Cross Dynamic Range And Cross Resolution Objective Image Quality Assessment With Applications

by

Hojatollah Yeganeh

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2014

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

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

Hojatollah Yeganeh

Abstract

In recent years, image and video signals have become an indispensable part of human life. There has been an increasing demand for high quality image and video products and services. To monitor, maintain and enhance image and video quality objective image and video quality assessment tools play crucial roles in a wide range of applications throughout the field of image and video processing, including image and video acquisition, communication, interpolation, retrieval, and displaying. A number of objective image and video quality measures have been introduced in the last decades such as mean square error (MSE), peak signal to noise ratio (PSNR), and structural similarity index (SSIM). However, they are not applicable when the dynamic range or spatial resolution of images being compared is different from that of the corresponding reference images. In this thesis, we aim to tackle these two main problems in the field of image quality assessment.

Tone mapping operators (TMOs) that convert high dynamic range (HDR) to low dynamic range (LDR) images provide practically useful tools for the visualization of HDR images on standard LDR displays. Most TMOs have been designed in the absence of a well-established and subject-validated image quality assessment (IQA) model, without which fair comparisons and further improvement are difficult. We propose an objective quality assessment algorithm for tone-mapped images using HDR images as references by combining 1) a multi-scale signal fidelity measure based on a modified structural similarity (SSIM) index; and 2) a naturalness measure based on intensity statistics of natural images. To evaluate the proposed Tone-Mapped image Quality Index (TMQI), its performance in several applications and optimization problems is provided. Specifically, the main component of TMQI known as structural fidelity is modified and adopted to enhance the visualization of HDR medical images on standard displays. Moreover, a substantially different approach to design TMOs is presented, where instead of using any pre-defined systematic computational structure (such as image transformation or contrast/edge enhancement) for tone-mapping, we navigate in the space of all LDR images, searching for the image that maximizes structural fidelity or TMQI.

There has been an increasing number of image interpolation and image super-resolution

(SR) algorithms proposed recently to create images with higher spatial resolution from low-resolution (LR) images. However, the evaluation of such SR and interpolation algorithms is cumbersome. Most existing image quality measures are not applicable because LR and resultant high resolution (HR) images have different spatial resolutions. We make one of the first attempts to develop objective quality assessment methods to compare LR and HR images. Our method adopts a framework based on natural scene statistics (NSS) where image quality degradation is gauged by the deviation of its statistical features from NSS models trained upon high quality natural images. In particular, we extract frequency energy falloff, dominant orientation and spatial continuity statistics from natural images and build statistical models to describe such statistics. These models are then used to measure statistical naturalness of interpolated images. We carried out subjective tests to validate our approach, which also demonstrates promising results. The performance of the proposed measure is further evaluated when applied to parameter tuning in image interpolation algorithms.

Acknowledgements

First and foremost, I thank God almighty for giving me grace and privilege to pursue my dream and his blessing in giving me life, health and intelligence.

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me. It is my great pleasure to take this opportunity to thank them.

Above all, I would like to express my deep appreciation and gratitude to my tremendous supervisor, Professor Zhou Wang. This thesis would not have been possible without the priceless help, support and engagement of him during my research carrier. His knowledge and his supportive supervision always gave me encouragements, and I am very grateful for being a member of his research group.

Thank you to the members of my committee, Dr. Oleg Michailovich, Dr. Pin-Han Ho, Dr. Edward Vrscay, and Dr. Tiago Falk. I express my sincere gratitude for your help, cooperation, and reviewing this thesis.

The University of Waterloo offers a rich and productive environment to explore new ideas. I am grateful to have the chance to study in the middle of a greatly supportive community and be surrounded by wonderful colleagues. In particular I would like to thank Dr. Edward Vrscay for his invaluable support and cooperation, and my great colleagues Kai Zeng, Abdul Rehman, Nima Nikvand, Mohammad Rostami, Kede Ma, and Jiheng Wang to name but a few.

Last but not the least, I would like to thank my wife Bahareh for her personal support and great patience at all times, and I lovingly dedicate this thesis to her. My family has been an integral part of my academic life. Despite being many thousands of kilometers away, their support and constant courage gave me the strength to persevere me through these many years as a student. There are no words that can express my gratitude and appreciation for all you have done and been for me.

Dedication

I would like to dedicate this thesis to my parents, and my wife for their endless love, limitless encouragement, and sacrifice throughout my PhD.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ARSD	Adaptive autoRegression and Soft-Decision estimation
CT	Computerized Tomography
DFDF	Directional Filtering and Data Fusion
DICOM	Digital Imaging and Communications in Medicine
FR	Full-Reference
GGD	Generalized Gaussian Distribution
HDR	High Dynamic Range
HDTV	High Definition TeleVision
HR	High Resolution
HVS	Human Visual System
ICBI	Iterative Curvature-Based Interpolation
IND	Interpolated Natural image Distortion measure
IQA	Image Quality Assessment
IW − SSIM	Information-Weighted Structural SIMilarity index
JND	Just Noticeable Difference
KRCC	Kendall Rank Correlation Coefficient
LDR	Low Dynamic Range

LPF	Low Pass Filter
LR	Low Resolution
MOS	Mean Opinion Score
MS − SSIM	Multi-Scale Structural SIMilarity index
MSE	Mean Square Error
MSF	Maximum Structural Fidelity
NARM	Nonlocal AutoRegressive Modeling
NEDI	New Edge-Directed Interpolation
NIQE	Natural Image Quality Evaluator
NN	Nearest Neighbour
NR	No-Reference
NSS	Natural Scene Statistics
PDF	Probability Density Function
PLCC	Pearson's Linear Correlation Coefficient
PSNR	Peak Signal to Noise Ratio
RR	Reduced-Reference
SAE	Sum of Absolute Difference
SDR	Standard Display Range
SR	Super-Resolution
SRCC	Spearman's Rank Correlation Coefficient
SSIM	Structural SIMIlarity index
SVD	Singular Value Decomposition
TAFC	Two-Alternative Forced Choice

TMO	Tone-Mapping Operator
TMQI	Tone-Mapped image Quality Index
TVI	Threshold Versus Intensity
VDP	Visual Difference Predictor
VIF	Visual Information Fidelity measure
VQA	Video Quality Assessment
WIND	Weighted Interpolated Natural image Distortion measure

Chapter 1

Introduction

1.1 Motivation

With the advances of image acquisition and display devices and communication networks, digital image and video products and services have become ubiquitous, ranging from personal cameras and visual communications to remote sensing and medical imaging. Since quality has always been among the most important performance measures, there has been a drastically increasing interest in accurate, reliable and practical Image Quality Assessment (IQA) and Video Quality Assessment (VQA) tools.

According to the availability of a reference image, objective IQA metrics can be divided into three main categories: full-reference (FR), no-reference (NR), and reduced-reference (RR) methods [18]. In order to evaluate the quality of a distorted image, FR IQA methods always assume full access to the original image, and thus FR methods usually provide the most reliable and accurate evaluation results. A number of successful algorithms have been proposed to predict human visual perception presuming that the pristine reference signal is available. Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), Visual Information Fidelity (VIF) [19], Structural SIMilarity index (SSIM) [20], and its derivative Multi-Scale SSIM [21] are among the best known FR IQA methods, and they have attracted significant attention in recent years. However, these methods cannot be applied in

many present and emerging practical real-world applications where the reference signal is unavailable or is in a different acquisition format or spatial resolution.

High Dynamic Range (HDR) images have a greater dynamic range of luminance levels between its brightest and darkest regions than standard dynamic range (SDR) or low dynamic range (LDR) images [5, 22]. Such a large dynamic range enables more accurate representations of the intensity levels in the real scene. Nevertheless, HDR images cannot be visualized on regular displays without the aid of Tone-Mapping Operators (TMO) that convert high dynamic range (HDR) to low dynamic range (LDR) images. Tone-mapping algorithms reduce the dynamic range of HDR images, and are able to preserve most of the details captured in the original HDR images. Figure 1.1 demonstrates the advantage of tone-mapping HDR images where the images of two scenes were taken multiple times with slightly different exposure settings, which may be subsequently merged to an HDR image. On the left, images taken with the best exposure are illustrated. Due to the existence of both light and dark areas in the same scene, even the best exposure shots fail to capture the detail and color appearance of the sky in the background and the bricks in the foreground. Apparently the sky in both scenes is overexposed and at the same time the foreground is dark. The same scenes shown on the right were captured using HDR format and were mapped by TMOs. It can be observed that the structural details are much better preserved in both the background and the foreground.

Since the original HDR images cannot be visualized on standard displays, subjective evaluation may not provide a golden standard to validate the performance of TMOs in terms of preserving structural details. Therefore, without a reliable objective quality measure for tone-mapped images, the design of traditional TMOs can only be based on intuitive ideas, lacking clear goals for further improvement. Although a number of TMOs have been proposed in recent years, little has been done in developing objective methods to assess and compare the quality of TMOs. The challenging task is due to the fact that the dynamic range of HDR and LDR images are different, and thus existing IQA methods cannot be applied.

Modern communication devises enable users to adaptively modify visual signals for better visualization purposes. We often need to view images or videos at a different resolution

Figure 1.1: Optimally exposed images (left) versus HDR tone mapped images (right).(Image courtesy of Reinhards book [3])

from the resolution of the original content. High resolution images should be decimated to fit on small displays. For example, a video is downsampled if it is originally produced in HDTV resolution, but watched on smart phones with lower resolutions. On the other hand, interpolation and super-resolution algorithms are used to visualize low resolution content

on high resolution screens. Therefore, measuring the performance of resizing algorithms plays an important role in the field of visual communication. Typical FR IQA methods fail to estimate the quality of interpolated images because the reference and the distorted images have different spatial resolution. Consequently how to evaluate the performance of interpolation algorithms is a challenging but important problem.

1.2 Objectives

The main objective of this thesis is to develop automatic quality assessment algorithms for images undergoing changes in dynamic range and spatial resolution. Furthermore, we aim to explore the potentials of these quality assessment algorithms in real-world applications.

1.3 Contributions

The main contributions of this thesis are summarized as follows:

- An objective quality assessment algorithm for tone-mapped images is presented. The method is built upon combining a multi-scale signal fidelity measure on the basis of a modified structural similarity index (SSIM) and a naturalness measure on the basis of intensity statistics of natural images. Validations using independent subject-rated image databases show good correlations between subjective ranking score and the proposed Tone-Mapped image Quality Index (TMQI). Furthermore, we demonstrate the extended applications of TMQI using two examples: parameter tuning for TMOs and adaptive fusion of multiple tone-mapped images

- Medical images are typically captured and stored using formats that allocate more bits to each pixel than those assumed by standard displays. As such, they are high dynamic range (HDR) images. To visualize HDR medical images, a so-called "windowing" procedure is typically employed by which the structural details within the intensity region of interest is mapped to the dynamic range of regular displays. A

parameter selection scheme is proposed to adaptively find the optimal windowing function for different medical images.

- We have developed a novel framework in designing new TMOs by making use of either the structural fidelity component in TMQI or TMQI itself as the objective. Unlike the design of traditional TMOs, we do not start from a predefined computational structure. Instead, we explicitly treat tone-mapping as an optimization problem in the space of images and propose an iterative search approach that starts from any initial image and moves step-by-step in the image space towards the direction of improving the proposed objective measure until a (local) maximal point is reached.

- We have proposed an NSS-based distortion measure to objectively assess the quality of interpolation algorithms. Our NSS model uses statistics of three features trained from high-quality natural HR images, and uses LR images as reference. The proposed method can be used when the interpolation scaling factor is an integer number. A subjective experiment was conducted to validate the performance of the quality measure where subjects were asked to rate images created from different interpolation algorithms. The results have demonstrated that the proposed objective measure well predicts subjective ratings. Moreover, to demonstrate a potential application of such IQA methods, the model is adopted for tuning the parameters of an existing image interpolation algorithm.

1.4 Thesis Organization

The organization of the rest of the thesis is as follows. Chapter 2 discusses related existing works on the topics addressed in the thesis. We briefly introduce HDR images and tone-mapping operators. Moreover, the details of well-known relevant IQA models are provided. The fundamentals of interpolation algorithms and related perceptual quality assessment methods are discussed. An overview of previous relevant works to the thesis is presented throughout the chapter. Chapter 3 introduces tone-mapped image quality index (TMQI)

as a novel objective quality measure for tone-mapped images, and shows its potential applications in parameter tuning and adaptive fusion of tone-mapping operators. A modified version of the proposed measure is described in Chapter 4 to enhance the visualization of HDR medical images. In Chapter 5, a perceptual optimization frameworks based on the proposed TMQI and structural fidelity measures is presented to obtain better tone-mapping operators for natural images. Chapter 6 focuses on developing IQA method for interpolation algorithms. Finally, Chapter 7 provides concluding remarks and discusses future research directions.

Chapter 2

Background

2.1 High Dynamic Range Image

High dynamic range image (HDR) can store a much wider gamut than standard 24-bit RGB where the range of intensity levels could be on the order of 10,000(cd/m^2) to 1(cd/m^2) [5, 22]. The difference between HDR and low dynamic range (LDR) images is more than the bit depth and the difference between minimum and maximum intensity values. In fact, HDR format enable us to store a wide range of luminance with much higher precision than LDR images do, and thus represent information more accurately than low dynamic range images. Typically, low dynamic range image is categorized as an *output-referred* standard since its luminance levels are associated with the dynamic range of the display device. On the contrary, most HDR images are *scene-referred* as the high dynamic range allows us for accurate representations of the luminance variations in real scenes, ranging from direct sunlight to faint starlight [5].

