

MSR

OCTM

Proposed

CASE 4

Input

HE

Arici *et al.*

MSR

OCTM

Proposed

CASE 5

Figure 3.9: Results of enhancement algorithms on poor contrast images. The presented algorithms include: HE, modified HE by Arici *et al.* [3], multiscale Retinex algorithm, OCTM [72] and the proposed method.

3.7 Conclusion

In Chapter 2 and 3, we address the long-time overlooked problem of colour fidelity preservation in image enhancement algorithms. We discuss the means to retain colour fidelity in several aspects, including: construction of the colour space that best suits for enhancement, optimal solution to out-of-gamut problems and the optimization-based framework for enhancement that maximizes output global contrast while containing tone and chrominance distortion. Based on them, we propose a novel algorithm solving the optimization problem efficiently by dynamic programming. Finally, our objective and subjective evaluations indicate that the proposed enhancement algorithm meets its design goal and is largely free of the chrominance errors that plague the existing enhancement methods without sacrificing output contrast.

Chapter 4

Learning-based Restoration of Backlit Images

4.1 Introduction

For many users of mass-marketed consumer cameras, the most annoying type of image quality degradation is, arguably, that of backlighting, which is characterized by under- and/or over-exposed object surfaces in patchy appearances of abrupt intensity changes. Illuminating objects from behind tends to cause incorrect, extremely varied exposures of different regions in an image, with backlit surfaces being severely underexposed and possibly at the same time front-lit surfaces being overexposed, resulting in loss of details and poor contrast. Although backlighting may render surface details illegible and it is a commonly encountered ill shooting condition that is beyond photographers' control and defies cameras' exposure settings, scanty research has been done on the restoration of backlit images. The vast majority of published papers on image restoration were devoted to denoising and superresolution, despite the fact

that noise corruption and resolution insufficiency are nowadays hardly any issues for ever improving modern digital cameras.

Largely motivated to combat the problem of backlighting and compensate for extreme lighting conditions, high-dynamic-range (HDR) imaging techniques [10] have been researched and developed. In HDR imaging, cameras take two or more successive photos with different exposure settings. The resulting multiple photos are then fused to generate an image of much improved visual quality. However, HDR is susceptible to motion artifacts when fusing differently exposed images; furthermore, HDR can only be used for image acquisition in backlighting, not restoration of existing backlit images of poor quality.

In this work, we investigate how to repair single-exposure backlit images and improve their visual quality. A novel approach of spatially adaptive tone mapping in conjunction with soft segmentation is proposed to solve the problem of backlit image restoration. Conventional image enhancement methods can be applied to improve backlit images, but their effectiveness is quite limited albeit without the shortcomings of multi-image HDR. Global enhancement methods, such as histogram equalization (HE), can hardly correct the problems of underexposure and overexposure with a single tone mapping function. Local context-sensitive enhancement methods, such as contrast limited adaptive histogram equalization (CLAHE) [78] and the Retinex algorithm [35], are more suitable for enhancing backlit images than HE because they can adapt to different local intensity distributions of backlit images. However, these methods are susceptible to halo artifacts. The halo artifact problem can be somewhat

alleviated by edge-aware image enhancement techniques, *e.g.* those based on edge-preserving decomposition [19] and local Laplacian filtering [55]. Although the edge-aware image enhancement approach can amplify image details by stretching dynamic range around edges, it tends to destroy tone smoothness when applied to backlit images.

A common flaw with the existing enhancement methods, if directly used to restore backlit images, is that they operate on back-lit and front-lit regions indiscriminately, ignoring the fact that these two types of regions exhibit drastically different colour distributions. Relatively few papers were devoted explicitly to backlit image restoration, although backlighting is a big culprit for poor image quality and a common nuisance for almost all camera users. In [69] Tsai and Yeh proposed a simple method of backlit region detection by thresholding the luminance. The detected backlit regions were linearly stretched to increase the contrast. This method was refined in [43] by quad-tree region growing and guided filtering [30], but still relying on thresholding. Although simple thresholding meets intuition that backlit objects have lower luminance, it could misclassify dark pixels inside front-lit regions into the class of backlit regions. Another work [32] was inspired by the dark channel method, which was initially proposed for haze removal [29]. However, light transmittance is not a reliable indictor to distinguish the back-lit and front-lit surfaces. Lu and Sun proposed an automatic exposure correction algorithm that was used on backlit image restoration [75]. This is a hybrid approach of tone mapping and edge-aware filtering.

In this research, we develop a new, spatially adaptive technique to restore backlit images. The key innovation of the new technique is learning-based classification of backlit and frontlit regions, denoted by ♦-regions and ♢-regions. This classifier allows

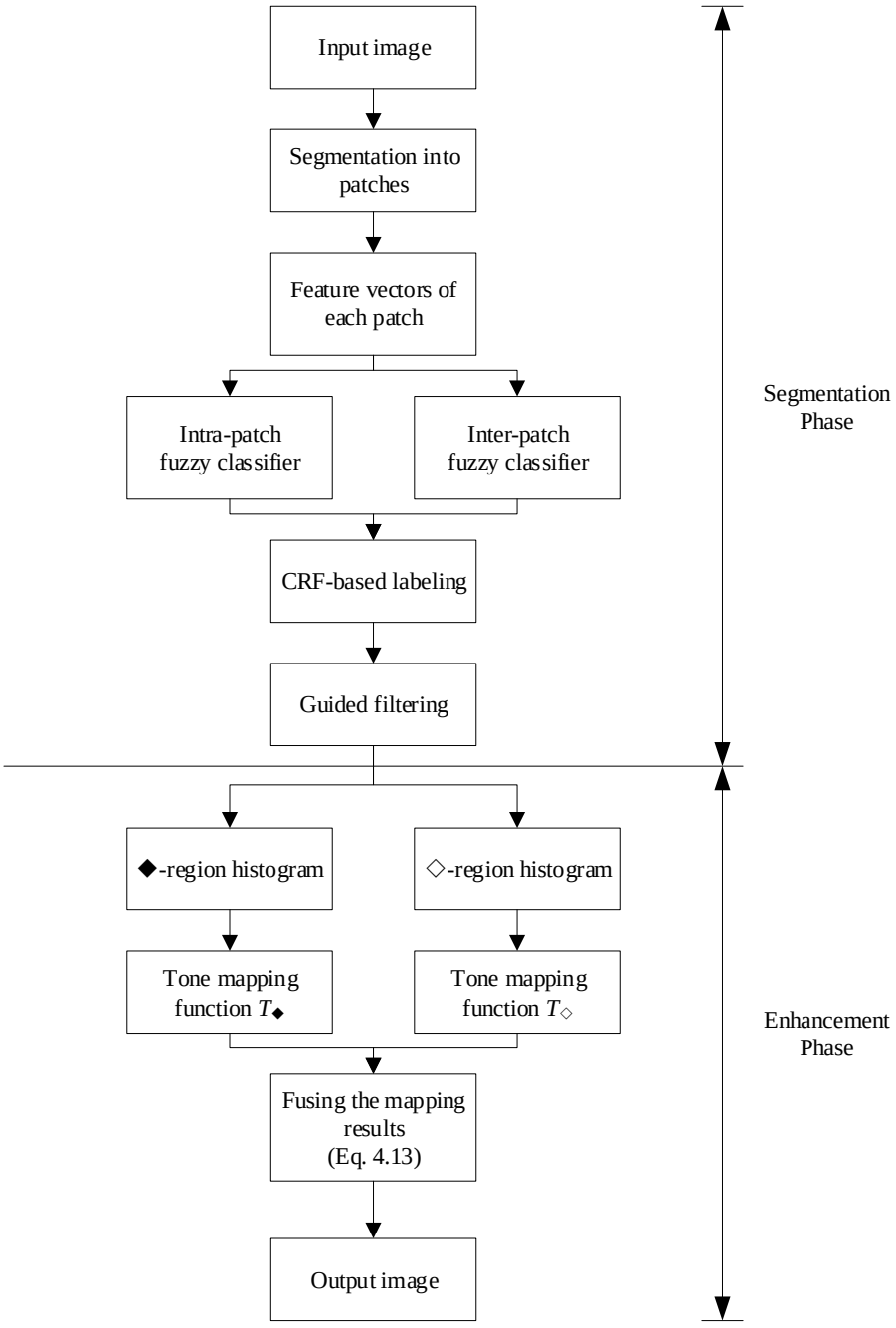


Figure 4.1: Flow of the proposed backlit image restoration method.

us to design two optimal tone mapping functions, one for enhancing each type of ill-illuminated regions. Following this line of thinking, we propose a two-phase backlit image restoration technique, as illustrated in Fig. 4.1, that performs an object-guided segmentation of ♦- and ♢-regions followed by spatially adaptive tone mapping.

The main difficulty of the proposed backlit restoration approach lies in accurate segmentation of ♦-regions, due to the ambiguity between ♦-regions and the dark areas that belong to ♢-regions, such as dark coloured surfaces and shadows. The problem of ♦- and ♢-regions separation is different from that of shadow detection; the latter has been studied far more extensively than the former [25, 39, 53, 77]. Shadow detection algorithms cannot be used for the purpose of restoring backlit images, because shadows in ♢-regions are often meaningful and desired characteristics of the scene; they should not be confused with ♦-regions and consequently get removed or diluted. In terms of physical causes backlit areas have much in common with shadows, namely, both are the results of low exposure to the light sources. But the two phenomena differ in semantics. Shadows and ♦-regions have different characteristics in their contours. As the shadow of an object is cast on other objects, its contour does not coincide with an object boundary in general. As a result, pixels on both sides of shadow contours belong to the same surface, so they are almost the same in many features except luminance, such as texture, chrominance, etc. This property has been exploited by shadow detection methods [25, 77]. On the other hand, the contours of backlit regions generally agree with object boundaries, so pixels on opposite sides of the boundary of a backlit region differ significantly in more features in addition to luminance, due to different surface properties. The above correlation between natural object boundaries and ♦-♢ boundaries offers useful features for detecting the

♦-♢ borders and for distinguishing ♦-regions from shadows.

In the segmentation phase, the input backlit image is partitioned into ♦- and ♢-regions by a soft binary classifier. Two support vector machine (SVM) based sub-classifiers are formed to locate backlit patches and ♦-♢ borders respectively, trained via supervised learning. These SVM results are further combined by a conditional random field (CRF) [42] based labeling technique to generate the final segmentation.

Next in enhancement phase, we apply different tone mappings on both regions respectively. The mapping functions are obtained by optimization maximizing contrast gain with bounded distortions in tone and chrominance. The separate enhanced results are finally fused into one to form the output.

The chapter is organized as follows: Sec. 4.2 details the segmentation phase, Sec. 4.3 concentrates on the enhancement phase, and Sec. 4.4 presents the results of the proposed method on a variety of backlit scenes, together with a comparison with related methods, and we come to conclusions in Sec. 4.5.

4.2 Backlit Region Detection

In this section, we present a data-driven scheme for segmenting input images into ♢- and ♦-regions by exploiting statistical differences between ♦- and ♢-regions and the characteristics of the boundaries between these two types of regions.

We first segment the image into small patches $\{\mathbf{x}_i\}$ by simple linear iterative clustering (SLIC) algorithm [1], such that all pixels of a given patch are either in or outside of a ♦-region. A graph is constructed in which the patches $\mathbf{x}_i$ form the vertices and the borders between every pair of adjacent patches $\{\mathbf{x}_i, \mathbf{x}_j\}, \forall\{i,j\} \in \mathcal{N}$ form the edges, $\mathcal{N}$ being the neighbourhood system which is a set of all adjacent

patches.