2.1.1 Applications of HDR Images

With recent advances in imaging and computer graphic technologies, the HDR images are becoming more widely available. Camera companies are already developing scene-referred

data by providing tools to combine images captured in different exposures. Moreover, there are some applications in which individuals need to record more than what is seen of a scene with naked eyes, and the HDR format is able to accomplish this goal. Some application areas of HDR images are outlined as follows:

- **Digital Photography:** Taking a single photo from a natural scene with high contrast in both very bright and very dark regions is still a challenging technical problem. Instead, photographers can capture multiple images at different exposures using conventional cameras and then merge the images into a single HDR image. While none of the images being physically captured contains all structural details in all regions, a transformation that mapps HDR to LDR images may be able to reproduce almost all the details. This transformation is called a "tone-mapping operation" and is a topic of major interest in this thesis.

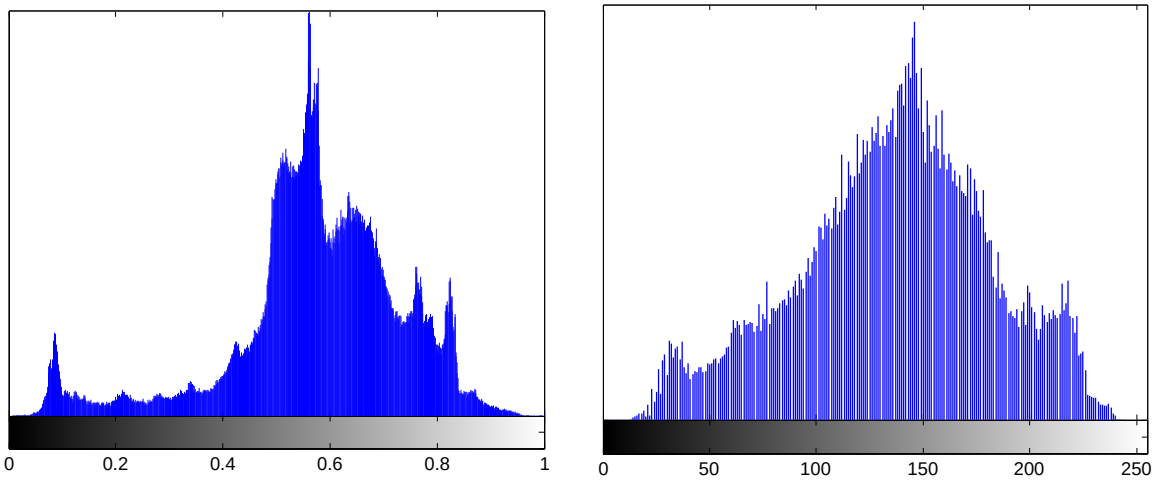
- **Remote Sensing:** Satellite imagery typically contains pixel intensities much more than what is visible to the naked eyes, where the images with different wavelengths may be combined together into a single HDR image [23].

- **Medical imaging:** DICOM standard is widely used in medical communities where the precision of intensity levels of certain image modality may occupy up to 16 bits per pixel. Therefore, DICOM is essentially a HDR image format. How to visualize DICOM images on standard displays is a practically important problem.

- **Computer Game:** Game engines are applying image-based rendering techniques and tone mapping algorithm to generate natural scenes, where new standards are emerging for HDR images as a critical element in this pipeline [24].

2.2 Tone Mapping Operators

Using different techniques mentioned above, we are able to obtain HDR images. However, most available display devices can only show images with moderate dynamic range



(a) (b)

Figure 2.1: Histograms of (a) "Office" HDR image and (b) its corresponding LDR image using TMO in [4].

of less than $100(cd/m^2)$ to $1(cd/m^2)$. The discrepancy between the wide luminance range of the HDR image and the small ranges reproduced by standard displays raises a common practical problem : how to visualize HDR images on standard display devices. A straightforward solution is simply linearly mapping the intensity levels of HDR images to that of the standard displays, but such a mapping often removes the structural information in low intensity regions, resulting in almost entirely dark image [25]. Thus, the ultimate goal of tone-mapping is to visually match between the observed scene and the tone-mapped LDR images, whereas simple linear scaling inevitably causes information loss. Figure 2.1 illustrates the histogram of "Office" HDR image and the histogram of its tone-mapped LDR image using the TMO given in [4]. It can be observed that tone mapping operation modifies the relative intensity of pixels in bright and dark regions while linear mapping results in the same histograms for HDR and LDR images. Many other tone-mapping algorithms have been proposed to solve this problem. Basically there are two main categories of tone-mapping algorithms: global and local tone mapping operators. Global TMOs map all the image pixel values to a display value without taking into consideration the spatial location of the pixel in question. The mapping function can be a gamma function, a power

function, a logarithmic function or a function derived according to certain characteristics of the Human Visual System (HVS). On the other hand, local mapping operator is spatially dependent and varying transformations are applied to each pixel depending on its surrounding image structures. This section describes typical methodologies to resolve the tone reproduction issue. A comprehensive review on TMOs can be found in [5, 22].

2.2.1 Human Visual Adaptation

The HVS needs to deal with the same issue. Although the dynamic range of individual channel in the visual pathway is limited, it gives us the ability to perceive the details during the course of day and night. Therefore, imitating the adaptation characteristics of HVS may help us solve the tone-mapping problem. There are different approaches to model the human adaptation behaviour. The most relevant ones include the threshold versus intensity function and the photoreceptor response model.

In psychophysical studies, human visual adaptation is evaluated by the minimum amount of incremental light that allows observers to detect an object from the background luminance. This minimum is called "Just Noticeable Difference" or JND. Figure 2.2 plots the JND versus various background intensity. This curve is also called a Threshold Versus Intensity (TVI) function. It can be seen that over a wide range of background intensity, the ratio $\frac{\Delta I}{I}$ is nearly constant. This relation is known for more than 140 years as the Weber's law. The Weber's law shows that human visual adaptation system scales scene intensities to preserve our ability to detect contrasts within a wide range of intensities.

Human eyes accomplish visual adaptation to the varying range of illumination with the coordination of the pupil, the rod-cone cells and the photoreceptor mechanism. The photoreceptor cells including rods and cones which convert the absorbed light energy into neural responses which have been measured with photoreceptor mechanism. While the visual system performs over a broad range of background light intensities, the photoreceptor rod-cone cells respond logarithmically to a narrow range of luminance. This range is only about three log units as shown in Figure 2.3. The shapes of the response curves of the cones

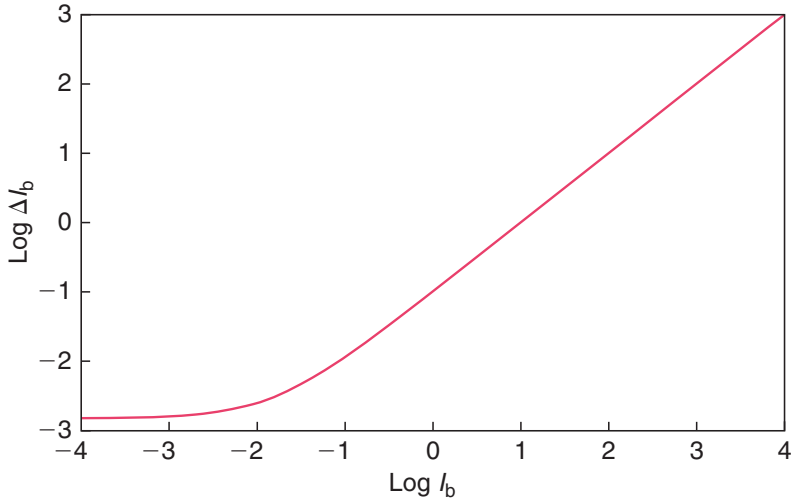


Figure 2.2: Plot of visual threshold ΔI_b versus intensity I_b, TVI [5].

and rods are the same; however, since rod cells are more sensitive to light, the response curve for cone is shifted to the right, as shown in Figure 2.3.

The response curve can be fitted with the following equation,

$$\frac{R}{R_{max}} = \frac{I^n}{I^n + \sigma^n}, \tag{2.1}$$

where R is the photoreceptor response, R_{max} is the maximum response, I is the light intensity, and σ is a semisaturation constant, which corresponds to the intensity causing half-maximum response. The role of σ in (2.1) is to control the position of the response curve on the horizontal intensity axis, and thus it is possible to represent the response curves of rods and cones by simply using two different values of σ. Finally, n is called sensitivity-control exponent whose value is in the range of 0.7 and 1 [6]. The response curve explains when a photoreceptor is exposed to relatively high intensity with respect to the background luminance, the response reaches its maximum and the photoreceptor is saturated. In other words, the photoreceptor loses sensitivity to any additional light intensity. However, this saturation does not continue for long. When the HVS is exposed

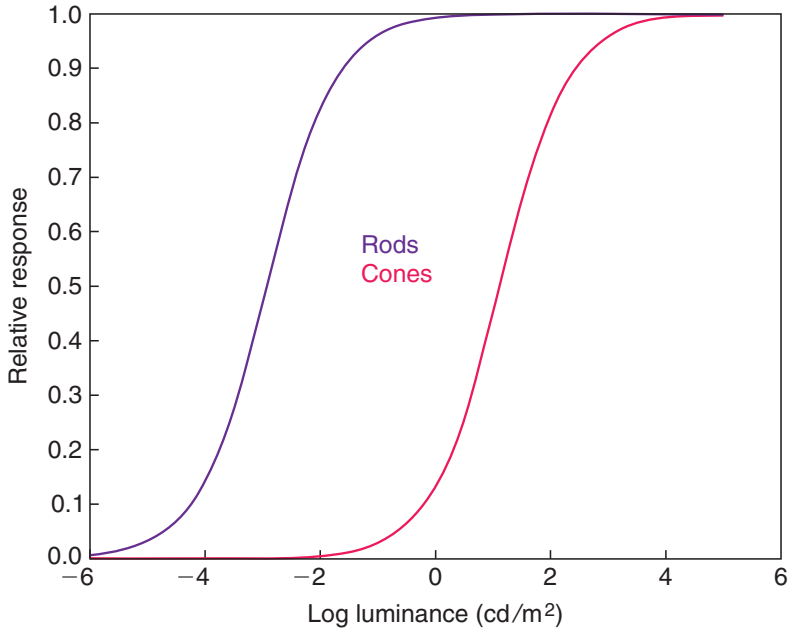


Figure 2.3: The response of dark-adapted rod and cone cells to various intensities [6].

to a new environment, the eye soon starts to adapt to the background luminnace. There is a strong relationship between the threshold adaptation and photoreceptor adaptation models. In Figure 2.4, the graph at the top illustrates photoreceptor response function at three different background luminance L_i, spaced approximately two *log* units apart. The ΔL_is are the luminance increments required to extract a fixed ΔR response increment. The bottom graph shows the ΔL_i values as a function of the background luminance L_i which is quite similar to the TVI curve in Figure 2.2.

Visual Adaptation Model for HDR Tone Mapping

Photoreceptor adaptation plays an important role in HVS adaptation. An appropriate mathematical model of this adaptation can be useful in developing tone-mapping operators. Schlick uses the following mapping function to determine display pixel values from pixel

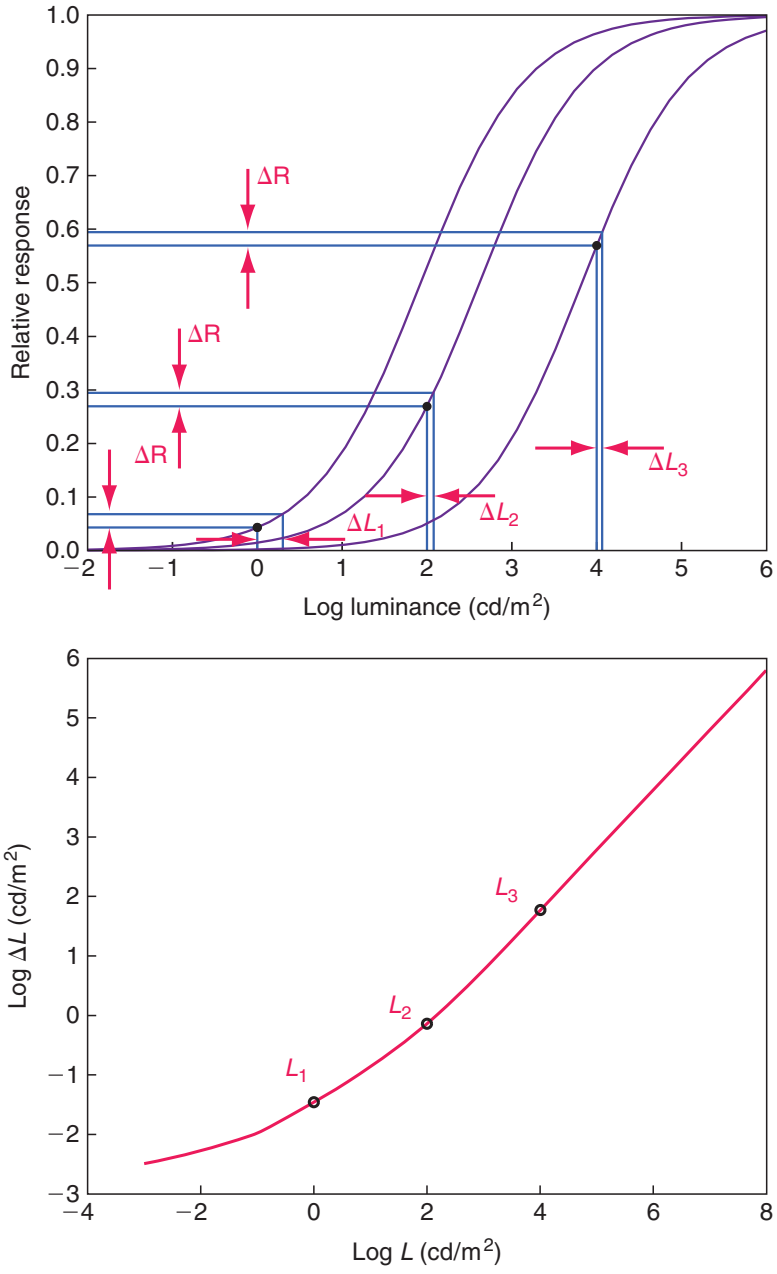


Figure 2.4: The response-threshold relation.

intensity, I [26].

$$F(I) = \frac{I}{I + \frac{I_{max} - I}{p}}\,, \tag{2.2}$$

where I_{max} is the maximum intensity and p takes a value in the range of 1 to ∞. We can easily relate this equation to (2.1) by setting the exponent n to 1, and substituting $\frac{I_{max}-I}{p}$ for σ. Tumblin et al. [27] proposed an S-shape curve (sigmoid) as their tone-mapping function:

$$F(I) = [\frac{I^n}{I^n + kI_b^n} + \frac{I_b^n}{k(I^n + kI_b^n)}] \cdot D\,, \tag{2.3}$$

which is inspired by Schlick's work. The first term is identical to (2.1) and the second term is to create an S-shaped function on a log-log plot. Pattanaik et al. [28,29], Reinhard and Devlin [30] explicitly make use of (2.1) to map high dynamic range image onto display dynamic range. Pattanaik et al. in [28] introduce two separate equations for rods and cones. The σ values for rods and cones are computed from the background intensity as

$$\sigma_{rod} = \frac{c_1 I_{rod}}{c_2 j^2 I_{rod} + c_3(1 - j^2)^4 I_{rod}^{\frac{1}{6}}}\,, \tag{2.4}$$

$$\sigma_{cone} = \frac{c_4 I_{cone}}{k^4 I_{cone} + c_5(1 - k^4)^2 I_{cone}^{\frac{1}{3}}}\,, \tag{2.5}$$

where

$$j = \frac{1}{c_6 I_{rod} + 1}\,, \tag{2.6}$$

$$k = \frac{1}{c_7 I_{cone} + 1}\,, \tag{2.7}$$

and I_{rod} and I_{cones} are the background intensities for the rods and cones, respectively. Reinhard and Devlin employed a simpler formula for defining σ at a given background intensity that is given by [30]

$$\sigma = (fI_b)^m \,, \tag{2.8}$$

where f and m are constants and are provided by users as desired parameters in their tone mapping algorithms. Reinhard then extended the function and provided a function which bears a strong resemblance to (2.1):

$$F(I) = \frac{I}{I + \frac{I_b}{a}} \tag{2.9}$$

where a is a scaling constant chosen according to the luminance range of image scene [13].

Threshold Versus Intensity (JND) Model For Tone-mapping

As mentioned before, the JND curve can be derived from photoreceptor responses. Therefore, the threshold versus intensity curve (TVI) can also be used in tone-mapping problems. Ward [31] utilized the TVI model for tone reproduction. From scene pixel luminance I_{scene} and the scene background luminance $I_{b,scene}$, the ratio

$$k = \frac{I_{scene} - I_{b,scene}}{\Delta I_{b,scene}} \tag{2.10}$$

is computed. This ratio calculates the number of JNDs which the pixel differs from the background. Therefore, substituting display background luminance $I_{b,display}$ and display adaptation threshold $\Delta I_{b,display}$, (2.10) can be rewritten to compute display pixel luminance:

$$I_{display} = k\Delta I_{b,display} + I_{b,display}, \tag{2.11}$$

Ferwerda et al. [32] later adapted this concept to compute JNDs specifically for rods and cones and used them for tone mapping images with a wide range of intensities.

Although HVS models have inspired numerous algorithms for tone reproduction, it is possible to devise algorithms without explicitly taking human vision into account. His-

togram adjustment [33], bilaterial filtering [12], multiscale optimization framework [11] and exposure fusion [34] belong to this class. Fattal et al. in [35] developed gradient domain dynamic range compression method based on compressive function to the gradient field. Fattal et al. observed that any drastic change in luminance in the HDR images generates large magnitude in luminance gradient, while fine details correspond to much smaller gradients. Hence, they propose to identify gradients across the HDR image first and then attenuate their magnitudes adaptively, i.e. larger gradients are attenuated more than smaller gradients.

2.3 Image Quality Assessment Methods

Image quality evaluation plays a critical role in many image processing problems, from image acquisition, synthesis and compression to restoration, enhancement and reproduction. The quality measurement methods are divided into two major categories, subjective and objective assessment. Although, human observers are very good at judging image quality, subjective assessment is expensive and time consuming. Objective quality metrics are appealing because they are able to incorporate the quality assessment method in various applications to measure and optimize the quality of images automatically. Moreover, they can be implemented by either software or hardware and computed very quickly.

Objective assessment methods are often classified into three major classes: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR) methods. Typically, FR quality metrics are used in image processing tasks where the reference image is available. RR algorithms attempt to measure the quality of a test image using certain features from the reference image, while in NR quality assessment, there is no information available from the reference image.

The usage of image and video quality measures in the design and optimization of image/video processing algorithms and systems is highly desirable. To incorporate in image/video processing algorithms, image/video quality assessment methods are desired to have the following properties,

- High correlation with subjective scores
- Low computational complexity
- Accurate local quality prediction that can help evaluate varying quality based on local content
- Good mathematical properties such as convexity that can reduce the difficulty in the design of optimization algorithms

To the best of our knowledge, there are only two image and video quality assessment methods that satisfy or nearly satisfy all of the above mentioned requirements: Mean Square Error (MSE) and the Structural SIMilarity (SSIM) index [20]. The remainder of this section introduces MSE and SSIM.

2.3.1 Mean Square Error (MSE)

The most widely used FR quality metrics are mean square error (MSE) and peak-signal to noise ratio (PSNR). MSE and PSNR are easy to apply, simple in calculation, and also have clear physical meanings. For a pair of images x and y, the MSE is computed by averaging the squared intensity differences of the test and reference image pixels as

$$\mathrm{MSE(x,y)} = \frac{1}{L_1 L_2} \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} [x(i,j) - y(i,j)]^2 \,, \tag{2.12}$$

where L_1 and L_2 are the length and width of the images, respectively. When MSE is computed based on the error signal, between the reference image x and its distorted version y, it can be employed as a measure for image quality. MSE is associated with attractive features such as simplicity, low computational cost, and memorylessness [7]. Moreover, MSE serves very well in solving optimization problems for the following reasons: it is a valid distance metric in $\mathcal{R}^N$; it preserves energy after any orthogonal transformation (Parseval's theorem); it is convex; and it is differentiable.

MSE is often converted to PSNR using the following expression

$$\text{PSNR} = 10\log_{10}\left(\frac{R^2}{\text{MSE}}\right) = 20\log_{10}\left(\frac{R}{\sqrt{\text{MSE}}}\right)[dB] \tag{2.13}$$

where R is the maximum pixel value, which for example takes 255 for an 8-bit pixel. The advantage of PSNR over MSE, as a perceptual quality measure, is its capability to handle images with different dynamic ranges.

It has been shown that MSE does not account for a number of important psychological features of the HVS [36]. Moreover, MSE is built based on several assumptions which may not be applicable in providing accurate perceived image/video quality prediction. For example, MSE is insensitive to the sign of the error signal. Furthermore, MSE treats all the pixels of an image equally, and thus the distortion in each pixel is calculated independently. It is claimed that these assumptions fail MSE as a perceptually meaningful quality assessment tool [20].

2.3.2 Structural SIMilarity index (SSIM)

The Structural SIMilarity index (SSIM) started a new paradigm in image quality assessment [7]. The basic assumption is that the HVS is highly adapted for extracting structural information, and thus the loss of structural information can cause perceptual distortions [20] [7]. The SSIM is a combination of three comparisons - luminance, contrast and structure (2.14):

$$\text{SSIM}(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y), \tag{2.14}$$

Let x and y be two images, the luminance component compares the luminance of the images using (2.15),

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \tag{2.15}$$

where μ_x and μ_y denote the mean intensities of image x and y, respectively, defined by (2.16).

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2.16}$$

where N is the number of pixels. The constant C_1 is included to avoid instability when the denominator is very close to zero, and the value is determined with $C_1 = (K_1 \cdot L)^2$, where L is the dynamic range of the pixel values and $K_1 \ll 1$ [20]. The luminance term is qualitatively consistent with Weber's law for its sensitivity to the relative luminance change. Contrast comparison is then performed which is specified as:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \tag{2.17}$$

where σ_x and σ_y represent the standard deviations of images x and y, respectively. For image x the standard deviation is defined by:

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2} \tag{2.18}$$

The constant C_2 has the same role as C_1 in luminance comparison and is computed with $C_2 = (K_2 \cdot L)^2$, where L is the dynamic range of pixel intensities and $K_2 \ll 1$ [20]. Finally, the structure-comparison function is given by:

$$s\,(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{2.19}$$

where σ_{xy} denotes the cross correlation between x and y defined by:

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y), \tag{2.20}$$

and C_3 is a constant to avoid division by zero. Since the correlation computation is

performed on normalized signals, the structure component is less sensitive to local image luminance and contrast. To simplify the expression C_3 is set to $C_2/2$, resulting in

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{2.21}$$

To obtain a local measure of structural similarity, the parameters are calculated within a local window that slides over the whole image, resulting in a quality map which indicates the quality of the distorted image across spatial locations.

The luminance term of the SSIM index is related to Weber's Law [20] [36], that indicates the perception of any stimulus change is proportional to the intensity of the stimulus. As explained in [20], Weber's law not only applies to the luminance but is also applicable to the image contrast i.e., the ratio of contrasts is constant for a constant SSIM value. The SSIM index gained significant attention in recent years and outperforms MSE and PSNR due to its good correlation with subjective quality assessment result [20] [7].

Figure 2.5 demonstrates a comparison between the performance of MSE and SSIM using an illustrative example. Figure 2.5(a) is the reference image and the rest of the images are contaminated with different types of artifacts. It can be observed that although the perceptual quality of the distorted images differs quite significantly, MSE gives the same score to all distorted images and predict a similar quality. By contrast, the SSIM values are better correlated to human perception. Several important mathematical properties of the SSIM index have been investigated in [37]. It has been proved that like the MSE, the SSIM index is preserved under orthogonal or unitary transformations. In addition, it is shown that the SSIM index can be partitioned into two components, each of which may be transformed into a valid distance metric. Convexity, quasi-convexity, and generalized convexity have also been shown to hold locally for the metrics derived from SSIM [37]. In brief, SSIM achieves the best compromise between accurate prediction of image quality and good mathematical properties, and thus is preferred in numerical optimization frameworks.

The SSIM index does not take into account the viewing distance of the human subject, and thus the scale of the images have impact on SSIM performance. To consider this

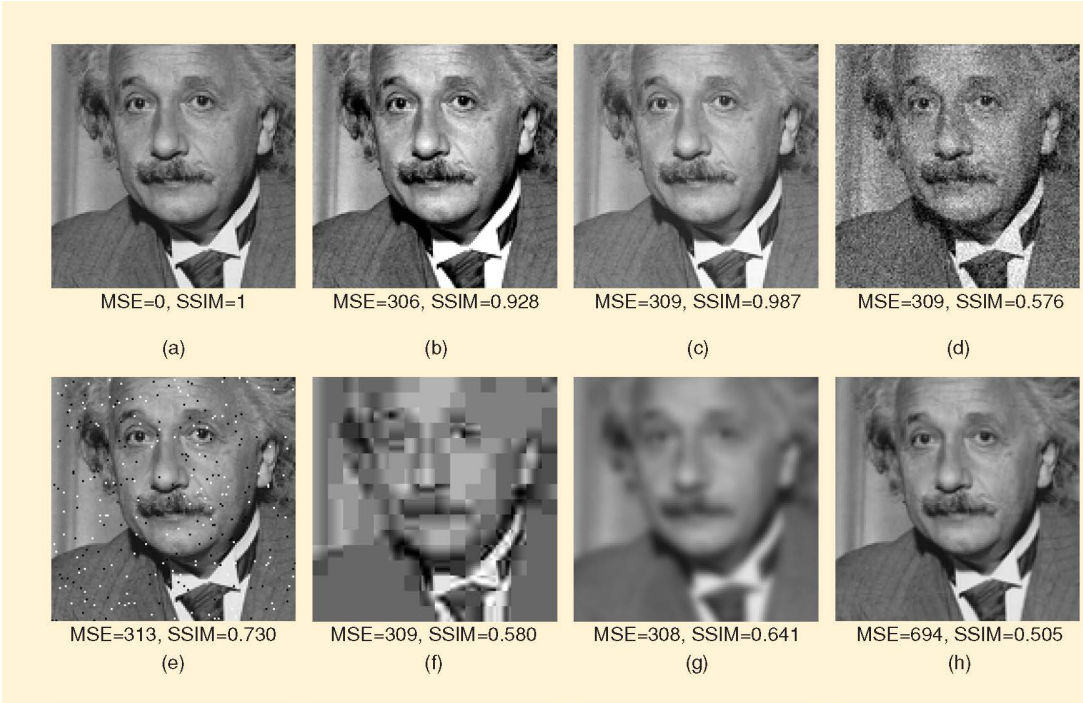


Figure 2.5: Comparison of SSIM and MSE performances for "Einstein image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise. (e) Impulsive noise. (f) JPEG compression. (g) Blurring. (h) Spatial scaling. (Image courtesy of Wang and Bovik's paper [7])

issue, Multi-Scale Structural SIMilarity index (MS-SSIM) has been proposed in [21]. In the general form, the MS-SSIM is given by

$$\mathrm{MS-SSIM(x,y)} = \prod_{r=1}^{R} [l(x_r, y_r)]^{\alpha_r} [c(x_r, y_r)]^{\beta_r} [s(x_r, y_r)]^{\gamma_r} \,, \tag{2.22}$$

where x_r and y_r are the image x and y, respectively, at resolution r. The α, β and γ are the relative importance/weight of each scale that were decided based on psychovisual experiments. It is worth noting that the weights were determined based on the subjective experiments and were found to be consistent with the general shape of Contrast Sensitivity Function (CSF) [21].