Next we develop a labeling technique to assign every patch $\mathbf{x}_i$ a corresponding label y_i, which takes 0 if the patch lies in an $\diamondsuit$-region, or 1 if it lies in a $\blacklozenge$-region. We solve the labeling problem by seeking for the maximum a posteriori (MAP) solution:

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} p(\mathbf{y}|\mathbf{x}) \tag{4.1}$$

We model the conditional probability $p(\mathbf{y}|\mathbf{x})$ by CRF [42], in which the labeling of $\mathbf{x}_i$ given the observation $\mathbf{x}$ is completely dependent on that of $\mathbf{x}_i$'s neighbouring patches, due to the Markovianity:

$$p(y_i|y_j, \mathbf{x}, j \neq i) = p(y_i|y_j, \mathbf{x}, \forall j \ s.t. \{i, j\} \in \mathcal{N}) \tag{4.2}$$

The joint distribution over the labels $\mathbf{y}$ given $\mathbf{x}$ in CRFs has the form [42]

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\left(\sum_i V_1(y_i|\mathbf{x}) + \lambda \sum_{\{i,j\}\in\mathcal{N}} V_2(y_i|y_j, \mathbf{x})\right)\right) \tag{4.3}$$

where the exponent (called Gibbs energy [5]) consists of a unary potential $V_1(\cdot)$ and a pairwise potential $V_2(\cdot)$ weighed by a coefficient λ, representing intra-patch and inter-patch confidence of labeling respectively. Both of the potentials should be non-negative. Hence the MAP solution can be obtained by minimizing the Gibbs energy function above; in other words, the smaller the Gibbs energy is, the closer the labeling is to the ground truth. We detail the definitions of $V_1(\cdot)$ and $V_2(\cdot)$ in the following discussions.

4.2.1 Intra-patch Term V_1

As shown in Fig. 4.1, we train a binary SVM classifier with Gaussian radial basis function (RBF) kernel (in which $\sigma = 1$) for whether an arbitrary patch $\mathbf{x}_i$ belongs to a backlit or frontlit region. The training backlit images are manually labeled into backlit and frontlit regions. For patch $\mathbf{x}_i$, a feature vector $\mathbf{v}_i$ is chosen to drive the classifier. Features to discriminate ♦- and ♢-regions include the spatial distribution of illumination and several statistics of the patch, such as the bright channel value, the skewness and average saturation. The discrimination power of these features is demonstrated in Fig. 4.2, and they are explained and justified one by one in the following subsections.

Estimated Illumination

Illumination is a decisive cue for separating ♦- and ♢-regions. Recall the image formation model (4.5) that, the illumination ℓ satisfies

$$I(x, y, \lambda) = s(\lambda)\ell(x, y, \lambda)r(x, y, \lambda) \tag{4.4}$$

However ℓ is hard to extract from 2D images directly. As ℓ is white for natural scenes, we can imply from (4.4) that

$$\log \mathcal{L}(x, y) = \log \ell(x, y) + \log r(x, y) + C \tag{4.5}$$

where C is a constant. Hence we can apply homomorphic filtering which low-pass filters $\log \mathcal{L}(x, y)$ to make a coarse estimation of $\ell(x, y)$, because ℓ tends to be of low frequency, whereas the reflectance signal r dominates the high frequency part in most

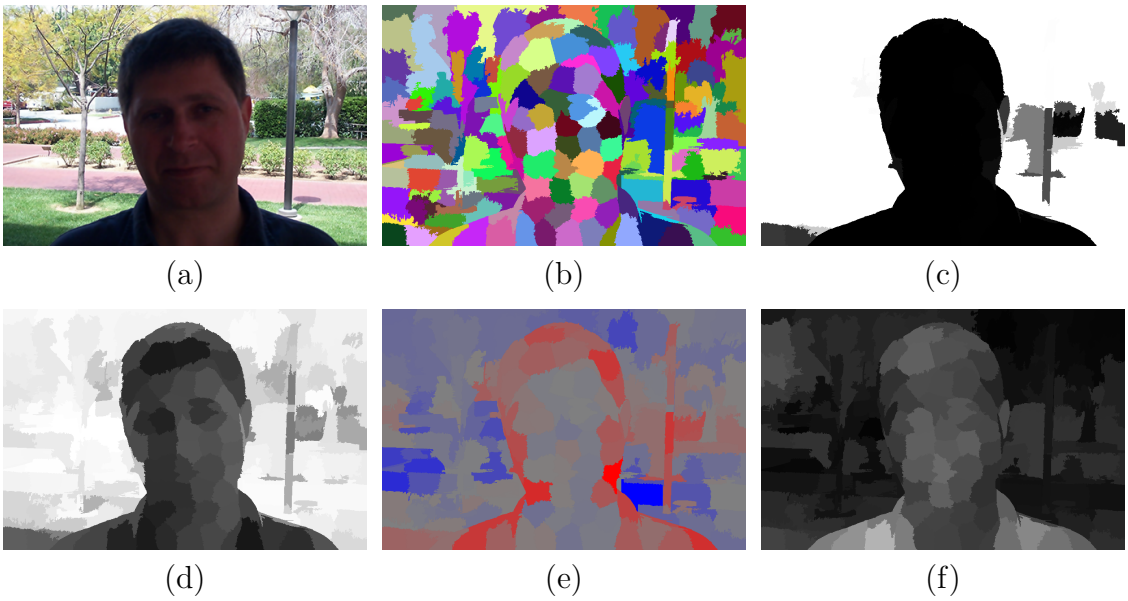


Figure 4.2: Features applied in the proposed SVM classifier. (a) Original image; (b) segmentation result by SLIC [1]; (c) estimated likelihood of being backlit by (4.7); (d) bright channel (4.9); (e) skewness; (f) saturation. In (c)(d)(f), patches with higher intensity are with larger values; in (e), red patches are with positive skewness and blue ones have negative skewness.

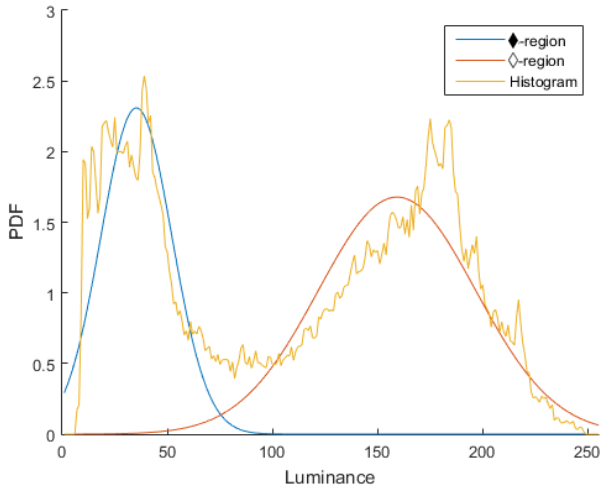


Figure 4.3: The luminance histogram of Fig. 4.2(a) (the yellow curve) modeled by 2-GMM (the blue curve for $w_{\blacklozenge}$ and the red curve for $w_{\lozenge}$).

natural scenes.

It can be observed that the histogram of the illumination signal ℓ is typically of two peaks, corresponding to ♦- (the low-intensity peak) and ♢-regions (the high-intensity peak) respectively. This distribution of luminance l can be effectively modeled by a mixture of two Gaussian components (2-GMM) $w_{\blacklozenge}$ and $w_{\diamondsuit}$ (see Fig. 4.3):

$$p(l) = \sum_{k} P(w_k) p(l|w_k), k \in \{\blacklozenge, \diamondsuit\} \tag{4.6}$$

where $p(l|w_k) \sim N(\mu_k, \sigma_k)$ is the PDF of the Gaussian component w_j (including $w_{\blacklozenge}$, the ♦-region component and the ♢-region component $w_{\diamondsuit}$) specified by mean $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_k$, and $P(w_k)$ is the prior probability of the pixels generated from the component w_k. The unknown parameters of the GMM model $\boldsymbol{\theta} = \{P(w_k), \mu_k, \sigma_k\}$ are estimated by expectation-maximization (EM) method [15].

The likelihood of being backlit for a patch $\mathbf{x}_i$ in a GMM sense, denoted by $\phi(\cdot)$, can be coarsely approximated by the posterior probability generated by the ♦-region component $w_{\blacklozenge}$ regarding the average patch illumination on the patch, denoted by $\bar{\mathcal{L}}_i$:

$$\phi(\bar{\mathcal{L}}_i) = P(w_{\blacklozenge}) p(\bar{\mathcal{L}}_i|w_{\blacklozenge}). \tag{4.7}$$

which ranges between 0 (if the patch luminance $\bar{\mathcal{L}}_i$ is generated solely by $w_{\diamondsuit}$) and 1 (if $\bar{\mathcal{L}}_i$ is generated by $w_{\blacklozenge}$ solely). The larger $\phi(\bar{\mathcal{L}}_i)$ is, the more likely the patch $\mathbf{x}_i$ belongs to the ♦-region.

As shown in Fig. 4.2(c), $\phi(\cdot)$ is statistically close to the ground truth though with some misclassified patches, thus adopted into the feature vector.

However, there are cases where the 2-GMM cannot fit the histogram of an input

backlit image well; the resulting segmentation result may not be sufficiently accurate (see Fig. 4.7 (d) for such examples). This motivates us to seek for more features to make the patch classifier more robust, which are introduced in the following subsections.

Bright Channel

Based on the observation that all three RGB channels have low values for pixels where little light arrives, Panagopoulos *et al.* proposed the concept of bright channel [53]. The bright channel $\beta(\cdot)$ with respect to an RGB pixel $c_j = \{c_j^{(r)}, c_j^{(g)}, c_j^{(b)}\}$ is defined by

$$\beta(c_j) = \max_{n\in\{r,g,b\}} \max_{k\in\Omega_j} c_k^{(n)} \tag{4.8}$$

where Ω_j is a rectangular neighbourhood patch centered at the pixel c_j. This metric was applied in shadow detection [53] as pixels with high bright channel values have little probability to be shaded.

In backlit scenes, we find the ⧫-regions have even lower bright channel value in contrast to ♢-regions. We extend the pixel-based definition to an arbitrary patch $\mathbf{x}_i$:

$$\beta(\mathbf{x}_i) = \max_{n\in\{r,g,b\}} \max_{\mathbf{c}\in\mathbf{x}_i} \mathbf{c}^{(n)} \tag{4.9}$$

Refer to Fig. 4.2(d) for a bright channel image.

Skewness

As discovered by Motoyoshi *et al.* in [50], the skewness of luminance histogram reflects the properties of object surfaces. We find there is significant difference in

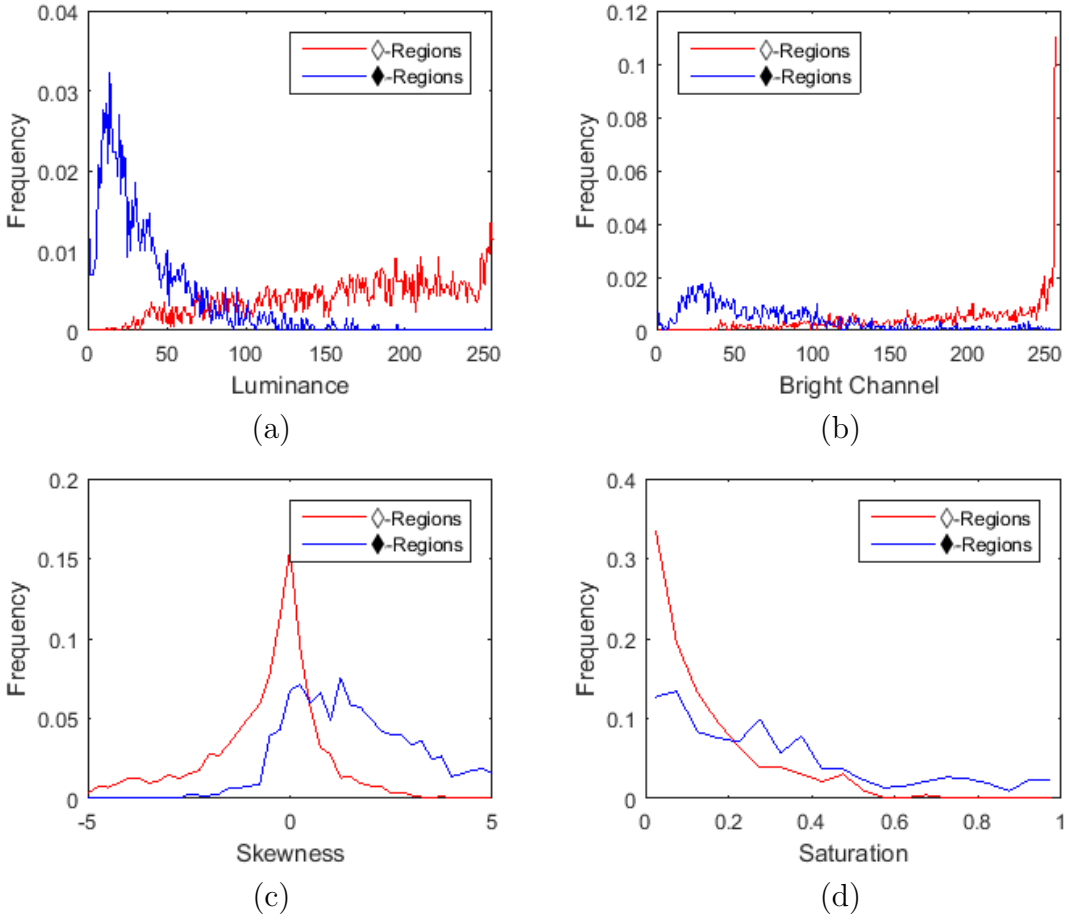


Figure 4.4: Statistics of all ♦- (blue) and ◊-region (red) patches in dataset images including the distribution of luminance (a), bright channel values (b), skewness (c) and saturation (d). Note that the luminance histograms of ◊- and ♦- regions are biased towards opposite direcitons, resulting in skewness of opposite signs.

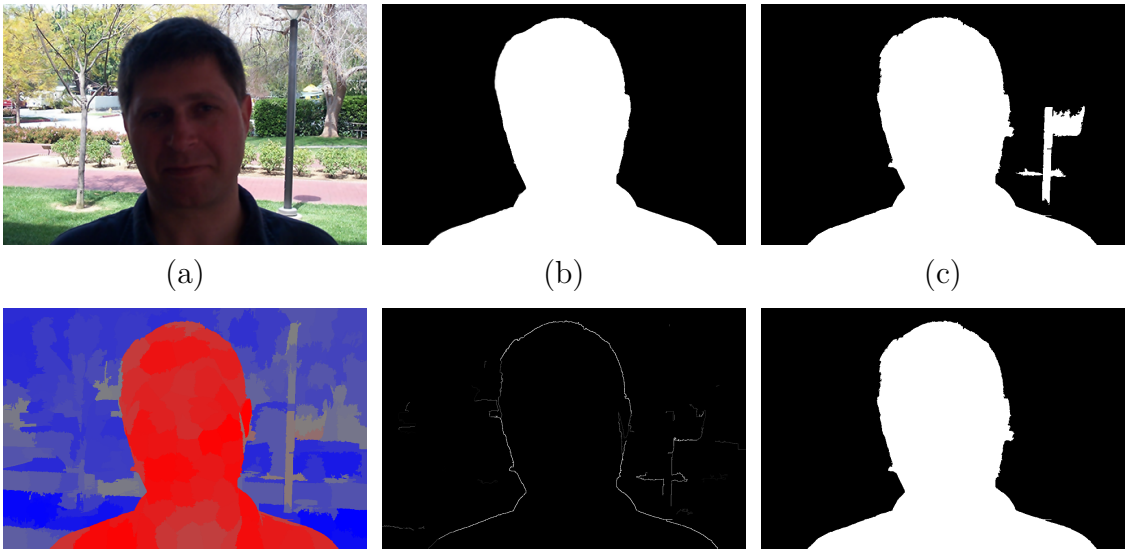


Figure 4.5: The segmentation phase. (a) Original image; (b) ground truth of segmentation; (c) binary output of the SVM classifier; (d) fuzzy output of the SVM classifier S_1 (4.10), where red coloured patches are classified as ♦-region, and blue coloured ones are ♢-regions; (e) fuzzy output of the ♦-♢ border classifier S_2 (Sec. 4.2.2); (f) segmentation result of (4.1).

the distribution of luminance skewness between ♦- and ♢-regions, as depicted in Fig. 4.4(c): skewness of patches in ♦-region are almost all positive, but that of ♢-region patches tends to be negative instead. This is mainly because skewness is indicative of the shape of the luminance histogram: as shown in Fig. 4.4(a), the luminance histogram of ♦-regions is typically tailed on the right, resulting in positive skewness, whereas the tail of ♢-region histogram is to the left, generating negative skewness.

Saturation

Apart from the luminance-based features above, we also consider the difference between ♦- and ♢-regions in terms of chrominance. In most natural scenes, the incident light is almost white so that it dilutes the colour of the objects illuminated, resulting in lower saturation than the ones receiving less illumination (see Fig. 4.2(f)). As revealed in Fig. 4.4(d), the saturation of ♢-regions is more biased toward low values, compared with the saturation of ♦-regions.

Formation of Feature Vectors

We include the aforementioned four features in the feature vector. In addition to the features of the patch $\mathbf{x}_i$ itself, we also include the four features regarding a neighbourhood of the patch $N(\mathbf{x}_i)$, with $N(\mathbf{x}_i)$ covering all adjacent pixels of $\mathbf{x}_i$ within a radius. Therefore, the feature vector of $\mathbf{x}_i$ is 11-dimensional: three dimensions for the colour centroid vector of $\mathbf{x}_i$, four for the features of $\mathbf{x}_i$, together with the other four for the features of $N(x_i)$.

In practice, instead of applying the binary output by the trained SVM directly,

we adopt the distance between the projection of the input vector in the feature space and the separating hyperplane, which implies higher confidence if larger, to provide a "fuzzy" segmentation. Let d_i be the distance to the separating hyperplane in the SVM feature space, ς_i be the binary SVM output (± 1, positive for being backlit), and then the soft SVM output $S_1(\mathbf{v}_i)$ is given by

$$S_1(\mathbf{v}_i) = \frac{1}{2}\left(1 + f(\varsigma_i \cdot \frac{d_i}{\max_i\{d_i\}})\right) \tag{4.10}$$

where $f(\cdot)$ is a sigmoid function, normalizing S_1 in the range $[0, 1]$. As d_i increases, S_1 approaches to either 0 or 1, meaning that the confidence level of the classifier increases.

We can see from Fig. 4.5(c) that the backlit regions estimated by the trained SVM classifier are rather close to the ground truth, while some isolated outliers still exist.

Finally, we combine two likelihood proxy terms, the soft SVM result S_1 and the GMM inference $\phi(\cdot)$, so that $\frac{1}{2}(S_1 + \phi(\cdot))$ is a better estimate of the probability that a patch is in a backlit region. The intra-patch potential $V_1(y_i|\mathbf{x})$ is then given by:

$$V_1(y_i|\mathbf{x}) = \left(y_i - \frac{1}{2}(S_1(\mathbf{v}_i) + \phi(\bar{L}_i))\right)^2 \tag{4.11}$$

which takes smaller value when the labelling result is closer to the estimate.

4.2.2 Inter-patch Term V_2

The segmentation result given by the patch classifier proposed in Sec. 4.2.1 is still imperfect, because the observation $\mathbf{x}$ is the only prior used; the correlation between the labelings of adjacent patches is neglected. Hence the pairwise potential V_2 in (4.3),

also called smoothness term in literature, is also important in the labeling scheme as it exploits such correlations.

Several existing CRF-based shadow detection methods embody $V_2(y_i|y_j, \mathbf{x})$ by $(y_i - y_j)^2$ [25, 39, 77] to impose penalty on the event that any adjacent pixels or regions are labeled (as shadows or not) differently, defying the common sense that the shaded area is mostly connected not scattered. However, the penalty should not be equal over all pairs, for the discontinuity in labeling is necessary at the boundaries of shadows, which is overlooked in those methods. In this work, we do consider the effect of the ♦-♢ borders in determining the penalty term $V_2(y_i|y_j, \mathbf{x})$.

The correlations between adjacent labels can be summarized into two points: 1) the labels should be constant over patches that belong to the same surface as they are almost equally exposed; 2) discontinuity of labels should only exist at ♦-♢ borders. So V_2 should be zero if both are satisfied, or take larger values to penalize the situation of discontinuity in labeling at any location but ♦-♢ borders. Based on this, V_2 is given by:

$$\begin{aligned} V_2(y_i|y_j, \mathbf{x}) =& (1 - S_2(\mathbf{v}_i, \mathbf{v}_j))(y_i - y_j)^2 + \\ & S_2(\mathbf{v}_i, \mathbf{v}_j)((1 - (y_i - y_j)^2) \\ =& S_2(\mathbf{v}_i, \mathbf{v}_j) + (1 - 2S_2(\mathbf{v}_i, \mathbf{v}_j))(y_i - y_j)^2 \end{aligned} \tag{4.12}$$

where S_2 is obtained from a ♦-♢ border classifier (to be shown later). V_2 imposes penalty against two situations in labeling: continuity at ♦-♢ borders, and discontinuity elsewhere.

The ♦-♢ border classifier is a linear SVM regarding the feature vectors of the patches on both sides $\mathbf{v}_i, \mathbf{v}_j$, namely a 22-dimensional feature vector for each pair of

adjacent patches. It also gives binary results as the intra-patch classifier in Sec. 4.2.1 does where 1 means the border of the given pair of patches is a ♦-♢ border and 0 means not. We apply the same normalization technique as (4.10) to provide a "fuzzy" classification, which forms the definition of S_2.

4.2.3 Solving the CRF

After constructing the graph, we apply graph cuts algorithm [41] to obtain the solution to (4.3), where the Lagrangian multiplier λ is empirically set to 0.3. The output is finally refined by guided filtering [30] so that it aligns closely with object boundaries in the original image. Meanwhile, as the labeling is binary, guided filtering can also "soften" the segmentation result to better suppress any possible artifacts at ♦-♢ borders.

Fig. 4.5 gives an example of how the proposed segmentation scheme prevents misclassifying dark objects and shadows into ♦-regions: although outliers still exist in the result given by the intra-patch classifier, they are effectively eliminated as the ♦-♢ border classifier does not output high confidence around them (see the pole and plants behind the person).

4.3 Restoration of Backlit Images

Based on the fuzzy binary segmentation of a backlit image I into the ♦-regions and ♢-regions, we enhance the image by applying respective tone mapping functions $T_{\blacklozenge}(\cdot)$ and $T_{\diamondsuit}(\cdot)$ on the ♦- and ♢-regions.

As shown in Fig. 4.4(a), the luminance histograms of ♦- and ♢- regions (denoted

by $\mathbf{p}_{\blacklozenge}$ and $\mathbf{p}_{\diamondsuit}$) differ significantly, both having a relatively small dynamic range. Thus image details in both $\blacklozenge$-regions and $\diamondsuit$-regions can be enhanced by separate tone mapping functions $T_{\blacklozenge}(\cdot)$ and $T_{\diamondsuit}(\cdot)$ that stretch the dynamic range of $\mathbf{p}_{\blacklozenge}$ and $\mathbf{p}_{\diamondsuit}$, respectively, using the contrast enhancement algorithm presented in Chapter 3. Specifically we construct $\mathbf{p}_{\blacklozenge}$ with the pixels where the fuzzy segmentation result $y_i \geq 0.5$ and $\mathbf{p}_{\diamondsuit}$ with the remaining pixels.

To prevent possible artifacts around $\blacklozenge$-$\diamondsuit$ borders due to different tone mappings $T_{\blacklozenge}(\cdot)$ and $T_{\diamondsuit}(\cdot)$, we apply both tone mappings $T_{\blacklozenge}(l_i)$ and $T_{\diamondsuit}(l_i)$ to pixels on or near the borders, where l_i is the luminance of the pixel $\mathbf{c}_i$, and then fuse the results to generate the output luminance l'_i:

$$l'_i = y_i T_{\blacklozenge}(l_i) + (1 - y_i) T_{\diamondsuit}(l_i) \tag{4.13}$$

so that each output pixel can be obtained using (3.24).

We present the proposed backlit image restoration algorithm in pseudocode at Algorithm 2.