Typically, a simple averaging over the local SSIM scores is employed for spatial pooling [20]. A new pooling approach based on information content is proposed in [38]. It was shown that Information Weighted SSIM (IW-SSIM) outperforms the original SSIM index [38]. Information content based weighting can yield more accurate quality prediction as compared to minkowski, local quality/distortion-based, saliency-based, and object-based pooling.

Popular quality metrics such as MSE, PSNR and SSIM assume the dynamic range of the distorted image is the same as the dynamic range of the reference image, and thus cannot be used in quality comparison of two images with different luminance dynamic ranges. More specifically, in the case of SSIM, the luminance and contrast terms cannot be used since there is a large difference in the luminance and contrast ranges of tone-mapped images and their corresponding HDR references. Moreover, HDR-LDR quality assessment does not directly fall into FR, RR and NR categories. The human eyes are unable to see the actual HDR image on regular displays, and thus the quality of tone-mapped images are judged without a reference with perfect visual quality. On the other hand, we cannot simply classify a HDR-LDR quality measure as an NR quality metric since the pixel values of the reference HDR image are fully provided and are available to check the signal fidelity of the LDR image. Therefore, quality assessment of tone-mapped images introduces a new challenge to the field of image quality assessment.

2.3.3 Quality Assessment of Tone-Mapping Operators

Because of the reduction in dynamic range, tone mapping procedures inevitably cause information loss. With multiple TMOs available, one would ask which TMO faithfully preserves the structural information in the original HDR images, and which TMO produces natural-looking realistic LDR images.

TMO assessment in the past mostly relied on human subjective evaluations. In [39], perceptual evaluations of six TMOs were conducted with regard to similarity and preferences. An overview and a subjective comparison of eight TMOs were reported in [40]. HDR capable monitor was employed in [41] to compare six TMOs in a subjective experiment using a paired comparison method. In [42], fourteen subjects were asked to rate two architectural interior scenes produced by seven TMOs based on basic image attributes as well as the naturalness of the LDR images. A more comprehensive subjective evaluation was carried out in [2], where tone mapped images generated by fourteen TMOs were shown to two groups of ten human observers to rate LDR images, concerning overall quality, brightness, contrast, detail reproduction and color. In [43], subjects were asked to choose the best LDRs derived from two TMOs with different parameter settings to optimally tune the algorithms. The value of subjective testing cannot be overestimated. However, they have fundamental limitations. First, subjective measurement is expensive and time consuming. Second, it is difficult to be incorporated into an optimization framework to automatically improve TMOs and adjust their parameter settings. Furthermore, important image structures contained in HDR images may be missing in tone mapped images, but human observers may not be aware of the existence of such missing information. In this sense, subjective evaluation should not be regarded as a golden standard for the quality of tone mapped images.

Typical objective image quality assessment (IQA) approaches assume the reference and test images to have the same dynamic range [36], and thus cannot be directly applied to evaluate tone mapped images. Only a few objective assessment methods have been proposed for HDR images. The HDR visible difference predictor (HDR-VDP) [44] is a HVS based fidelity metric aiming to distinguish between visible (or suprathreshold) and invisible

(or subthreshold) distortions. The HDR VDP takes two HDR images as the reference and test images and generates a map of probability as the output. Figure 2.6 depicts the block diagram of the visible difference predictor. First, a sophisticated processing procedure is made on both the reference and test images to discriminate visual masking effect by deriving threshold contrast in different spatial and orientation channels. Second, to consider the relative insensitivity of the HVS to the small shift of the signal phase, a low-pass filter with a small kernel is applied. The filtered masking maps are then normalized to measure the distortion between the reference and distorted images for every pixel and for every channel. Subsequently, the normalized distortion measure is subjected to the psychometric function[1] that estimates the probability of detecting the difference between each channel [45]. For each pixel, the estimated probability values are summed across all channels. Finally, the probability values are used to predict visible differences between the reference and the distorted images. It is assumed that the difference can be perceived for a given pixel when the probability value is greater than 0.75. The interpretation of the distortion maps is important, as the HDR VDP is sometimes misused to measure the magnitude of distortions, which is not its intended application. The HDR VDP is a threshold fidelity measure whose task is to distinguish between visible (or suprathreshold) and invisible (or subthreshold) degradations. The metric is complicated in implementation and should be regarded as an indicator of perceptual distortions in terms of probability of detection only.

2.3.4 Dynamic Range-Independent Image Quality Assessment

The HDR VDP assumes the dynamic ranges of the images being compared are similar. A dynamic range independent approach was proposed in [8], which improves upon HDR-VDP. The DRI measure follows the HVS model in HDR VDP, which provides detection of visible contrast changes. However, in the last block, the visibility information is employed to analyze only visible structural changes. Aydin et al. consider three classes of structural changes [8]:

[1]We will explain the psychometric function in detail in the next chapter.

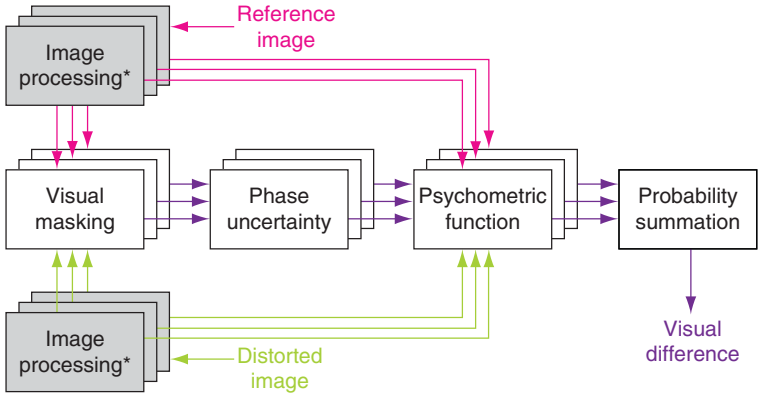


Figure 2.6: Framework of HDR-VDP approach (Image courtesy of [5]).

- Loss of visible contrast: Image details which can be seen in the HDR reference image disappear in the LDR image. This is a common problem in tone mapping processes.
- Amplification of invisible contrast: Image details which are not visible in the HDR image, but tone mapping operators make them visible in the LDR image.
- Reversal of visible contrast: Image details which can be seen in both HDR and tone mapped image but with different polarity.

Figure 2.7 illustrates each of the discussed distortion types for a simple signal. The three structural distortion types are computed using their probabilities, and each type produces a distortion map. Eventually, the output of this metric is three visibility maps each corresponds to one distortion type. These quality maps show good correlations with subjective classifications of image degradation types including blur, sharpening, contrast reversal, and no distortion. However, this approach suffers from its high implementation complexity. Moreover, it does not provide a single quality score for an entire image, making it impossible to be validated with subjective evaluations of the overall image quality.

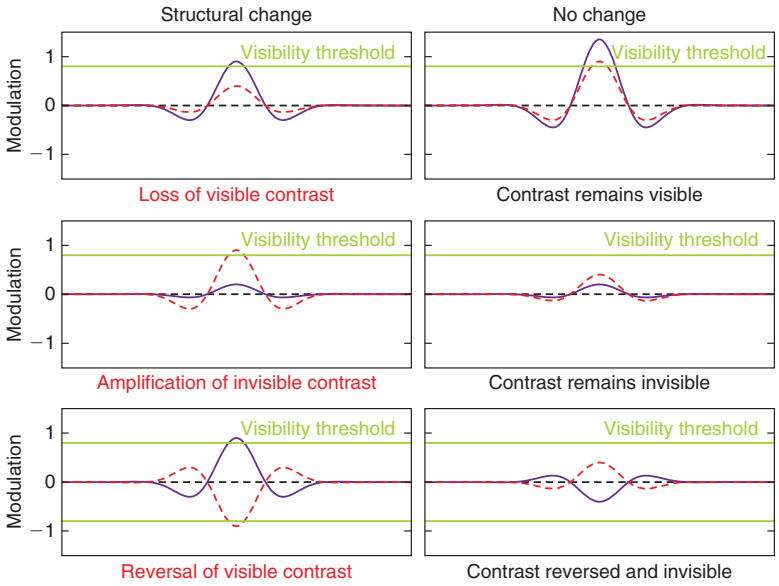


Figure 2.7: Specific cases of contrast modification that the DRI measure classifies as a structural change (left) or a lack of structural change (right). The solid and dashed lines depict the reference and test signals, respectively, while the horizontal lines denote the visibility threshold level [8] (Image courtesy of [5]).

2.4 Image Interpolation

Image interpolation techniques that can improve the spatial resolution of given low-resolution (LR) images are used in many real-world devices and systems such as web browsers, media players, photo editors, and high-definition television (HDTV) [46]. Over the past decades, an increasing number of interpolation algorithms have been proposed. They can be classified into two major categories- spatially invariant and spatially adaptive methods [47].

Spatially invariant techniques enlarge LR images by directly computing new pixel values without differentiating local image features. The typical method is to model local image signal by a low-order polynomial function. They are applied uniformly to all pixels and are computationally inexpensive. Classical interpolation algorithms such as nearest neighbour, bilinear and bicubic methods belong to this category. This type of interpolation methods

often produce artifacts such as blur, blockiness and halo specifically around the edges. To resolve these problems, several spatially adaptive interpolation methods have been proposed recently, which often adapt towards image edges or locally oriented structures [9, 14–17].

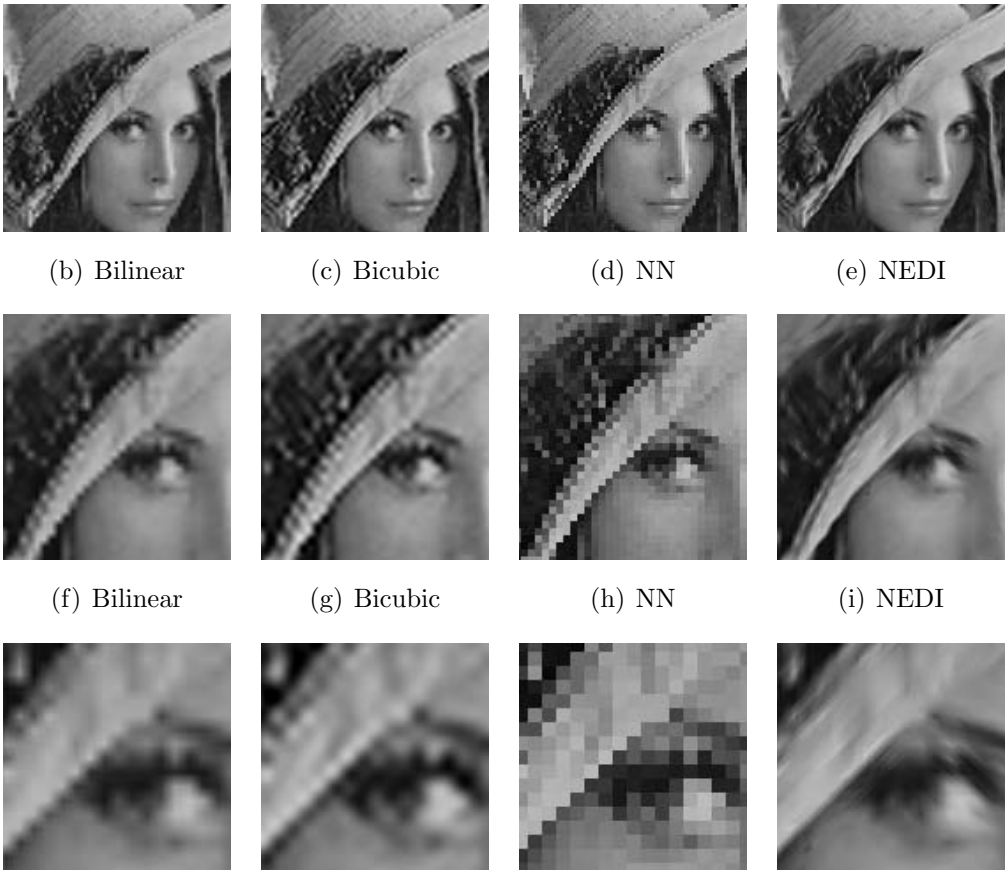
Edges are visually attractive to the human perceptual system, and thus spatially adaptive interpolation algorithms often attempt to address edge reconstructions. As a result, most spatially adaptive interpolation algorithms can be regarded as edge-directed methods. The key idea is to preserve the edge sharpness during the interpolation process. In particular, most of these methods explicitly estimate edge orientation and then interpolate along the edge orientation. To reduce the computational complexity, some methods further quantize the edge orientation [48] [49]. The performance of interpolated images using edge-directed methods is often determined by the estimation accuracy of the edge orientation. It has been shown that weighting the edge orientations, can improve the perceptual quality of interpolated images [47]. A comprehensive review on interpolation techniques can be found in [47] and [50].