4.4 Experimental Results

The experiments are conducted on a dataset of 64 images, captured in various backlit scenes by different devices. Half of the images are used to train the classifiers, and the remaining are for validations.

input : Input image I
output: Output image I'
begin
  $\{\mathbf{x}_i\} \leftarrow$ `slic(`x`)` ; `// Perform SLIC segmentation`
  $\{\mathbf{v}_i\} \Leftarrow \{\mathbf{x}_i\}$; `// Compute the feature vector of each patch` $\mathbf{x}_i$
  `// Construct the CRF`
  $\{S_1(\mathbf{v}_i)\} \Leftarrow \{\mathbf{v}_i\}$; `// Run Intra-patch SVM on each patch using` (4.10)
  $\{S_2(\mathbf{v}_i)\} \Leftarrow \{(\mathbf{v}_i, \mathbf{v}_j)\}$; `// Run Inter-patch SVM on each patch`
  `// Solve the CRF using graph-cut algorithm`
  $V_1(\mathbf{y}) \leftarrow S_1, V_2(\mathbf{y}) \leftarrow S_2$; `// Using` (4.11) `and` (4.12)
  $\mathbf{y} \leftarrow \arg\min_{\mathbf{y}}(V_1 + \lambda V_2)$;
  `// Perform guided filtering`
  $\{y_i\} \leftarrow$ `guidedfilter(`$\mathbf{y}$`)` ; `//` y_i `is classfication result for pixel` $\mathbf{c}_i$ `in` I
  `// Run Algorithm 1 on both regions`
  $R_B \leftarrow \{\mathbf{c}_i | y_i \geq 0.5\}$; $T_\blacklozenge \leftarrow$`algorithm1(`$R_B$`)`;
  $R_F \leftarrow \{\mathbf{c}_i | y_i < 0.5\}$; $T_\diamondsuit \leftarrow$`algorithm1(`$R_F$`)`;
  `// Perform luminance mapping together with inverse colour transform`
  foreach $\mathbf{c}_i \in I$ **do**
    $L'_i \leftarrow y_i T_\blacklozenge(l_i) + (1 - y_i) T_\diamondsuit(l_i)$;
    $\mathbf{c}'_i \leftarrow$ result of (3.24) on pixel $\mathbf{c}_i$;
  end
  return I';
end

Algorithm 2: Proposed backlit image restoration method (in pseudocode).

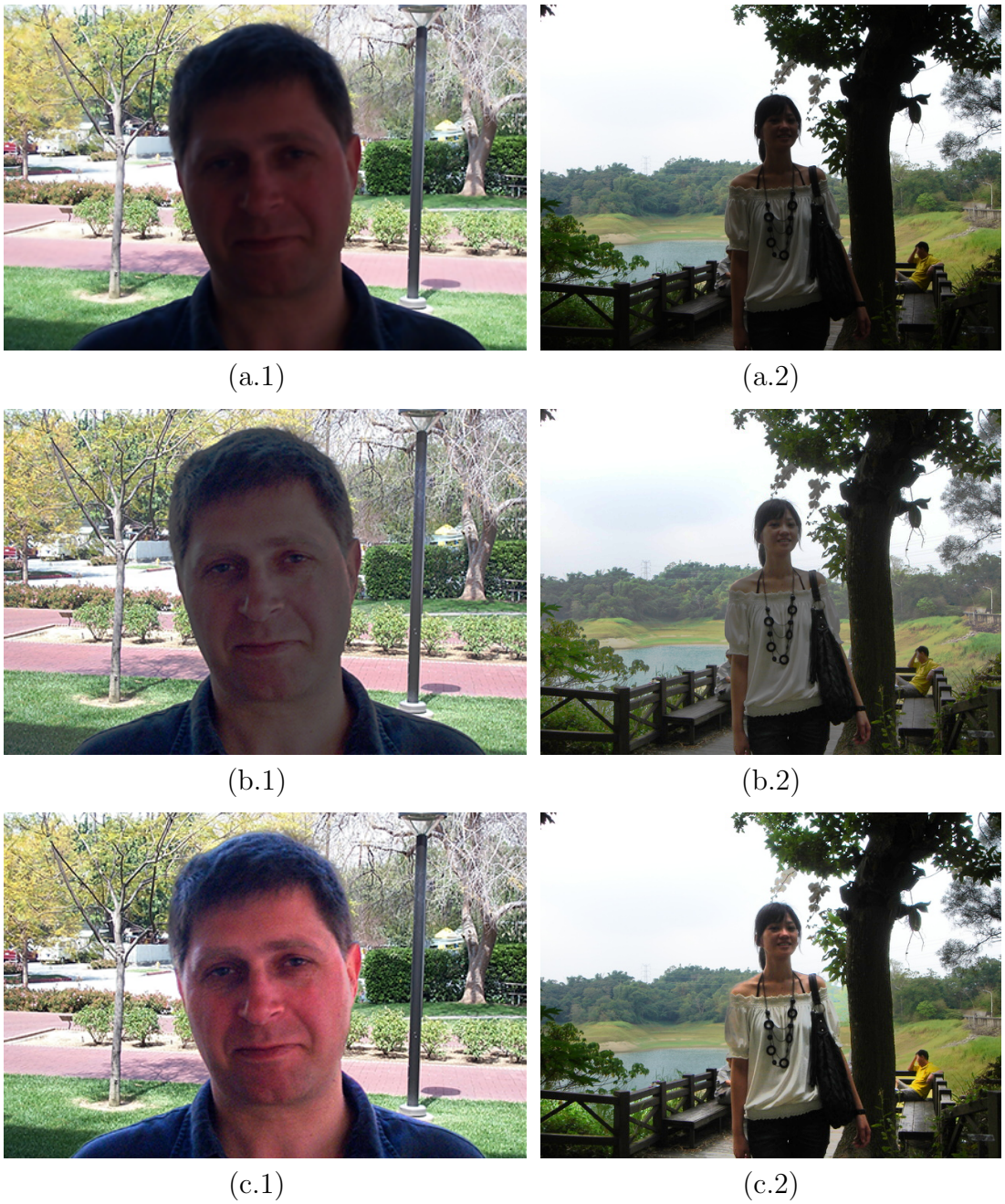
4.4.1 Accuracy of Segmentation

First, we demonstrate the power of the proposed learning-based segmentation in the restoration of backlit images. In Fig. 4.6 we compare restored backlit images that are generated by optimal tone mapping technique in Chapter 3, without versus with segmentation. It is evident that the new segmentation-based backlit image restoration method, by spatially adaptive tone mapping, achieves close to ideal harmonized lighting conditions, with all parts of the image properly exposed, leading to superior image quality. In contrast, the original OCTM method without segmentation fails to achieve the same level of visual quality in both ♦- and ♢-regions as the proposed method; for instance, the frontlit regions of Fig. 4.6 (b.2) have reduced contrast, low dynamic range and tone artifacts.

Next, we closely evaluate the performance of the segmentation method, and the final restoration results in comparison with those of existing methods. We compare results of our segmentation phase with several other methods mentioned above, including the 2-GMM illumination estimator (4.7), the intra-patch SVM classifier (Sec. 4.2.1), and the method by Lee *et al.* [43], in both quantitative and qualitative aspects.

The accuracy of classification and the root-mean-square error (RMSE) against the ground truth of them are tabulated in Table 4.1. As revealed by the comparison, machine-learning based classifiers have superior performance over methods based on image statistics only. Next we present more examples that further differentiate the capabilities of intra-patch classifier and the proposed method, though they are close in objective assessment.

Fig. 4.7 exhibits the results of these methods versus hand labeled ground truth on



(a.1) (a.2)

(b.1) (b.2)

(c.1) (c.2)

Figure 4.6: Backlit image restoration results without and with the proposed segmentation phase. First row: original images; second row: restored images without segmentation; third row: restored images with segmentation phase.

(a) (b) (c) (d) (e) (f)

Figure 4.7: Comparison of various backlit region detection methods. (a) Original image; (b) ground truth of segmentation; (c) method by Lee *et al.* [43]; (d) likelihood ϕ with respect to the estimated illumination (4.7); (e) binary output of the intra-patch SVM classifier (Sec. 4.2.1); (f) segmentation result (4.1).

several representative images in the dataset. These images contain patches that are prone to misclassification, for instance the dark green bushes in the background (in the first row), the white text on the sign (in the second row) and the light-coloured part on the girl's dress (in the third and fourth row), making the segmentation more challenging. We can see from Fig. 4.7 (c) that [43] makes most misclassifications by simply classifying all dark pixels into ♦-regions. The misclassifications get somewhat relieved by the 2-GMM classifier (Fig. 4.7 (d)), but it may fail in some cases with too much ambiguity (for instance, the dark-coloured sky in the third row is classified as ♦-region). The intra-patch SVM classifier (Fig. 4.7 (e)) provides more precise segmentation results than the former two methods, but the outputs are still vulnerable to the ambiguity brought by *e.g.* dark objects and shadows, with observable "orphans" in both ♦- and ♢-regions. With help of the ♦-♢ border classifier, the proposed method (Fig. 4.7 (f)) can effectively get rid of those outliers, producing results closest to the ground truth. Remarkably, the CRF-based labeling process still preserves isolated patches with high labeling confidence, as shown in the third row where the gap between the woman's arm and waist is not smoothened into the ♦-region, which demonstrates the robustness of the proposed segmentation scheme.

Table 4.1: Performance comparison of segmentation methods.

Method	Accuracy(%)	RMSE
Lee *et al.* [43]	74.61	0.3938
2-GMM	82.59	0.3423
Intra-patch SVM	94.36	0.1950
Proposed	96.10	0.1547

Input HE CLAHE Retinex

Farbman *et al.* Yuan and Sun's Paris *et al.* Proposed

Input HE CLAHE Retinex

Farbman *et al.* Yuan and Sun's Paris *et al.* Proposed

Input HE CLAHE Retinex

Farbman *et al.* Yuan and Sun's Paris *et al.* Proposed

Input HE CLAHE Retinex

Farbman *et al.* Yuan and Sun's Paris *et al.* Proposed

Figure 4.8: Backlit images enhanced by existing contrast enhancement methods and the proposed method. The presented existing methods include: HE, CLAHE, multi-scale Retinex method, edge-preserving decomposition [19] by Farbman *et al.* , Yuan and Sun's exposure correction method [75] and local Laplacian filtering by Paris *et al.* [55].

4.4.2 Subjective Evaluation

We invite the reader to compare the proposed method with existing mainstream enhancement methods in terms of the perceptual quality of the restored backlit images, by examining the experimental results in Fig. 4.8. It is evident that HE and the method of [75], which are based on global tone mapping, fail to mitigate the problems of underexposure and overexposure. Both algorithms CLAHE and Retinex, if applied to backlit images, suffer from severe halo artifacts; they hardly improve the visual quality of backlit images. The methods of [19] and [55], which are considered among the best image enhancement methods, are apparently not suited to the task either, producing severe artifacts in colour and tone; of course, this should not be a surprise because these methods are not meant to restore backlit images. In summary, none of the existing tone mapping and enhancement techniques is suited for backlit image restoration, justifying the need for dedicated backlit image restoration algorithms as developed in this paper. Indeed, the proposed method of backlit image restoration appears to produce visually most pleasing results, without introducing objectionable artifacts.

4.4.3 Objective Evaluation

We also evaluate these methods with two objective metrics of image quality for contrast enhancement operations. They are the histogram entropy and the luminance ordinal distortion (LOD). The histogram entropy is an information theoretical metric for the richness (variability) of image details, which may be considered as a proxy for contrast. Generally speaking, higher entropy corresponds to higher contrast and perceptual quality. LOD, on the other hand, is a measure on the distortion of order

statistics caused by the contrast enhancement process in sliding windows. It is defined as

$$LOD = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{\sum_{k=1}^{n}(\hat{v}_{ik} - v_{ik})^2}{n}} \tag{4.14}$$

where N is the total number of sliding windows, n is the number of pixels inside a window, and v_i and $\hat{v}_i$ are the ordinal vectors of the i-th window in the input and processed images, respectively. Ideally the processed image should not violate the order statistics of pixel values with a zero LOD; artifacts such as contours, halos and ghosts are associated with errors in order statistics. Therefore, LOD is an objective measure of artifacts due to over-enhancement.

We compute the above two objective quality metrics on the ♢- and ♦-regions, separately. The results over the set of 32 test images are reported in Table 4.2 and 4.3. Table 4.2 shows that the proposed method achieves the highest entropy in backlit regions, as it should because it is designed explicitly for enhancing underexposed regions. As frontlit regions typically do not suffer from low contrast problem, it is less critical to increase the entropy of the ♢-regions. All other methods perform almost as well as the proposed method in the entropy measure, except methods [35] and [75] actually reduce the entropy of ♦-regions. On the other hand, in the LOD

Table 4.2: Objective evaluation by histogram entropy [28].

Method	♢-region	♦-region
Input	1.7791	2.0159
HE	1.6857	2.2413
CLAHE	1.9560	2.2748
Retinex	1.9615	1.8696
Farbman *et al.*	1.9463	2.2830
Yuan and Sun's	1.6976	1.9519
Paris *et al.*	1.8675	2.1056
Proposed	1.7953	2.3096

Table 4.3: Objective evaluation by local ordinal distortion.

Method	♢-region	♦-region
HE	6.5594	6.1195
CLAHE	6.5141	6.0965
Retinex	7.7811	8.1528
Farbman *et al.*	6.9688	7.1189
Yuan and Sun's	5.7147	5.4613
Paris *et al.*	7.4335	6.8186
Proposed	2.3142	3.7818

measure, the competing methods are much inferior to the proposed method, reflecting the fact that they are prone to over-enhancement artifacts, as evident in Fig. 4.8 and discussed in the proceeding subsection. On balance the proposed method achieves superior perceptual quality by boosting contrast without introducing objectionable artifacts.

4.5 Discussions

We have presented a novel, learning-based, spatially adaptive technique of optimal tone mapping for the restoration of backlit images. The new technique first identifies the backlit regions using a binary soft classifier, and then enhances degraded image details due to poor illuminations by fusing the results of two optimal tone mapping functions, one designed for backlit regions and the other for the remainder of the image. The binary classifier for segmentation is trained via supervised learning on manually labeled segmentation data. Empirical results and objective measures are presented to establish the efficacy of the proposed technique.

Nonetheless, the new method has its limitations: the quality of output heavily relies on the accuracy of segmentation. Although techniques have been applied to

(a) (b) (c)

Figure 4.9: A failure case with incorrect segmentation. (a) Input image; (b) segmentation result; (c) output after enhancement.

refine the segmentation result, such as CRF optimization and guided filtering, failure may still occur. Fig. 4.9 is an example in which the complicated window grids confuse the classifier.

Chapter 5

Video Restoration against Yin-Yang Phasing

5.1 Introduction

5.1.1 The problem and background

A highly irritating type of video degradation, called Yin-Yang Phasing (YYP) in this paper, is spatially patchy and temporally inconsistent objects appearance, with parts of the scene turning in and out the state of under-exposure (yin) or over-exposure (yang), back and forth. Such effects are quite common in impromptu-made video materials in daily life. An example is presented in Fig. 5.1: in a short progression of few video frames, which are produced by an iPhone 6 camera, the person's face alters from normally lit to unintelligibly dark as she moves her head. Moreover, the involuntary, dramatic changes in the intensity (luminance) and possibly chrominance are compounded by low contrast due to wrongly exposed objects. The YYP effects

Figure 5.1: Top row: temporally inconsistent frames due to changes in exposure time. Bottom row: frames restored by the proposed method.

are caused by unevenness of illumination, significant changes in the directions of incident and reflection lights due to object and/or camera motions, and improper camera operations by amateur users. These scenarios tend to confuse the ubiquitous "dumb" cameras and lead to incorrect and time-lagged settings of auto-exposure and white point. Unfortunately, the above root cause of YYP is an inherent weakness of mass-produced cameras that will trouble average users for a foreseeable future, particularly when shooting videos in unevenly-lit scenes.

In difficult illumination conditions, disabling camera's auto-exposure functionality cannot cure but rather aggravate the YYP problem for most users. For example, in backlit scenes like in Fig. 5.1, without auto-exposure the person in the foreground, the very focus of the video session, will be severely underexposed in most frames. Also, out of question is the use of auxiliary lighting to compensate for problematic

illumination in the scene, as it is beyond the means and knowledge of amateur camera users.

The YYP type of video degradation is becoming a major culprit of poor video quality, much more so than insufficient spatial resolution, low frame rate, sensor noises, and compression distortions. Nowadays even consumer-grade cameras boast very high pixel counts, frame rates of 60Hz and above, and low senor noise level; furthermore, modern communication infrastructures can support high throughput visual data exchanges. While these hardware advances are making superresolution, frame rate upconversion, denoising, compression artifacts removal, etc., less important in practice, they can do nothing to repair the YYP degradation as the lighting conditions in the scene and non-expert video shooting behaviors are beyond the control of video acquisition and communication equipment. Despite the daily encountering of the YYP video degradation problem in a wide range of video applications, such as social media, on-line video sharing, video monitoring, spontaneous video reporting, etc., very little research has been carried out on the YYP phenomenon and its mitigation. In this paper we investigate how to algorithmically remove the undesired YYP effects, aiming to restore the video to the state as though it was shot in spatially uniform and temporally steady lighting conditions. For lack of better wording, we call such a process YYP removal.

Unlike in the studies of traditional image/video restoration problems, such as deshaking [49], denoising [48], deblurring [7], superresolution [61], etc., the source of YYP degradation is not purely physical and hence difficult to model analytically. As a result, casting YYP removal as an inverse problem or into a mathematical programming formulation is not as easy as for other restoration problems.

5.1.2 Our approach

As tonal reproduction primarily depends on global image statistics, it is difficult to perform the YYP removal solely in pixel domain by tracking motion trajectory across frames and enforcing temporal tone consistency guided by the motion flow. Instead, we formulate the YYP removal as an inter-frame, spatially-adaptive, optimal tone mapping problem and propose an optimization approach to solve it, aiming to neutralize temporally unsteady intensity levels of the objects. The basic premise of our approach is that the overall appearance of an image is governed by the shape of its intensity histogram [50]. Therefore, we propose to retain the temporal consistency of tonal reproduction by mapping the intensity histograms of input frames to a common target histogram per a given scene. The other reason for adopting a histogram-based tone mapping approach is that it can be tuned to boost contrast as well, hence unifying the tasks of tonal stabilization and image enhancement.

As the human visual system can rapidly adapt to different luminance levels, our histogram targeting strategy for YYP harmonization can be made more effective if the tone mapping is tailored to image regions under different illumination conditions. Specifically, we segment a YYP-affected video frame into two types of regions, by the likelihood of a pixel being on a weakly or strongly illuminated object surface, denoted by the ⧫-region and ♢-region.

In order to prevent the ⧫-regions from being underexposed and the ♢-regions from being overexposed sporadically, we temporally track the ⧫-regions and ♢-regions. For a fixed video scene of N frames, let $R_{\blacklozenge}^{(n)}$ and $R_{\diamond}^{(n)}$, $1 \leq n \leq N$, be the sets of ⧫-regions and ♢-regions in frame n, respectively. Using the statistics of the data set $\{R_{\blacklozenge}^{(n)}\}_{n=1}^{N}$, an anchor intensity histogram $\mathbf{p}_{\blacklozenge}^{*}$ is constructed as the target for the

YYP harmonization of ♦-regions through all N frames, which can be viewed as a generalized centroid in terms of Kullback-Leibler distance of probability distributions [8]. Then, for each frame n, a tone mapping (histogram transformation) $T_{\blacklozenge}^{(n)}$ is computed to best match the histogram of $R_{\blacklozenge}^{(n)}$ to $\mathbf{p}_{\blacklozenge}^{*}$. By forcing all N histogram transformations $T_{\blacklozenge}^{(n)}$, $1 \leq n \leq N$, to approach the same output histogram $\mathbf{p}_{\blacklozenge}^{*}$, we make the tone reproduction of an object temporally consistent. In case the anchor histogram $\mathbf{p}_{\blacklozenge}^{*}$ is ill-shaped, which is common in YYP-degraded videos, one can employ the contrast enhancement technique in Chapter 3 prior to histogram matching. The same procedure outlined above can be applied to the set of ♢-regions $\{R_{\diamondsuit}^{(n)}\}_{n=1}^{N}$ to compute $\mathbf{p}_{\diamondsuit}^{*}$ and $T_{\diamondsuit}^{(n)}$, $1 \leq n \leq N$. The main objective of the proposed YYP harmonization algorithm is to stabilize throbbing intensity levels in any given ♦- or ♢-region. But it can also stabilize shifting chrominance of any given region, if required. We only need to first decompose the original video signal into luminance and chrominance components, and then apply the same region-adaptive inter-frame harmonization method to the chrominance component of the input video.

To summarize our novel YYP harmonization approach, objects under similar lighting conditions are spatially grouped and temporally tracked, the luminance and chrominance distributions of the these similarly-lit objects are optimized in terms of perceptual quality, and set as the objectives of tone mapping for all frames in a given scene of the YYP-degraded video. For each frame n, to prevent possible boundary effects of region-based tone mapping functions $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$, the results of $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$ are weighted based on a fuzzy light field segmentation.

5.1.3 Related works

Thus far most of research efforts on video restoration against adverse acquisition conditions are devoted to the stabilization of shaky frames caused by large, irregular camera jitters [49]. Of most relevance to this work is a 2011 publication by Farbman and Lischinski on tonal stabilization of video [18]. The authors addressed the problem of tonal inconsistency in consecutive video frames caused by improper auto-exposure and white points, and proposed a method to make global, smooth tone transitions from one frame to the next guided by selected anchor frames. In comparison, YYP is a different and more challenging video degradation problem: severe tonal fluctuations *compounded* by low contrast, which are caused by uneven, incorrect exposures. In [18] two assumptions are made: 1) lighting conditions in the scene do not change abruptly; 2) the tonal fluctuations are of a global nature not spatially varying; moreover, there is no significant loss of contrast due to under and/or over exposures. The YYP problem differs from the one in [18] in both the effect and the cause. YYP removal requires an approach of joint tone stabilization and contrast enhancement, as proposed by this paper.

The YYP phenomenon has similar visual characteristics as poor tone reproduction in HDR images. If each video frame is treated in isolation, it is tempting to try some of the many HDR tone mapping methods [16, 62] to correct poor use of dynamic range and improve the visual quality. However, this naive approach is highly prone to objectionable temporal artifacts. Very recently, Aydin *et al.* addressed the problem of temporal coherence in tone mapping of HDR video [4]. They proposed a method of edge-aware filtering method through pixel motion paths to achieve temporal stability of the enhanced video. But this work, like other HDR tone mapping methods, is

mainly about how to compress the intensity dynamic range while maintaining contrast and preventing artifacts; hence it is not suited to compensate for drastic changes in object appearance in time as required by the restoration of YYP videos.

This chapter is structured as follows. Sec. 5.2 presents an algorithm for temporally-constrained fuzzy light field segmentation, which is a preparation step for the main task of inter-frame region-adaptive YYP harmonization. Sec. 5.3 details the YYP harmonization algorithm, in particular explaining the choice of an anchor frame and the histogram targeting process. Sec. 5.4 discusses how to improve the performance of YYP harmonization by using an enhanced anchor frame to drive the algorithm developed in Sec. 5.3. Sec. 5.5 reports experimental results and performance evaluations.

5.2 Temporally-constrained fuzzy light field segmentation

As reasoned in the introduction, our YYP harmonization method needs to analyze an unevenly-illuminated video scene, and separate weakly-illuminated ⧫-regions from strongly-illuminated ♢-regions in the input frame. This can be implemented by homomorphic filtering and the 2-GMM fitting as described in Sec. 4.2.1. For particular scenes we also apply more sophisticated methods to achieve more precise segmentation, for instance the learning-based segmentation algorithm in Chapter 4 for backlit scenes, and face matting algorithm [11] for human head-and-shoulder scenes.