2.5 Cross Spatial Resolution Image Quality Assessment

Figure 2.8 depicts examples of reconstructed high resolution (HR) images created from the LR "Lena" image for scaling factors of 2, 4 and 8, by means of bilinear, bicubic, nearest neighbor (NN), and new edge-directed interpolation (NEDI) [9] methods, respectively. It can be observed that as the scaling factor increases, the perceptual differences between different interpolation methods become more pronounced. A natural question arises here is: with a variety of interpolation methods available, which of them produces more natural-looking realistic HR images? To answer this question, IQA methods are highly desirable, without which, different interpolation methods cannot be compared and future improvement is pointless.

Subjective evaluation provides a direct and reliable method in assessing the quality of

(a) LR image



(b) Bilinear (c) Bicubic (d) NN (e) NEDI

(f) Bilinear (g) Bicubic (h) NN (i) NEDI

(j) Bilinear (k) Bicubic (l) NN (m) NEDI

Figure 2.8: (a): low-resolution (LR) image; (b-e): interpolated images by a scaling factor of 2; (f-i): interpolated images by a scaling factor of 4; (j-m): interpolated images by a scaling factor of 8. Column 1: bilinear interpolation; Column 2: bicubic interpolation; Column 3: nearest neighbor (NN) interpolation; Column 4: new edge-directed interpolation (NEDI) [9]. All interpolated images are cropped for better visualization.

interpolated images. In subjective IQA experiments, human subjects may be instructed to assign a "quality score" based on a linear scale or quality category for each given interpolated image, and then the mean opinion scores (MOSs) as well as the variations between multiple subjects can be calculated [51, 52]. The subjects may also be asked to compare pairs of images and pick the one with higher quality [53, 54]. Such a two-alternative-forced-choice (TAFC) approach has been shown to provide consistent test results in the literature of visual psychophysics [55], though with low efficiency.

Subjective studies of the trade-off among spatial resolution, temporal resolution, and quantization step size are carried out in [56–59]. Bae et al. in [56] conducted a subjective test to investigate the preferred spatial resolution for a given quantization error. They demonstrate that people prefer to observe larger images with less quantization error rather than low resolution images with no visible quantization error. Moreover, Bae et al. conclude that beyond certain quantization error, subjects accept more distortion as the spatial resolution decreases. Wang et al. [60] conducted a subjective study to examine the impact of jointly adjusting spatial resolution, temporal resolution, and quantization step-size. They demonstrate that people prefer a smaller image with smaller quantization errors compared to a larger image with larger quantization errors, for the same bit-rate. The effect of different spatial resolution, temporal resolution and quantization parameters on subjective quality on mobile platforms are studied in [57] and [58], where the experiments indicate that video content has impact on perceptual quality of subjects. Cermak et al. in [59] validated the test results of two VQEG projects. They used the Mean Opinion Score (MOS) obtained from human subjects for QCIF(176×144), CIF(352×288), VGA(640×480), and HD(1920×1200) resolution at several bit-rates, and determined the required bit-rate to achieve a given level of video quality for a given screen resolution. All these subjective evaluation methods are useful in comparing the performance of interpolation algorithms. However, they are often time-consuming and expensive, which largely constrains their applications when the volume of images becomes large or when one aims to incorporate them into the optimal design and parameter tuning of interpolation algorithms.

Very limited progress has been made in automatic or objective quality assessment of interpolated images. To employ existing IQAs to estimate the quality of two images with

different spatial resolution, researchers follow two straightforward methods. They either compare the low resolution image with the decimated high resolution reference image or compare the interpolated low resolution image with the high resolution reference image. Both approaches are inaccurate in predicting perceptual quality [61]. The difficulty lies in the fact that a perfect-quality HR image is unavailable to compare with. As a result, typical FR objective IQA approaches such as PSNR and SSIM are not directly applicable. It is worth noting that the pixels in the LR image constitute a subset of the HR image pixels and are available to the IQA system. This well fits into the scenario of RR IQA, where only partial information about the perfect-quality original image is accessible [36].

Few studies have been performed to adopt existing objective quality measures to understand the effect of resolution on the quality. These studies focus on providing the best quality video transmission under the constraints of available bandwidth, and the resolution of the viewer's display. Reed and Lim proposed an algorithm in [62] to explore the best trade-off between spatial resolution, temporal resolution, and encoding quantization parameters by using Sum of Absolute Error (SAE) as the objective measure. Akyol et al. presented a framework in [63] to choose the best settings for a scalable encoder using an NR objective measure that quantifies different image artifacts such as blockiness and blurriness.

In [61], Demirtas et al. proposed a FR objective quality assessment algorithms to estimate the quality of a distorted image with a lower resolution using high resolution reference image. They assume that the LR image is generated from the HR reference image by performing Low Pass Filtering (LPF) followed by downsampling. Their objective measure is based on wavelet representation and measuring the mutual information motivated by Visual Information Fidelity (VIF) measure [19]. More specifically, they decompose both HR and LR images using bi-orthogonal wavelets, and then compute the mutual information between the corresponding subbands using similar model described in the mutual information calculation of VIF. Moreover, they compare HR reference image an LR image as if they are being viewed at identical visual angle. Using Equation 2.23, and utilizing CSF

function in [64], they model the effect of resolution change to the perceptual quality by

$$f(l) = \frac{\pi \times d \times n}{180 \times h \times 2 \times 2^l}\,, \tag{2.23}$$

where d, h and n represent viewer distance, height of the screen, and the number of pixels in the vertical direction, respectively. Moreover, l indicates the level of subband decomposition. Finally, they conducted a subjective test and demonstrated that their method well correlates with subjective data. Although the proposed method is technically sound, it is limited by the use of dyadic wavelet transforms, implying that the ratio between the size of the input image and that of the reference image can only be a power of 2. In addition, their method is computationally expensive.

Chapter 3

Objective Quality Assessment of Tone-Mapped Images

Tone mapping operators (TMOs) that convert HDR to LDR images provide practically useful tools for the visualization of HDR images on standard LDR displays. Different TMOs create different tone-mapped images, and a natural question is which image has the best quality. Without an appropriate quality measure, different TMOs cannot be compared and further improvement is directionless. Subjective rating may be a reliable evaluation method, but is expensive and time-consuming, and more importantly, is difficult to embed into optimization frameworks.

This chapter develops an objective IQA model for tone mapped LDR images using their corresponding HDR images as references. The work is inspired by the success of two design principles in the IQA literature. The first is the Structural SIMilarity (SSIM) approach [20] and its multi-scale derivations [21], [38], which asserts that the main purpose of vision is to extract structural information from the visual scene and as a consequence structural fidelity is a good predictor of perceptual quality. The second is the natural scene statistics (NSS) approach, which maintains that the visual system is highly adapted to the natural visual environment and uses the departure from natural image statistics as a measure of perceptual quality [65]. Here we propose a method that combines a multi-scale structural

fidelity measure and a statistical naturalness measure, leading to the Tone-Mapped image Quality Index (TMQI). Moreover, we demonstrate that TMQI can be employed to optimize parameters in TMOs and to adaptively fuse multiple tone-mapped images.

3.1 Tone-Mapped Image Quality Index

Due to the reduction in dynamic range, TMOs cannot preserve all information in HDR images, and human observers of the LDR versions of these images may not be aware of the loss. Therefore, structural fidelity plays an important role in assessing the quality of tone-mapped images. On the other hand, structural fidelity alone does not suffice to provide an overall quality evaluation. A good-quality tone- mapped image should achieve a good compromise between structural fidelity preservation and statistical naturalness, which are sometimes competing factors.

3.1.1 Structural Fidelity

The SSIM approach provides a useful design philosophy as well as a practical method for measuring structural fidelities between images [7]. The original SSIM algorithm is applied locally and contains three comparison components − luminance, contrast and structure. Since TMOs are meant to change local intensity and contrast, direct comparisons of local contrast are inappropriate. Let x and y be two local image patches extracted from the HDR and the tone-mapped LDR images, respectively. The local structural fidelity measure is defined as

$$S_{\text{local}}(x, y) = \frac{2\sigma'_x\sigma'_y + C_1}{{\sigma'_x}^2 + {\sigma'_y}^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2}\,, \tag{3.1}$$

where σ_x, σ_y and σ_{xy} are the local standard deviations and cross correlation between the two corresponding patches in HDR and LDR images, respectively, and C_1 and C_2 are positive stabilizing constants. Parameters σ' is to distinguish significant and insignificant local contrast, and is explained in details below. Compared with the SSIM definition [20], the luminance comparison component is missing, and the structure comparison component (the

second term in (3.1)) is exactly the same. The first term in (3.1) compares signal strength and is modified from that of the SSIM definition based on two intuitive considerations. First, the difference of signal strength between HDR and LDR image patches should not be penalized when their signal strengths are both significant (above the visibility threshold) or both insignificant (below the visibility threshold). Second, the algorithm should penalize the cases in which the signal strength is significant in one of the image patches but insignificant in the other. This differs from the corresponding term in the original SSIM definition where any change in signal strength is penalized.

To distinguish between significant and insignificant signal strength, the local standard deviation σ is passed through a nonlinear mapping, which results in the σ' value employed in (3.1). The nonlinear mapping should be designed so that significant signal strength is mapped to 1 and insignificant signal strength to 0, with a smooth transition in between. Therefore, the nonlinear mapping is related to the visual sensitivity of contrast, which has been extensively studied in the literature of visual psychophysics [45]. Practically, the HVS does not have a fixed threshold of contrast detection, but typically follows a gradual increasing probability in observing contrast variations. Psychometric functions describing the detection probability of signal strength have been employed to model the data taken from psychophysical experiments. Generally, the psychometric function resembles a sigmoid shape [66, 67], and the sensory threshold is usually defined at the level of 50% of detection probability. A commonly adopted psychometric function is known as Galton's ogive [45], which takes the form of a cumulative normal distribution function given by

$$p(s) = \frac{1}{\sqrt{2\pi}\theta_s} \int_{-\infty}^{s} \exp\left[-\frac{(x-\tau_s)^2}{2\theta_s^2}\right] dx\,, \tag{3.2}$$

where p is the detection probability density, s is the amplitude of the sinusoidal stimulus, τ_s is the modulation threshold, and θ_s is the standard deviation of the normal distribution that controls the slope of detection probability variation. It was found that the ratio

$$k = \frac{\tau_s}{\theta_s}\,, \tag{3.3}$$

is roughly a constant, known as Crozier's law [45, 68]. Typical values of k range between 2.3 and 4, and $k = 3$ makes the probability of false alarm relatively small [45].

The reciprocal of the modulation threshold τ_s is often used to quantify visual contrast sensitivity, namely the contrast sensitivity function (CSF), which is a function of spatial frequency [45]. A CSF formula that fits well with data collected in various psychological experiments is given by [10]

$$A(f) \approx 2.6[0.0192 + 0.114f] \exp[-(0.114f)^{1.1}], \tag{3.4}$$

where f denotes spatial frequency. This function is normalized to have peak value 1 as shown in Figure 3.1, and thus provides only relative sensitivity across the frequency spectrum.

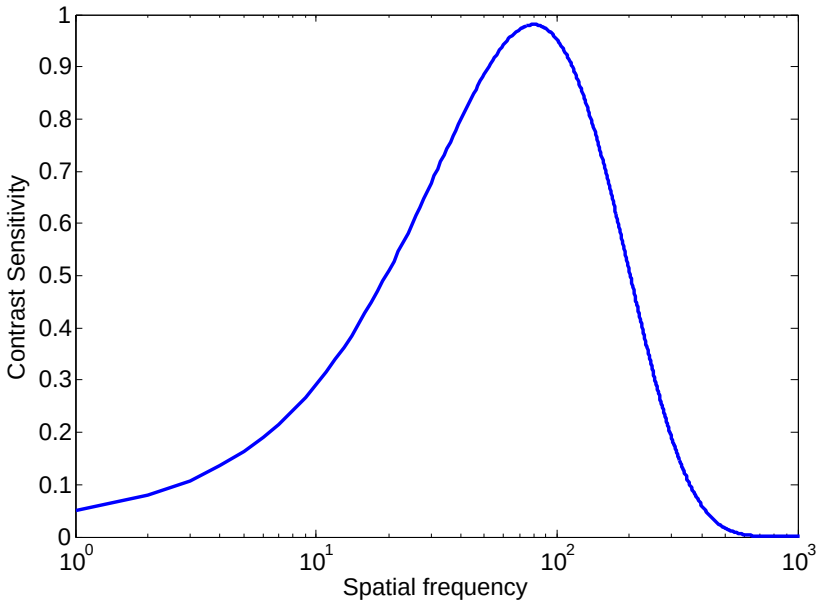


Figure 3.1: Plot of Contrast Sensitivity Function given in [10]

In practice, it needs to be scaled by a constant λ to fit psychological data. The implementation proposed in this thesis follows Kelly's CSF measurement [69]. Combining this

with (3.4), we obtain

$$\tau_s(f) = \frac{1}{\lambda\, A(f)}\,. \tag{3.5}$$

where the subscript s denotes the threshold defined using signal strength. This threshold value is calculated based on contrast sensitivity measurement that assumes pure sinusoidal stimulus. To convert it to a signal strength threshold measured using the standard deviation of the signal, we need to take into account that signal amplitude scales with both contrast and mean signal intensity, and there is a $\sqrt{2}$ factor between the amplitude and standard deviation of a sinusoidal signal. As a result, a threshold value defined on signal standard deviation, σ, can be computed as

$$\tau_\sigma(f) = \frac{\overline{\mu}}{\sqrt{2}\,\lambda\, A(f)}\,, \tag{3.6}$$

where $\overline{\mu}$ is the mean intensity value and the subscript σ indicates that the modulation threshold is defined using the standard deviation of signals. Based on Crozier's law [45,68], we have

$$\theta_\sigma(f) = \frac{\tau_\sigma(f)}{k}\,. \tag{3.7}$$

We can then define the mapping between σ and σ' as

$$\sigma' = \frac{1}{\sqrt{2\pi}\theta_\sigma}\int_{-\infty}^{\sigma} \exp\left[-\frac{(x-\tau_\sigma)^2}{2\theta_\sigma^2}\right] dx\,, \tag{3.8}$$

In (3.1), σ'_x and σ'_y are the mapped versions of σ_x and σ_y, respectively. They are bounded between 0 and 1, where 0 and 1 represent completely insignificant and completely significant signal strengths, respectively.