As the reader will appreciate shortly when we present the details of the YYP harmonization algorithm, the precision of region boundaries is less critical for the

algorithm performance. For visual quality of video restoration, more important than boundary precision is that the segmentation results should be consistent in time. Under the assumption that in a given video scene, the foreground object motion or/and camera motion is modest, we can achieve the desired temporal consistency by the following simple and fast inter-frame segmentation technique:

$$\mathcal{R}_j^{(n)} = ((\mathcal{R}_j^{(n-1)} \cap R_j^{(n)}) \oplus \tau) \cap (\mathcal{R}_j^{(n-1)} \cup R_j^{(n)}), \quad j \in \{\blacklozenge, \diamondsuit\} \tag{5.1}$$

where τ is the core for morphologic dilation, $\mathcal{R}_{\blacklozenge}^{(n)}$ ($\mathcal{R}_{\diamondsuit}^{(n)}$) is the temporally smoothed $R_{\blacklozenge}^{(n)}$ ($R_{\diamondsuit}^{(n)}$) with respect to its counterpart $\mathcal{R}_{\blacklozenge}^{(n-1)}$ ($\mathcal{R}_{\diamondsuit}^{(n-1)}$) in the previous frame. Set $\mathcal{R}_{\blacklozenge}^{(n)}$ (or reciprocally set $\mathcal{R}_{\diamondsuit}^{(n)}$) deterministically classifies every pixel in frame n to be on a weakly-illuminated (or strongly-illuminated) object surface or not. Such a hard-decision classification may, due to segmentation errors, generate boundary artifacts after pixels in $\mathcal{R}_{\blacklozenge}^{(n)}$ and $\mathcal{R}_{\diamondsuit}^{(n)}$ are gone through two different tone mappings $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$. Like Chapter 4, we introduce a simple fuzzy classification technique to eliminate the boundary artifacts. The idea is to fuse the results of $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$ by the likelihood of a pixel being on a weakly-illuminated object surface. The likelihood is determined by a fuzzy classifier $w^{(n)}$, given by the output of guided filtering of the hard segmentation result $\mathcal{R}_{\blacklozenge}^{(n)}$.The value of $w^{(n)}(x, y)$ will be used as the likelihood for pixel (x, y) being on weakly-illuminated object surface. Fig. 5.2 summarizes the temporally-constrained fuzzy segmentation-classification process presented in this section.

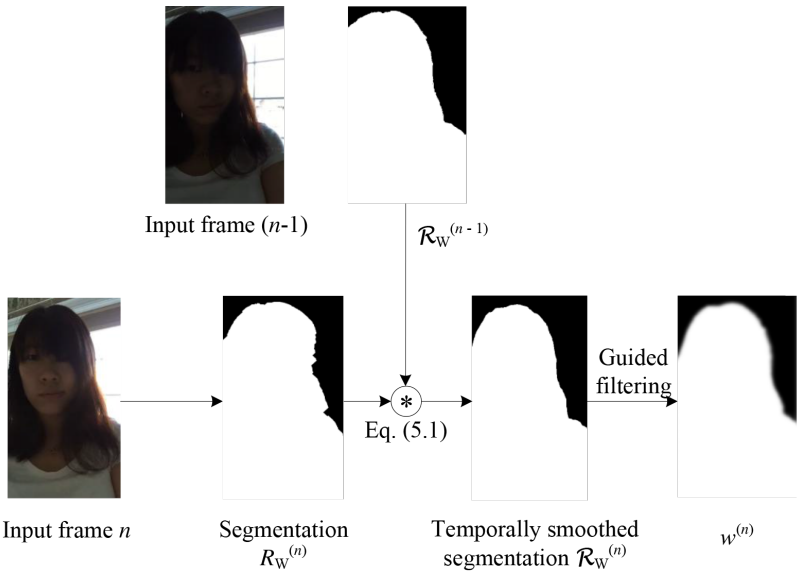


Figure 5.2: Steps of temporally-constrained fuzzy segmentation.

5.3 Interframe region-adaptive YYP harmonization

Now we discuss how to construct the tone mappings $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$ for restoring the interframe consistency of frame n in a YYP-degraded video. All of the following technical developments, conclusions and methods apply exactly the same way to the construction of $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$. Therefore, we can drop the subscripts in previous notations $T_{\blacklozenge}$, $\mathcal{R}_{\blacklozenge}^{(n)}$, $T_{\diamondsuit}$, $\mathcal{R}_{\diamondsuit}^{(n)}$, etc. to avoid symbol clutter, and discuss the case of $T_{\blacklozenge}^{(n)}$ only.

Let $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \cdots, \mathbf{p}^{(N)}$ be the intensity histograms of $\blacklozenge$-regions for the N input frames in a given video scene. We need to select, among the N input frames, frame n^* such that $\mathbf{p}^{(n^*)}$ is statistically the best representative of all other $N-1$ histograms.

Histogram $\mathbf{p}^{(n^*)}$ can be viewed a generalized centroid of the set $\{\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \cdots, \mathbf{p}^{(N)}\}$ and is used as the target intensity distribution for all tone-mapped frames by $T^{(n)}$, $1 \leq n \leq N$, to closely obey.

Using the Kullback-Leibler distance $D(\cdot||\cdot)$ between distributions, the generalized centroid histogram $\mathbf{p}^{(n^*)}$ can be computed as below:

$$\begin{aligned} \mathbf{p}^{(n^*)} &= \underset{\mathbf{p}\in\{\mathbf{p}^{(n)}\}_{n=1}^{N}}{\arg\min} \sum_{j=0}^{N} D(\mathbf{p}^{(j)}||\mathbf{p}) = \underset{\mathbf{p}}{\arg\min} \sum_{j=0}^{N} H(\mathbf{p}^{(j)}, \mathbf{p}) \\ &= \underset{\mathbf{p}}{\arg\min}\, H(\bar{\mathbf{p}}, \mathbf{p}) \end{aligned} \tag{5.2}$$

In other words, among all N frames of the scene, frame n^* is the one whose histogram has the minimum cross entropy with respect to the average histogram $\bar{\mathbf{p}}$ of the group.

Upon having selected the anchor intensity histogram $\mathbf{p}^{(n^*)}$, the tone mapping function $T^{(n)}$ for ♦-regions of frame n, $1 \leq n \leq N$, is computed via histogram matching:

$$T^{(n)}(k) = \underset{j}{\arg\min}\, |C^{(n)}(k) - C^{(n^*)}(j)| \tag{5.3}$$

where C stands for the cumulative density function of the corresponding histograms.

The next step is to eliminate possible boundary artifacts caused by separate tone mappings $T_{\blacklozenge}^{(n)}$ and $T_{\diamondsuit}^{(n)}$ on ♦- and ♢-regions, respectively, as explained in the previous section. Recall that the proposed fuzzy segmentation assigns each pixel (x, y) a likelihood value $w(x, y) \in [0, 1]$, with the pixels around the border of ♦- and ♢-regions being far off from 0 and 1. In the interest of robustness, the input intensity value $I(x, y)$ is finally mapped to $\tilde{I}(x, y)$ via the following affine weighting by the

region likelihood $w(x, y)$ of the pixel:

$$\tilde{I}(x,y) = w(x,y)T_{\blacklozenge}^{(n)}\left(I(x,y)\right) + (1 - w(x,y))T_{\diamondsuit}^{(n)}\left(I(x,y)\right) \tag{5.4}$$

In the same approach to harmonizing time-varying luminance in YYP-degraded video, we can neutralize temporal variations of chrominance, if necessary, by matching the chrominance distributions of all frames to that of the anchor frame n^*. Any two-dimensional chrominance space, such as (U, V) in YUV or (H, S) in HSI, can be used; in this paper we adopt the (U, V) chrominance space. In order for the (U, V) distribution of frame n with mean vector $\mu^{(n)}$ and covariance matrix $\mathbf{\Sigma}^{(n)}$ to match that of the anchor frame n^* with mean vector μ^* and covariance matrix $\mathbf{\Sigma}^*$, we solve the following optimization problem

$$\begin{aligned} &\{\mathbf{A}^{(n)}, \mathbf{t}^{(n)}\} = \\ &\arg\min_{\mathbf{A},\mathbf{t}} \|\mathbf{A}\mu^{(n)} + \mathbf{t} - \mu^*\|_2^2 + \lambda\|\mathbf{A}\mathbf{\Sigma}^{(n)}\mathbf{A}^T - \mathbf{\Sigma}^*\|_F^2. \end{aligned} \tag{5.5}$$

to determine the affine transform $\mathbf{A}^{(n)}$ and the translation vector $\mathbf{t}^{(n)}$ which achieves the best chrominance match between anchor frame n^* and input frame n in a least-squares sense when applying to all pixel chrominance vectors (U, V) of frame n. Fig. 5.3 is an example of performing the matching on images to fit the chrominance distributions of given anchor images by solving (5.5), λ being 0.05.

(a) (b)

(c) (d)

Figure 5.3: Chrominance matching by solving the optimization problem (5.5). (a)(b) Input images; (c) result of matching (a) to the anchor image (b); (d) result of matching (b) to the anchor image (a).

5.4 Contrast enhancement in YYP restoration

Often after the YYP degradation, the representative frame n^* chosen by (5.2) is still of poor visual quality, with ill-shaped histograms $\mathbf{p}_{\blacklozenge}^{(n^*)}$ and $\mathbf{p}_{\lozenge}^{(n^*)}$. In such cases we can and should first enhance the anchor frame n^*, and then drive the YYP harmonization algorithm with the histograms of the enhanced anchor frame. Any contrast enhancement technique can be applied, for example the one presented in Chapter 3.

Finally, the flow of the proposed YYP harmonization method is presented in Fig. 5.4.

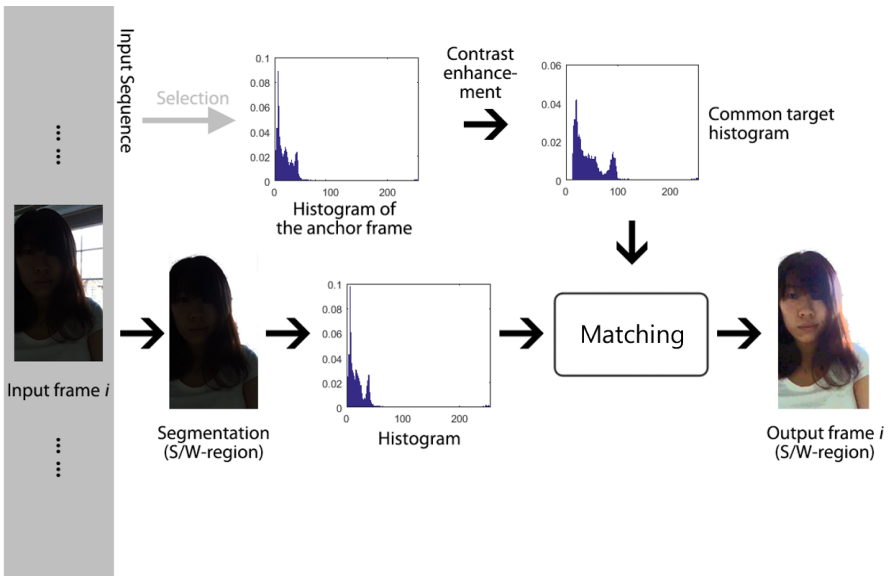


Figure 5.4: Pipeline of the proposed YYP harmonization method.

5.5 Experimental Results

We conducted extensive experiments with the proposed YYP restoration method on videos captured by smartphones and laptops under poor, uneven illumination conditions and with both camera and object motions. Some samples of our experimental results are presented below; more example videos are available as supplementary materials on the internet.

Because the YYP-type of video degradation as identified by this paper has hardly been treated in the literature, there are no previous YYP video restoration methods to compare with. A technique of close spirit to ours is the one on video tonal stabilization by Farbman *et al.* [18]; this work is compared with the proposed YYP harmonization method. Since a simple way of attacking the YYP problem is to perform tone mapping on each input frame, we add to our comparison and evaluation group two single-frame methods: the contrast limited adaptive histogram equalization (CLAHE) [78], which represents the family of histogram transform-based global tone mapping methods; and the enhancement method based on edge-preserving decomposition (EPD) by Farbman *et al.* [19], which represents the family of edge-aware local filter-based tone mapping methods. The video tonal stabilization method also requires an anchor frame; this anchor frame is chosen manually to be an input frame of best visual quality for fair comparison.