The local structural fidelity measure S_{local} is applied to an image using a sliding window that runs across the image space. This results in a map that reflects the variation of structural fidelity across space. The visibility of image details depends on the sampling density of the image, the distance between the image and the observer, the resolution of the display, and the perceptual capability of the observer's visual system. A single-scale method cannot capture such variations. Following the idea used in multi-scale [21] and

information-weighted SSIM [38], we adopt a multi-scale approach, where the images are iteratively low-pass filtered and downsampled to create an image pyramid structure [70], as illustrated in Figure 3.2. The local structural fidelity map is generated at each scale. Figure 3.3 shows two examples of such maps computed at multiple scales for the LDR images created from two different TMOs. It is interesting to observe these fidelity maps and examine how they correlate with perceived image fidelity. For example, the structural details of the brightest window regions are missing in Image (b), but are more visible in Image (a). For another example, there are detailed structures in the top-right dark regions that are not easily discerned in Image (a), but are better visualized in Image (b). All of these observations are clearly reflected in the structural fidelity maps.

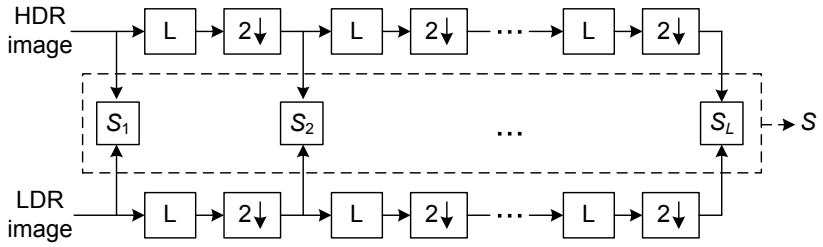
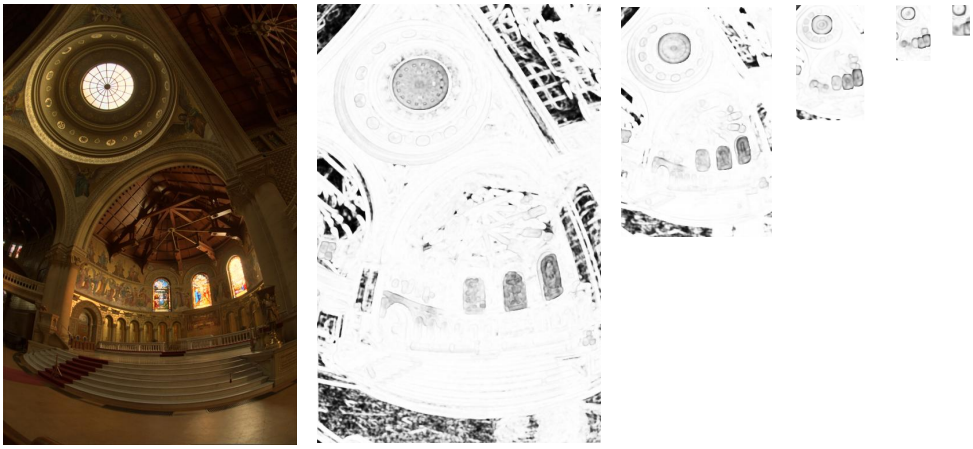


Figure 3.2: Framework of multi-scale structural fidelity assessment

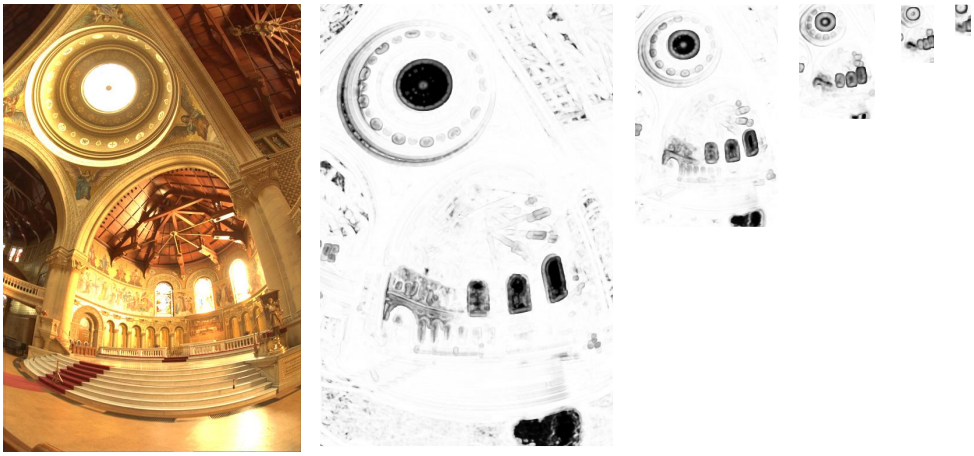
At each scale, the map is pooled by averaging to provide a single score:

$$S_l = \frac{1}{N_l} \sum_{i=1}^{N_l} S_{\text{local}}(x_i, y_i)\,, \tag{3.9}$$

where x_i and y_i are the i-th patches in the HDR and LDR images being compared, respectively, and N_l is the number of patches in the l-th scale. In the literature, advanced pooling strategies such as information content based pooling [38] have been shown to improve the performance of IQA algorithms. However, in our current experiment, these advanced pooling methods did not result in notable performance gain in the proposed structural fidelity measure. The overall structural fidelity is calculated by combining scale level structural



(a) $S = 0.9152$ ($S_1 = 0.8940$; $S_2 = 0.9341$; $S_3 = 0.9428$; $S_4 = 0.9143$; $S_5 = 0.8277$)



(b) $S = 0.8614$ ($S_1 = 0.9161$; $S_2 = 0.9181$; $S_3 = 0.8958$; $S_4 = 0.8405$; $S_5 = 0.7041$)

Figure 3.3: Tone mapped LDR images and their structural fidelity maps in five scales. The images were created using Adobe Photoshop "Highlight compression" and "Exposure and Gamma" methods (not optimized for quality), respectively.

fidelity scores using the method in [21]

$$S = \prod_{l=1}^{L} S_l^{\beta_l}\,, \tag{3.10}$$

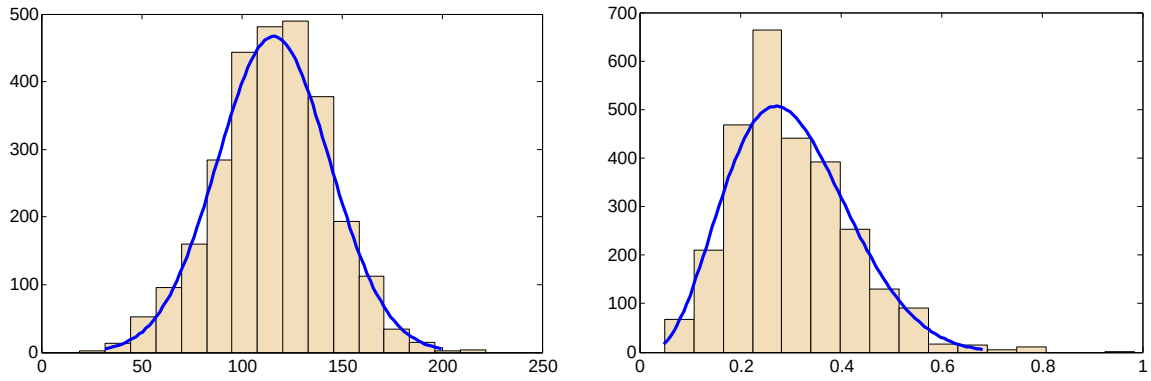
where L is the total number of scales and β_l is the weight assigned to the l-th scale.

There are several parameters in the implementation of our structural fidelity model. First, when computing S_{local}, we set $C_1 = 0.01$ and $C_2 = 10$, and find that the overall performance of the structural fidelity model is insensitive to these parameters within an order of magnitude. Second, to create the fidelity map at each scale, we adopt the same setting as in the SSIM algorithm [20] by employing a Gaussian sliding window of size 11×11 with standard deviation 1.5. Third, as in [21], we assume a viewing distance of 32 cycles/degree, which can represent signals up to 16 cycles/degree of resolution without aliasing, and thus we use 16 cycles/degree as the spatial frequency parameter when applying the CSF in (3.4) to the finest scale measurement. The spatial frequency parameters applied to the subsequent finer scales are then 8, 4, 2, 1 cycles/degree, respectively. Fourth, the mean intensity value in (3.6) is set to be the mean of the dynamic range of LDR images, i.e., $\overline{\mu} = 128$. Fifth, when combining the measures across scales, we set $L = 5$ and $\{\beta_l\}$ = {0.0448, 0.2856, 0.3001, 0.2363, 0.1333}, which follows the psychophysical experiment results reported in [21]. Finally, in order to assess the quality of color images, we first convert them from RGB color space to Yxy space and then apply the proposed structural fidelity measure on the Y component only.

3.1.2 Statistical Naturalness

A high quality tone-mapped LDR image should not only faithfully preserve the structural fidelity of the HDR image, but also look natural. Naturalness, however, is a subjective quantity that is difficult to define quantitatively. A large body of literature has been dedicated to the statistics of natural images. These statistics have important significance to both image processing applications and the understanding of biological vision [71]. An interesting study of naturalness in the context of subjective evaluation of tone-mapped images was carried out in [72], and provided useful information regarding the correlations between image naturalness and different image attributes such as brightness, contrast, color reproduction, visibility and reproduction of details. The results showed that among all attributes being tested, brightness and contrast have more correlation with perceived

naturalness. This finding motivates us to build our statistical naturalness model based on these two attributes. This choice may lead to oversimplification in defining the general concept of statistical image naturalness (and may not generalize to other image processing applications that use the concept of naturalness). It does however provide an ideal compromise between the simplicity of our model and the capability of capturing the most important ingredients of naturalness that are related to the tone- mapping evaluation problem we are trying to solve, where brightness mapping is an inevitable issue in all tone-mapping operations. The choice also best complements the structural fidelity measure described in Section 3.1.1, where brightness modeling and evaluation are missing.



(a) (b)

Figure 3.4: Histograms of (a) means (fitted by Gaussian PDF) and (b) standard deviations (fitted by Beta PDF) of natural images.

Our statistical naturalness model is built upon statistics conducted on about 3,000 8bit/pixel gray-scale images obtained from [73, 74]. These images represent many different types of natural scenes. Figure 3.4 shows the histograms of the means and standard deviations of these images, which are useful measures that reflect the global intensity and contrast of images. We found that these histograms can be well fitted using a Gaussian and a Beta probability density functions, respectively, which are given by

$$P_m(m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{(m-\mu_m)^2}{2\sigma_m^2}\right] \tag{3.11}$$

and

$$P_d(d) = \frac{(1-d)^{\beta_d - 1} d^{\alpha_d - 1}}{\mathrm{B}(\alpha_d, \beta_d)}\,, \tag{3.12}$$

where $\mathrm{B}(\cdot,\cdot)$ is the Beta function. The fitting curves are shown in Figure 3.4, where the model parameters are estimated by least-square regression. The best values we found are $\mu_m = 115.94$ and $\sigma_m = 27.99$ in (3.11), and $\alpha_d = 4.4$ and $\beta_d = 10.1$ in (3.12), respectively.

Recent studies suggest that brightness and contrast are largely independent quantities in terms of both natural image statistics and biological computation [75]. As a result, their joint probability density function can be well approximated by the product of the two. Therefore, we define our statistical naturalness measure as

$$N = \frac{1}{K} P_m\, P_d\,, \tag{3.13}$$

where K is a normalization factor given by $K = \max\{P_m\, P_d\}$. This normalization constrains the statistical naturalness measure to be bounded between 0 and 1.

3.1.3 Quality Assessment Model

The structural fidelity measure S introduced in Section 3.1.1 and the statistical naturalness measure N described in Section 3.1.2 characterize different aspects of the quality of tone mapped images. They may be used individually or jointly as a vector valued measure. In many practical applications, however, users prefer a single score that indicates the overall quality of the image. Therefore, these parameters should be combined in some manner. In the literature of IQA, earlier work combined image statistics and measures of structure and contrast [76], although in a different context to ours. Here we define a three-parameter function to scalarize the joint measure, resulting in a Tone-Mapped image Quality Index (TMQI)

$$Q = aS^{\alpha} + (1-a)N^{\beta}\,, \tag{3.14}$$

where $0 \le a \le 1$ adjusts the relative importance of the two components, and α and β determine their sensitivities, respectively. Since both S and N are upper-bounded by 1,

the overall quality measure is also upper-bounded by 1.

The parameters in (3.14) are left to be determined. In our implementation, they are tuned to best fit the subjective evaluation data provided by [1], where the subjects were instructed to look simultaneously at two LDR images created by two different TMOs applied upon the same HDR image. Then they had to pick the one with better overall quality. Two studies have been done, involving two groups of subjects. The first study was carried out at Zheijang University, where 59 naive volunteers were invited to do the pair-wise comparison task and fill in the preference matrix. The second study was conducted using Amazon Mechanical Turk, an online service of subjective evaluation. Each paired comparison was assigned to 150 anonymous subjects. The database includes six data sets, each of which contains images generated by five well-known TMOs, introduced by Drago *et al.* [4], Durand & Dorsey [12], Fattal *et al.* [35], Reinhard *et al.* [13] and Mertens*et al.* [34]. The subjective ranking scores in each folder were then computed using the preference matrix.

Finding the best parameters in (3.14) using subjective data is essentially a regression problem. The major difference from traditional regression problems is that here we are provided with relative ranking data between images only, but not quality scores associated with individual images. We have developed a learning method whereby the parameters are learnt from an iterative method. At each iteration, one pair of images is randomly selected from one randomly selected data set. If the model generates objective scores that place the pair the same order as in the subjective rank order, then there is no change to the model parameters; otherwise, each parameter is updated in the direction of correcting the model error by a small step. In other words, if both the structural fidelity and the statistical naturalness measures produce correct ranking scores, no parameter update is required. Otherwise we give more weight to the component that results in correct rank order and reduces the contribution of the other. The iteration continues until convergence. In our experiment, we observe a good convergence property in this iterative learning process. To ensure the robustness of our approach, we conducted a leave-one-out cross validation procedure, whereby the database (of six data sets) was divided into five training sets and one testing set, and the same process was repeated six times, each with a different division

between the training and testing sets. Although each time ends up with a different set of parameters, they are fairly close to one another and result in the same ranking orders for all the training and testing sets. In the end, we selected $a = 0.8012$, $\alpha = 0.3046$ and $\beta = 0.7088$ as our final model parameters.

3.2 Validation

The validation process is conducted by comparing our objective quality assessment results with subjective data. The following evaluation metrics are employed:

- Spearman's rank-order correlation coefficient (SRCC) is defined as

$$\text{SRCC} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{3.15}$$

  where d_i is the difference between the i-th image's ranks in subjective and objective evaluations. SRCC is a non-parametric rank-order based correlation metric, independent of any monotonic nonlinear mapping between subjective and objective scores.

- Kendall's rank-order correlation coefficient (KRCC) is another non-parametric rank correlation metric computed as:

$$\text{KRCC} = \frac{N_c - N_d}{\frac{1}{2}N(N-1)}, \tag{3.16}$$

  where N_c and N_d are the numbers of concordant (of consistent rank order) and discordant (of inconsistent rank order) pairs in the data set, respectively.

The proposed TMQI is the only objective quality measure being tested. To the best of our knowledge, almost no other method has been proposed to compare images with different dynamic ranges. The only exception is the method proposed in [8], which creates probability maps to distinguish between visible (suprathreshold) and invisible (subthreshold)

degradations. The probability maps are shown to be useful in classifying image distortion types but are not meant to be pooled to produce an overall quality score for a tone-mapped image. As a result, direct comparison with the proposed method is not possible.

Three experiments have been carried out in our validation process, each using a different subject-ranked database. The first database is from [1], and was also used in the parameter training step discussed in Section 3.1.3. Our leave-one-out cross validation method described in Section 3.1.3 creates SRCC and KRCC values for each of the six testing data sets, where for each data set, the parameters were trained using the other five data sets. Table 3.1 shows the means and standard deviations of KRCC and SRCC values between subjective rankings and our model predictions, respectively.

Table 3.1: Cross validation results using data from [1]

	KRCC	SRCC
Mean	0.7333	0.8333
Std	0.1632	0.1211

In the second experiment, we use the database introduced in [2,3], from which we employ the overall quality ranking data by ten naive subjects, of 14 tone-mapped images created from the same HDR image. The KRCC and SRCC values between subjective rankings of the images and our structural fidelity, statistical naturalness and overall quality scores are given in Table 3.2, where we observe that the structural fidelity measure alone can provide reasonable predictions of subjective rankings. The statistical naturalness measure by itself is not a good predictor of the overall quality ranking, but it complements the structural fidelity measure. When the two measures are combined, better prediction of the overall image quality is achieved. It is worth mentioning that the test data here is not used in the training process, but the resulting KRCC and SRCC values are comparable with those obtained in the test using the first database, which is used for training. This implies good generalization capability of the training method described in Section 3.1.3.

Table 3.2: Performance evaluation using data from [2, 3]

	KRCC	SRCC
Structural Fidelity	0.6923	0.7912
Statistical Naturalness	0.3846	0.5385
Overall Quality	0.7179	0.8187

The third experiment was conducted using a database developed by ourselves. Twenty subjects were provided with 15 sets of tone mapped images, each of which included eight images generated by eight TMOs from the same HDR image. The results created by five of the TMOs developed by Reinhard *et al.* [13], Drago *et al.* [4], Durand & Dorsey [12], Mantiuk *et al.* [11] and Pattanaik *et al.* [28] were computed using the publicly available software Luminance HDR [77]. In addition, three other images were created using the built-in TMOs from Adobe Photoshop, namely "Exposure and Gamma", "Equalize Histogram", and "Local Adaptation", respectively. The parameters used in all eight TMOs were set as their default values and are not optimized. The reference HDR images were selected to represent different indoor and outdoor scenes and are all available online [2, 78–80]. In the subjective test, each of the 20 observers was asked to rank the eight images in each image set from the best to the worst. The subjective rankings for each image were then averaged, resulting in its mean ranking score within the set.

To evaluate the TMQI method, we calculate the KRCC and SRCC values between the mean ranking scores and the objective quality measures for each image set. The results are given in Table 3.3. To provide an anchor in evaluating the performance of TMQI, we compare it with the behavior of an average subject. To do this, we first compute the KRCC and SRCC values between the mean ranking scores and the ranking scores given by each individual subject for each image set. We then compute the mean and standard deviation of these KRCC and SRCC values over subjects(Table 3.3). The average KRCC and SRCC values over all 15 image sets are given in the last row. It can be seen that for all image sets, the KRCC and SRCC values of TMQI are well within the range of ± 1

Table 3.3: Performance evaluations using 15 image sets and 8 TMOs

Image Set	KRCC			SRCC		
	mean subject performance	std of subject performance	TMQI performance	mean subject performance	std of subject performance	TMQI performance
1	0.8071	0.1038	0.7857	0.9071	0.0650	0.9048
2	0.7269	0.2072	0.6429	0.8251	0.1709	0.7857
3	0.7642	0.1064	0.6429	0.8797	0.0758	0.8095
4	0.8107	0.1141	0.7143	0.9130	0.0746	0.8571
5	0.4714	0.2116	0.6429	0.6000	0.2030	0.7381
6	0.6464	0.1646	0.7857	0.7630	0.1707	0.9048
7	0.7250	0.1275	0.5714	0.8285	0.1006	0.6905
8	0.7000	0.1862	0.5714	0.8023	0.1813	0.6905
9	0.6607	0.1978	0.5714	0.7857	0.1625	0.7619
10	0.8418	0.0991	0.7857	0.9276	0.0581	0.9048
11	0.7428	0.1815	0.7143	0.8523	0.1352	0.8810
12	0.6250	0.2084	0.5714	0.7595	0.2055	0.7143
13	0.5637	0.2298	0.5455	0.6970	0.2343	0.6587
14	0.6214	0.1720	0.6429	0.7702	0.1474	0.7381
15	0.8142	0.0994	0.7857	0.9035	0.0705	0.9048
Average	**0.7014**	**0.1606**	**0.6649**	**0.8143**	**0.1368**	**0.7963**

standard deviation from the KRCC and SRCC values of the mean over all subjects. This finding indicates that TMQI behaves quite similarly to an average subject.

Since the TMQI algorithm does not involve any expensive searching or iterative procedures, it is computationally efficient. Our unoptimized MATLAB implementation on an Intel Quad-Core 2.67GHz computer takes on average around 0.75 and 2.7 seconds to evaluate images of sizes 512×512 and 1024×1024, respectively. Figure 3.5 illustrates the scatter plot of runtime versus the number of image pixels for 20 HDR-LDR comparisons. It shows that the computational complexity of the TMQI algorithm is approximately linear with respect to the number of pixels in the image. The scatter plot of runtime versus the square root of window size is depicted in Figure 3.6. It can be seen that window size does not affect the runtime too much, and even for relatively large window size (35×35) the TMQI can be computed in almost a second. The relatively low computational cost makes TMQI easily adapted to practical applications that involve iterative optimization processes.

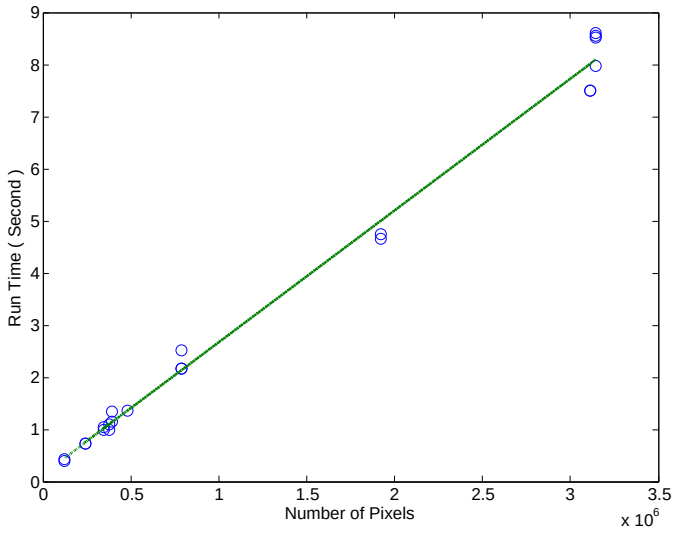


Figure 3.5: Run time versus the number of image of the proposed algorithm.

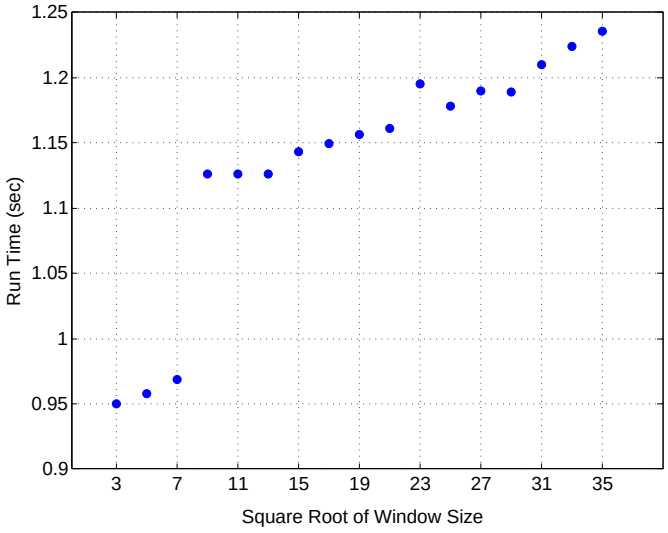


Figure 3.6: Run time versus window size of the proposed algorithm.

3.3 Applications

The application scope of objective IQA measures goes beyond evaluating images and comparing algorithms. A wider range of applications extends to developing novel image processing algorithms optimized for novel IQA measures. In this section, we use two examples to demonstrate the potential of TMQI.

3.3.1 Parameter Tuning in TMO Algorithm

Many TMOs contain one or more parameters whose optimal values are often image-dependent. Without human interference, it is often difficult to choose these parameters, while varying the parameters could lead to drastically different results. An objective quality measure provides a useful tool for picking these parameters automatically. Here, we use the TMO proposed in [4] as an example. It uses a logarithmic function with varying bases

in different locations to change the dynamic range adaptively. The algorithm is given by

$$L_{\mathrm{d}} = \frac{L_{\mathrm{dmax}} \cdot 0.01}{\log_{10}(L_{\mathrm{wmax}} + 1)} \cdot \frac{\log(L_{\mathrm{w}} + 1)}{\log\left(2 + \left(\left(\frac{L_{\mathrm{w}}}{L_{\mathrm{wmax}}}\right)^{\frac{\log(b)}{\log(0.5)}}\right) \cdot 8\right)}, \tag{3.17}$$

where L_{w} and L_{wmax} are the world luminance and maximum luminance of the scene, L_{d} and L_{dmax} are the display luminance and maximum luminance of the display, respectively, and b is a tuning parameter. The perceptual quality of the tone-mapped image varies significantly with b. However, in the literature, the b value is typically fixed at around 0.8 through empirical experimenting with multiple images [4, 77].