In Fig. 5.5 we compare the above four methods and demonstrate how they behave when applied to a YYP-degraded head-and-shoulder video. This scenario of face-to-face video communication is very common in social media (e.g., Apple's FaceTime), and it is highly susceptible to YYP degradation because the video is frequently shot in unfavorable indoor lighting, inexpensive cameras on mobile devices are limited in

optical capability, and users tend to be naive in operating the camera.

The video scene in Figures 5.5 and 5.1 is backlit. The uneven illumination, as explained in the introduction, causes the camera's auto-exposure to flip-flop accompanying the motions of the person or/and camera. The foreground person (♦-region) becomes severely underexposed from time to time, meanwhile the background (♢-region) is also unstable and suffers from overexposure intermittently (see the uploaded video supplementary materials). As shown in Fig. 5.5 and the supplementary video file, the two single-frame methods are ineffective to neutralize the temporal intensity fluctuations, particularly on the face. The method [18] performs much better than the single-frame methods in terms of the temporal consistency in the foreground intensity, but it fails to correct the overexposure problem in the background. The proposed method appears to be more effective and robust than all others; it removes frame-to-frame intensity drifting in both foreground and background and at the same time enhances under- and over-exposed regions, greatly boosting the video quality.

The YYP degradation shown in Fig. 5.6 has a different polarity from that in Fig. 5.5: the background is underlit and underexposed, where the person in the foreground is well lit. The background becomes even darker as the person moves closer to the light source (frame 170 in the figure). This generates serious halo artifacts in CLAHE output. The method [19] fails to bring any temporal consistency to the background intensity. The method [18] stabilizes the time-varying background intensity but in the process it leaves the foreground object overexposed. Again, the proposed method performs noticeably better than others in both ♦-region (the background) and ♢-region (the foreground object), retaining temporal consistency and rich spatial details.

Figure 5.5: Rows: two YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [78]; (c) output by method [19]; (d) output by method [18]; (e) output by the proposed method.

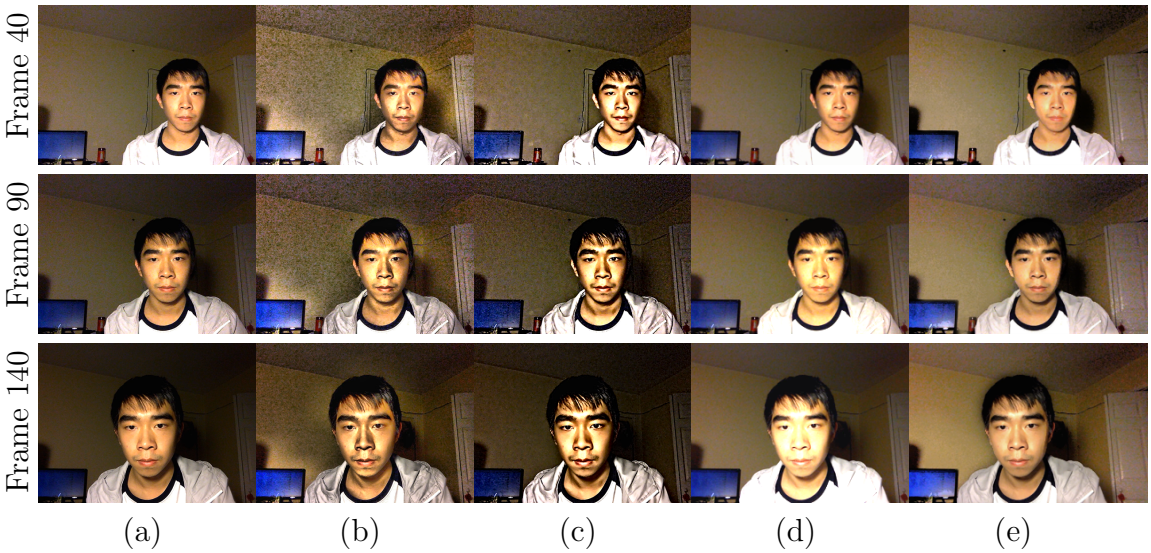


Figure 5.6: Rows: three YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [78]; (c) output by method [19]; (d) output by method [18]; (e) output by the proposed method.

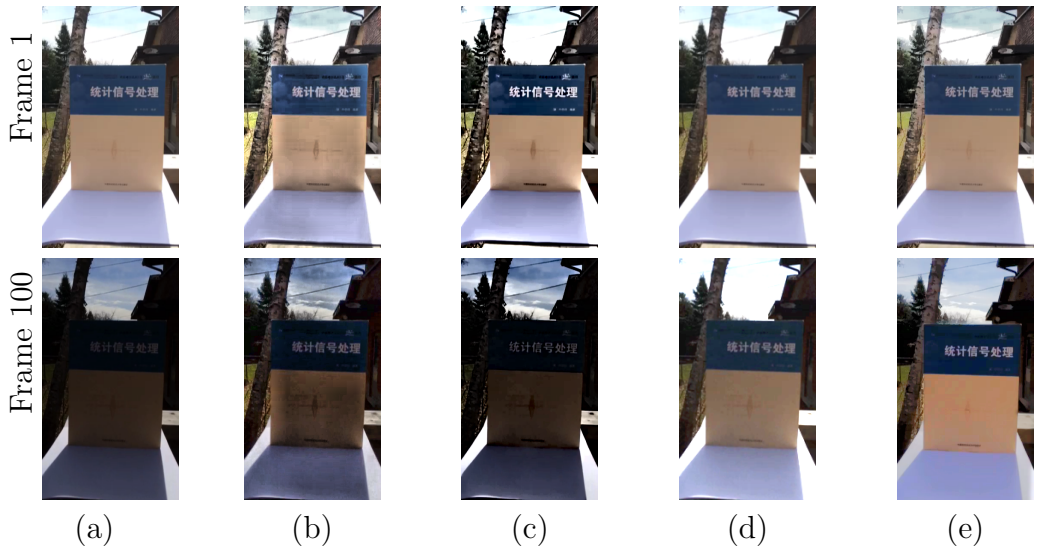


Figure 5.7: Rows: two YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [78]; (c) output by method [19]; (d) output by method [18]; (e) output by the proposed method.

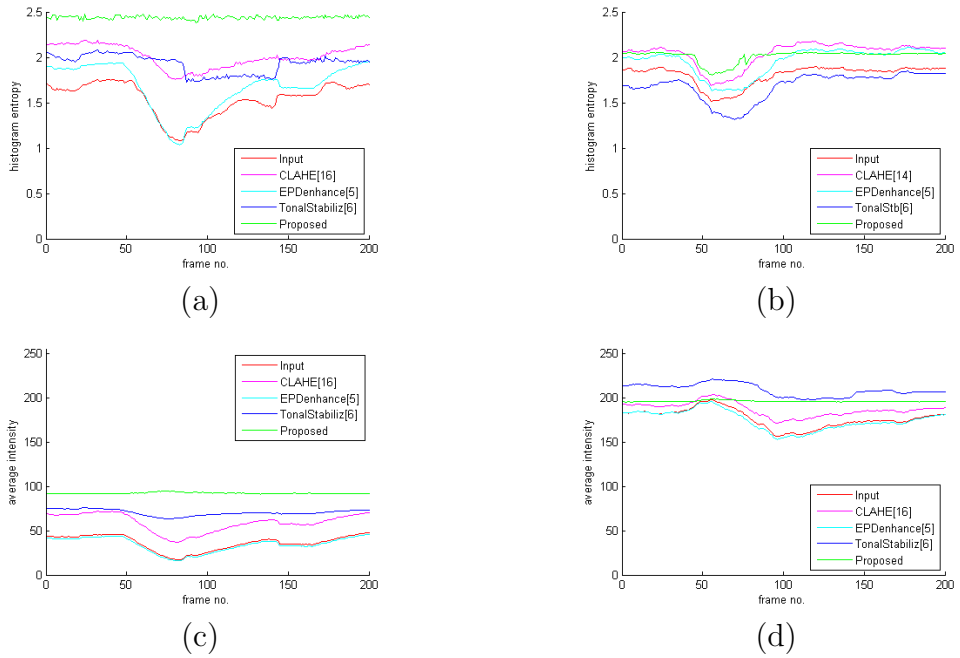


Figure 5.8: Performance comparison of different methods on the video sequence of Fig. 5.1. (a) Entropy of the ♦-region; (b) Entropy of the ♢-region; (c) Average intensity of the ♦-region; (d) Average intensity of the ♢-region.

Fig. 5.7 shows a YYP-degraded video shot in an outdoor situation, together with restored results by the four different methods. Here the YYP phenomenon happens when the camera focus moves from the book (frame 1) to the sky (frame 100). The two single-frame methods fail to correct the underexposure problem in the foreground (frame 100) and they leave the drastic intensity changes in time largely uncompensated. The method [18], on the other hand, does a better job in the temporal consistency of overall intensity; but it makes the sky severely overexposed, wiping out details such as the clouds. In comparison, the proposed method effectively mitigates the underexposure problem in the ♦-region (the book) without overexposing the ♢-region (the sky); its restored video has a more steady tone reproduction largely immune to camera motion.

To further validate the efficacy of the proposed YYP harmonization method, we evaluate it and its alternatives in two objective metrics as well. The first metric is the entropy of the restored video. The entropy can measure both the temporal consistency and the detail richness of the restored video. In Fig. 5.8 we plot separately the entropies of ♦-regions and ♢-regions that are restored by the four different methods. The plotted curves step through all frames of the video scene in Fig. 5.1 to demonstrate the temporal behaviors of the different methods. As being evident in the figure, the proposed method has the highest entropy in ♦-regions by a significant margin and almost ties for the highest entropy in ♢-regions, corroborating our empirical findings that the proposed method reproduces richer details than other methods. Moreover, the proposed method has nearly flat entropy curves, whereas other methods have much varied entropy values in time. This distinction is also clear by the second objective metric: the average frame luminance (Fig. 5.8(c)(d)). The

restored video by the proposed method keeps a nearly constant luminance in both ♦- and ♢-region, while the two single-frame methods suffer from intensity fluctuations, and the tonal stabilization method sits in between.

5.6 Conclusion

We address in this chapter a common yet understudied video degradation problem named YYP, characterized by involuntary and dramatic flip-flop in the luminance and chrominance of an object as the video plays. We investigate the problem and propose a video restoration technique to suppress YYP artifacts and retain temporal consistency of object appearance via inter-frame, spatially-adaptive, optimal tone mapping. Experimental results are encouraging, pointing to an effective, practical solution to degradations caused by YYP.

Chapter 6

Conclusion

Recent years have witnessed a rapid development in digital cameras. However, existing imaging technologies are still far from being able to guarantee high quality output in every condition, especially for poorly or unevenly illuminated scenes. Given an image of non-ideal lighting, algorithmically recovering the details of the scene by the means of image enhancement is necessary. Existing image enhancement methods however often introduce various observable artifacts, adversely affecting the naturalness of the contrast enhanced image. One of the most annoying but often overlooked artifacts is chrominance distortion, which is caused by the seemingly reasonable practice of enhancing the luminance channel without touching the chrominance channels. In this research, we analyze this common flaw in existing image enhancement techniques and propose the optimal clipping scheme for reducing the chrominance distortion artifacts. Furthermore we propose a novel optimization-based image enhancement framework, which generates optimal tone mapping by maximizing global contrast gain with constrained tone and chrominance distortions. Additionally, we devise a highly efficient

algorithm to solve the optimization problem using dynamic programming. Experimental results demonstrate the remarkable output perceptual quality of the proposed method against severe underexposure and/or overexposure.

Based on the enhancement framework above, we propose a new method that restores images acquired in more challenging backlit scenes. The idea of the method is to fuse our enhancement method with image segmentation. The segmentation is driven by feature-based classifiers constructed by supervised learning from our backlit image dataset. Our experiment shows the superior performance of the proposed method over several state-of-the-art methods in terms of segmentation accuracy and restoration results on backlit photographs.