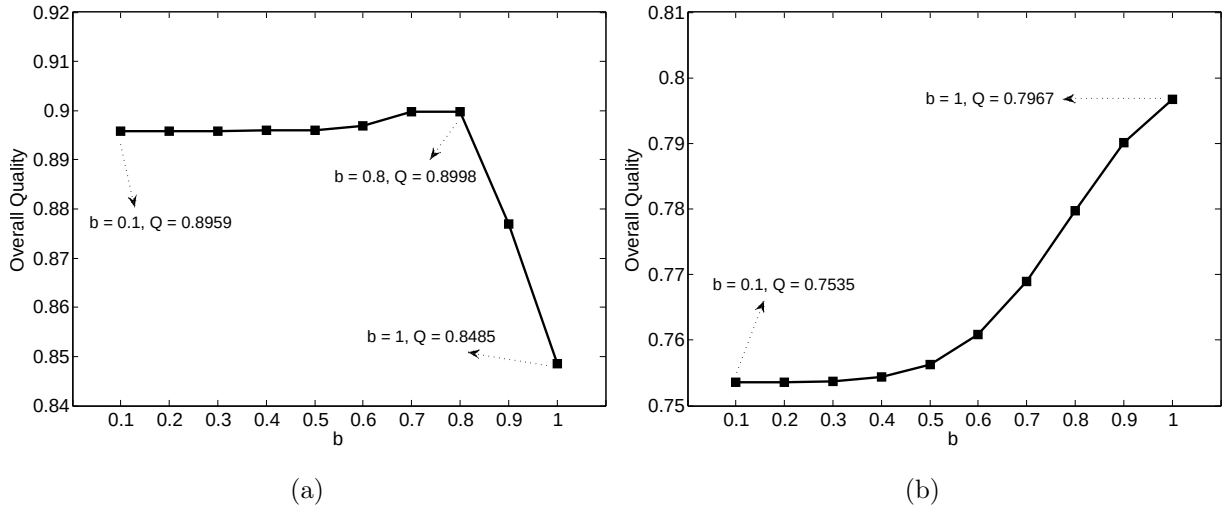
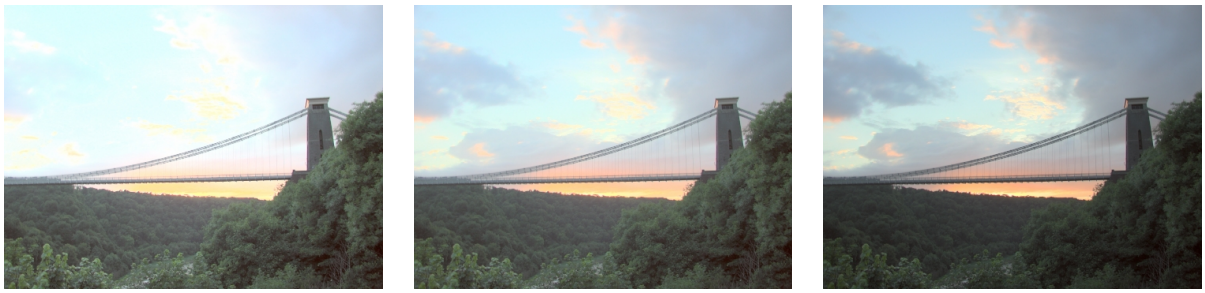


Figure 3.7: Overall quality measure Q versus parameter b for "Desk" (a) and "Bristol Bridge" (b) images. The tone mapped images corresponding to selected b values are shown in Figures 3.8 and 3.9, respectively.

In Figures 3.7(a) and 3.7(b), we plot how TMQI varies as a function of b for the images "Desk" and "Bristol Bridge", respectively (No computation beyond $b = 1$ is conducted because it is beyond the value range suggested by the algorithm). It appears that the quality score behaves quite differently as a function of b. Based on the plots, $b = 0.8$ and $b = 1$ are picked as the optimal values for the two images, respectively. These results confirm

(a) (b) (c)

Figure 3.8: LDR images generated with different parameter b in (3.17). (a) $b = 0.1$, $S = 0.8344$, $N = 0.4599$ and $Q = 0.8959$; (b) $b = 0.8$, $S = 0.8448$, $N = 0.4874$ and $Q = 0.8998$; (c) $b = 1.0$, $S = 0.8337$, $N = 0.1423$ and $Q = 0.8485$.



(a) (b) (c)

Figure 3.9: LDR images generated with different parameter b in (3.17). (a) $b = 0.1$, $S = 0.5214$, $N = 0.0249$ and $Q = 0.7535$; (b) $b = 0.7$, $S = 0.8137$, $N = 0.1136$ and $Q = 0.7690$; (c) $b = 1.0$, $S = 0.8856$, $N = 0.2923$ and $Q = 0.7967$.

that the optimal b value is close to the empirical value (around 0.8) selected in previous studies, but varies for different images. The tone mapped LDR images corresponding to the three selected b values are shown in Figure 3.8 and Figure 3.9, respectively. Careful inspection of these images shows that the best b values lead to a good balance between preserving structural details and producing natural-looking images.

3.3.2 Adaptive Fusion of Tone-mapped Images

When experimenting with different TMOs on different HDR images, we often find it difficult to pick a single TMO that produces the best results for all HDR images. Furthermore, within a single HDR image, the best TMO may also vary when different regions in the image are under consideration. To take advantages of multiple TMOs, image fusion techniques may be employed to combine multiple tone-mapped images, and an objective quality measure can play an important role in this process.

Of the various fusion schemes, we employ the Laplacian pyramid method, which has strong compatibility with the multi-scale fidelity measure. The concept of an image pyramid was discussed in the early 1980s as a fast method of representing the multi-resolution information contained within an image. It also facilitates multi-scale processing similar to that in the HVS [81]. Before long, the relevance of this technique to image-fusion was realized [82]. The Laplacian image fusion scheme starts by constructing Laplacian pyramids for each of the source images. The Laplacian band-pass images from each level are combined by a fusion algorithm. Eventually the final fused image is obtained through inverse Laplacian pyramid decomposition [83].

Given multiple tone-mapped images created by different TMOs, we first apply a Laplacian pyramid transform that decomposes these images into different scales. In the pyramid domain, this step results in multiple coefficients at the same scale and the same spatial location, each corresponding to a different TMO. Examples are given in the first two rows in Figure 3.10, and demonstrate four-scale Laplacian pyramid decompositions, where the fine scale coefficients (Scales 1-3) represent image details and the coarsest scale coefficients (Scale 4) preserve local mean intensities across space. A fusion strategy can then be applied to combine multiple coefficients into one at each location in each scale before an inverse Laplacian pyramid transform is employed to reconstruct a fused image. Typical fusion schemes aim to locally select the most salient image features [84]. The most widely adopted approaches include averaging the coefficients or picking one of the coefficients with the largest absolute value. The former is more appropriate for locations where the source images are similar, while the latter is often used in locations where the source images are

distinctly different.

Here, we propose a different fusion scheme. The general idea is to use the TMQI as the weighting factor in the fusion process. Let S_j and c_j be the local structural fidelity measure and the Laplacian pyramid transform coefficient computed from the j-th tone-mapped image being fused, respectively. The fused coefficient is computed as

$$c^{(fused)} = \frac{\sum_j S_j c_j}{\sum_j S_j} . \tag{3.18}$$

This is applied to all scales except for the coarsest one, for which we use the statistical naturalness measure as the weighting factor:

$$c^{(fused)} = \frac{\sum_j N_j c_j}{\sum_j N_j} , \tag{3.19}$$

where N_j denotes the statistical naturalness score of the j-th tone-mapped image.

The proposed Laplacian pyramid domain fusion method is demonstrated in the bottom row of Figure 3.10, where the fused image preserves the details in the brightest region (the light area on the top) as in (f), while at the same time maintains higher contrast in relatively darker regions, as in (a). Figure 3.11 provides an example with a natural scene, where one tone-mapped image (a) better preserves structural details, and another (b) gives a more natural overall appearance (but loses structural information, especially in the brightest areas). Three fused images created by three different image fusion algorithms are given in (c), (d) and (e), respectively. The image created by the proposed method achieves the best balance between structure preserving and statistical naturalness, and also results in the best quality score using TMQI.

To further validate the proposed fusion scheme, we have conducted an additional subjective experiment, where ten subjects were asked to rank five sets of tone-mapped images, each of which includes eight images. Seven of these images are generated using the TMOs employed in the third experiment in Section 3.2. Two of these seven TMOs are chosen to produce the eighth image using the proposed fusion method. Table 3.4 compares the

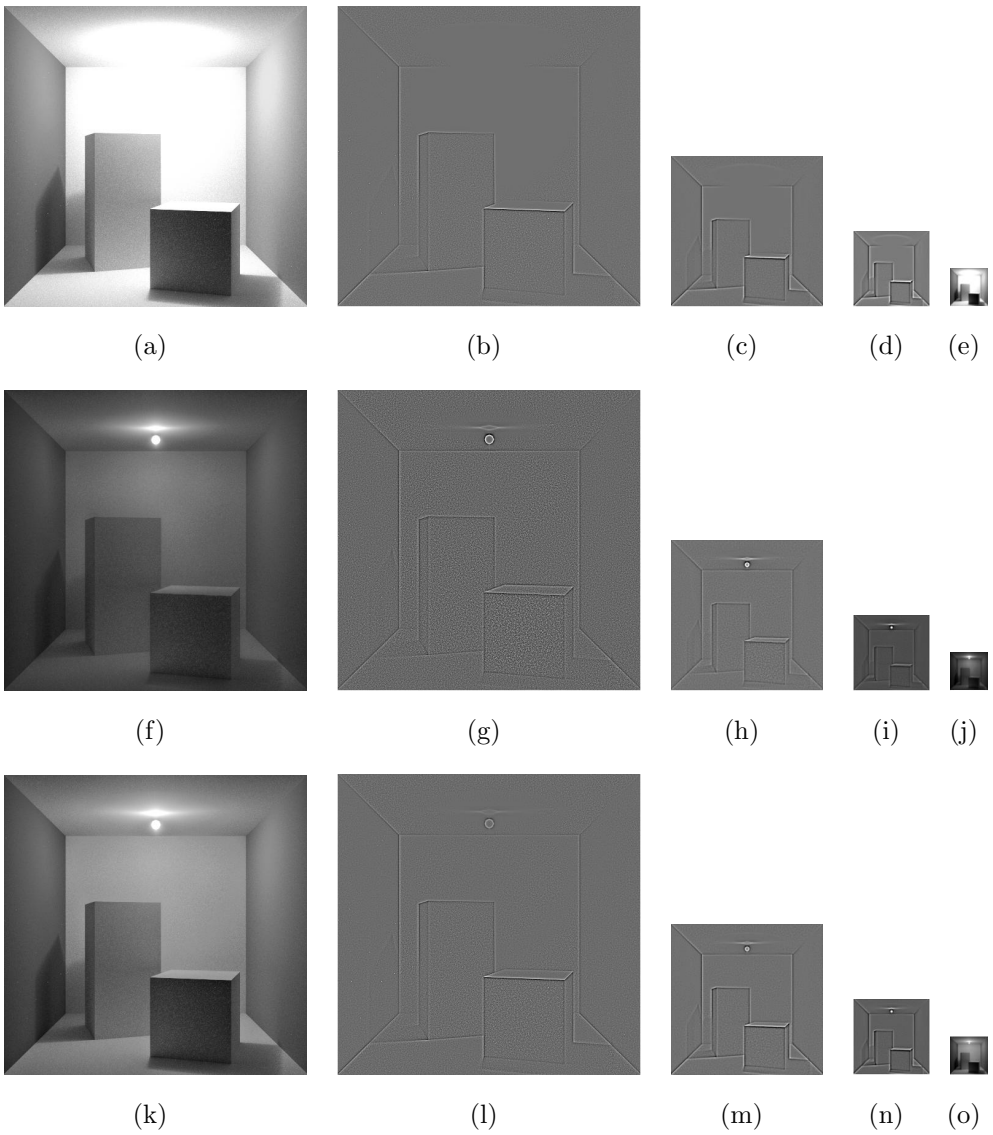


Figure 3.10: Image fusion in Laplacian pyramid domain. Top row: first tone mapped image (a) created by TMO proposed in [11], and its Laplacian pyramid subbands (b)-(e), $S = 0.5034$, $N = 0.1263$, $Q = 0.6937$; Middle row: second tone mapped image (f) using "Exposure and Gamma" method in Adobe Photoshop, and its Laplacian pyramid subbands (g)-(j), $S = 0.6642$, $N = 0.0786$, $Q = 0.7386$; Bottom row: fused image by the proposed method (k), and its its Laplacian pyramid domain representation (l)-(o), $S = 0.7419$, $N = 0.3080$, $Q = 0.8167$.

Table 3.4: Average ranking scores made by 10 subject for each set.

Image Set	Source 1	Source 2	Fused Image
1	4.3	7	1.8
2	5.2	4	1.5
3	3.7	5.9	2.3
4	4.1	6.1	2.2
5	2.7	6.9	3

average subjective rankings of the source images and their corresponding fused images, where lower ranking scores correspond to better quality. It can be seen that the fused image is almost always ranked significantly higher than the two source images being fused.

(a) (b)

(c) (d)

(e)

Figure 3.11: Fusion of tone mapped images. (a) First tone-mapped image using [12], S = 0.8168, N = 0.1631, Q = 0.8075; (b) Second tone-mapped image using the "Exposure and Gamma" method in Adobe Photoshop, S = 0.6315, N = 0.8657, Q = 0.8744; (c) Fused image by coefficient averaging in Laplacian pyramid domain, S = 0.7561, N = 0.7409, Q = 0.8955; (d) Fused image by selecting coefficient of maximal absolute value in Laplacian pyramid domain, S = 0.7685, N = 0.9428, Q = 0.9290; (e) Fused image by the proposed method, S = 