We also address the problem of Yin-Yang Phasing (YYP), a common but largely untreated video degradation problem characterized by involuntary and severe inconsistency in the exposure of objects between frames. We propose a video restoration method based on inter-frame, spatially-adaptive optimal tone mapping, to suppress YYP artifacts; it can maintain temporal consistency of luminance and chrominance of surfaces at the same time. Experimental results demonstrate significant performance gains of the new approach over existing video stabilization methods.

The above two restoration methods are examples of extending a context-free enhancement method to gain more desired properties such as spatial adaptivity and temporal consistency without causing artifacts. These extensions follow the same philosophy: 1) the ranking of pixel luminance can be changed, if done properly, to further increase the output contrast; 2) the adjustment (tone mapping) on pixels under the same lighting condition should be consistent. Following the same philosophy, the optimization-based enhancement method in Chapter 3 can be further improved

by combining more cutting-edge image segmentation algorithms, especially semantic segmentation based on deep learning. This is a promising research direction for future studies in the area of image and video enhancement.

Bibliography

[1] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **34**(11), 2274–2282.

[2] Agaian, S., Silver, B., and Panetta, K. (2007). Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *Image Processing, IEEE Transactions on*, **16**(3), 741–758.

[3] Arici, T., Dikbas, S., and Altunbasak, Y. (2009). A histogram modification framework and its application for image contrast enhancement. *IEEE Trans. Image Process.*, **18**(9), 1921–1935.

[4] Aydin, T. O., Stefanoski, N., Croci, S., Gross, M., and Smolic, A. (2014). Temporally coherent local tone mapping of hdr video. *ACM Transactions on Graphics (TOG)*, **33**(6), 196.

[5] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236.

[6] Celik, T. and Tjahjadi, T. (2011). Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, **20**(12), 3431–3441.

[7] Cho, S. and Lee, S. (2009). Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, page 145. ACM.

[8] Cover, T. M. and Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

[9] Dahlman, E., Oestges, C., Bovik, A. C., Fette, B. A., Jack, K., Dowla, F., Parkvall, S., Skold, J., DeCusatis, C., da Silva, E., *et al.* (2009). *Communications engineering desk reference*. Academic Press.

[10] Debevec, P. E. and Malik, J. (2008). Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM.

[11] Deng, X., Shen, Y., Wu, X., and Zhao, L. (2017). A model-based approach for human head-and-shoulder segmentation. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 3315–3319. IEEE.

[12] Dippel, S., Stahl, M., Wiemker, R., and Blaffert, T. (2002). Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. *IEEE Transactions on medical imaging*, **21**(4), 343–353.

[13] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003). Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum*, volume 22, pages 419–426. Wiley Online Library.

[14] Duan, J. and Qiu, G. (2004). Novel histogram processing for colour image enhancement. In *Multi-Agent Security and Survivability, 2004 IEEE First Symposium on*, pages 55–58.

[15] Duda, R. O., Hart, P. E., and Stork, D. G. (2012). *Pattern classification*. John Wiley & Sons.

[16] Eilertsen, G., Wanat, R., Mantiuk, R. K., and Unger, J. (2013). Evaluation of tone mapping operators for hdr-video. In *Computer Graphics Forum*, volume 32, pages 275–284. Wiley Online Library.

[17] Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.

[18] Farbman, Z. and Lischinski, D. (2011). Tonal stabilization of video. *ACM Transactions on Graphics (TOG)*, **30**(4), 89.

[19] Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. (2008). Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM.

[20] Fattal, R., Agrawala, M., and Rusinkiewicz, S. (2007). Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph.*, **26**(3), 51.

[21] Fossum, E. R. (1998). Digital camera system on a chip. *IEEE micro*, **18**(3), 8–15.

[22] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., and Paisley, J. (2016a). A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, **129**, 82–96.

[23] Fu, X., Zeng, D., Huang, Y., Zhang, X.-P., and Ding, X. (2016b). A weighted

variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790.

[24] Gonzalez, R. C. (2009). *Digital image processing*. Pearson Education India.

[25] Guo, R., Dai, Q., and Hoiem, D. (2013). Paired regions for shadow detection and removal. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **35**(12), 2956–2967.

[26] Han, J.-H., Yang, S., and Lee, B.-U. (2011). A novel 3-d color histogram equalization method with uniform 1-d gray scale histogram. *Image Processing, IEEE Transactions on*, **20**(2), 506–512.

[27] Hanmandlu, M., Verma, O., Kumar, N., and Kulkarni, M. (2009). A novel optimal fuzzy system for color image enhancement using bacterial foraging. *Instrumentation and Measurement, IEEE Transactions on*, **58**(8), 2867–2879.

[28] Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, **67**(5), 786–804.

[29] He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **33**(12), 2341–2353.

[30] He, K., Sun, J., and Tang, X. (2013). Guided image filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **35**(6), 1397–1409.

[31] Huang, K., Wang, Q., and Z, W. (2006). Natural color image enhancement and

evaluation algorithm based on human visual system. *Computer Vision and Image Understanding*, **103**(1), 52–63.

[32] Im, J., Yoon, I., Hayes, M. H., and Paik, J. (2013). Dark channel prior-based spatially adaptive contrast enhancement for back lighting compensation. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 2464–2468. IEEE.

[33] ITU-R Recommendation BT.709-6 (2015). Parameter values for the hdtv standards for production and international programme exchange. *ITU Radiocommunication*.

[34] Jaya, V. and Gopikakumari, R. (2013). Iem: a new image enhancement metric for contrast and sharpness measurements. *International Journal of Computer Applications*, **79**(9).

[35] Jobson, D. J., Rahman, Z.-U., and Woodell, G. A. (1997a). A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, **6**(7), 965–976.

[36] Jobson, D. J., Rahman, Z.-u., and Woodell, G. A. (1997b). Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, **6**(3), 451–462.

[37] Kaiser, P. K. and Boynton, R. M. (1996). Human color vision.

[38] Katoh, N., Ito, M., and Ohno, S. (1999). Three-dimensional gamut mapping using various color difference formulae and color spaces. *Journal of Electronic Imaging*, **8**(4), 365–380.

[39] Khan, S. H., Bennamoun, M., Sohel, F., and Togneri, R. (2014). Automatic feature learning for robust shadow detection. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1939–1946. IEEE.

[40] Kolb, H. (2011). Simple anatomy of the retina.

[41] Kolmogorov, V. and Zabin, R. (2004). What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **26**(2), 147–159.

[42] Lafferty, J., McCallum, A., and Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

[43] Lee, S., Kim, N., and Paik, J. (2015). Adaptively partitioned block-based contrast enhancement and its application to low light-level video surveillance. *SpringerPlus*, **4**(1), 1–11.

[44] Li, Z. and Wu, X. (2014). Contrast enhancement with chromaticity error bound. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 4507–4511. IEEE.

[45] Li, Z. and Wu, X. (2018). Learning-based restoration of backlit images. *IEEE Transactions on Image Processing*, **27**(2), 976–986.

[46] Li, Z., Cheng, K., and Wu, X. (2015). Soft binary segmentation-based backlit image enhancement. In *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, pages 1–5. IEEE.

[47] Luo, M. R., Cui, G., and Rigg, B. (2001). The development of the cie 2000

colour-difference formula: Ciede2000. *Color research and application*, **26**(5), 340–350.

[48] Mahmoudi, M. and Sapiro, G. (2005). Fast image and video denoising via non-local means of similar neighborhoods. *Signal Processing Letters, IEEE*, **12**(12), 839–842.

[49] Matsushita, Y., Ofek, E., Ge, W., Tang, X., and Shum, H.-Y. (2006). Full-frame video stabilization with motion inpainting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **28**(7), 1150–1163.

[50] Motoyoshi, I., Nishida, S., Sharan, L., and Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, **447**(7141), 206–209.

[51] Naik, S. K. and Murthy, C. A. (2003). Hue-preserving color image enhancement without gamut problem. *IEEE Trans. Image Process.*, **12**(12), 1591–1598.

[52] Nakamura, J. (2017). *Image sensors and signal processing for digital still cameras*. CRC press.

[53] Panagopoulos, A., Wang, C., Samaras, D., and Paragios, N. (2012). Estimating shadows with the bright channel cue. In *Trends and Topics in Computer Vision*, pages 1–12. Springer.

[54] Panetta, K., Zhou, Y., Agaian, S., and Jia, H. (2011). Nonlinear unsharp masking for mammogram enhancement. *IEEE Transactions on Information Technology in Biomedicine*, **15**(6), 918–928.

[55] Paris, S., Hasinoff, S. W., and Kautz, J. (2011). Local laplacian filters: edge-aware image processing with a laplacian pyramid. *ACM Trans. Graph.*, **30**(4), 68.

[56] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, **39**(3), 355–368.

[57] Sara, J. J. (1984). *The automated reproduction of pictures with nonreproducible colors*. Ph.D. thesis, Massachusetts Institute of Technology.

[58] Shimizu, S., Kondo, T., Kohashi, T., Tsurata, M., and Komuro, T. (1992). A new algorithm for exposure control based on fuzzy logic for video cameras. *IEEE Transactions on Consumer Electronics*, **38**(3), 617–623.

[59] Shu, X. and Wu, X. (2013). Image enhancement revisited: From first order to second order statistics. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 886–890. IEEE.

[60] Shu, X. and Wu, X. (2018). Locally adaptive rank-constrained optimal tone mapping. *ACM Transactions on Graphics (TOG)*.

[61] Siu, W.-C. and Hung, K.-W. (2012). Review of image interpolation and super-resolution. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–10. IEEE.

[62] Smith, K., Krawczyk, G., Myszkowski, K., and Seidel, H.-P. (2006). Beyond tone

mapping: Enhanced depiction of tone mapped hdr images. In *Computer Graphics Forum*, volume 25, pages 427–438. Wiley Online Library.

[63] Sobol, R. (2004). Improving the retinex algorithm for rendering wide dynamic range photographs. *Journal of Electronic Imaging*, **13**(1), 65–75.

[64] Stark, J. A. (2000). Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.*, **9**(5), 889–896.

[65] Stone, M. C., Cowan, W. B., and Beatty, J. C. (1988). Color gamut mapping and the printing of digital color images. *ACM Transactions on Graphics (TOG)*, **7**(4), 249–292.

[66] Szeliski, R. (2010). *Computer vision: algorithms and applications*. Springer Science & Business Media.

[67] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE.

[68] Trahanias, P. and Venetsanopoulos, A. (1992). Color image enhancement through 3-d histogram equalization. In *Pattern Recognition, 1992. Vol.III. Conference C: Image, Speech and Signal Analysis, Proceedings., 11th IAPR International Conference on*, pages 545–548.

[69] Tsai, C.-M. and Yeh, Z.-M. (2010). Contrast compensation by fuzzy classification and image illumination analysis for back-lit and front-lit color face images. *Consumer Electronics, IEEE Transactions on*, **56**(3), 1570–1578.

[70] Wang, S., Zheng, J., Hu, H.-M., and Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, **22**(9), 3538–3548.

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

[72] Wu, X. (2011). A linear programming approach for optimal contrast-tone mapping. *IEEE Transactions on Image Processing*, **20**(5), 1262–1272.

[73] Wu, X., Li, Z., and Deng, X. (2015). Video restoration against yin-yang phasing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 549–557.

[74] Yang, C. C. and Kwok, S. H. (2003). Efficient gamut clipping for color image processing using lhs and yiq. *Optical Engineering*, **42**(3), 701–712.

[75] Yuan, L. and Sun, J. (2012). Automatic exposure correction of consumer photographs. In *Proceedings of the 12th European conference on Computer Vision-Volume Part IV*, pages 771–785. Springer-Verlag.

[76] Zhang, Y., Wu, L., Wang, S., and Wei, G. (2010). Color image enhancement based on hvs and pcnn. *Science China Information Sciences*, **53**(10), 1963–1976.

[77] Zhu, J., Samuel, K. G. G., Masood, S. Z., and Tappen, M. F. (2010). Learning to recognize shadows in monochromatic natural images. In *Computer Vision and Pattern Recognition*, pages 223–230.

[78] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc.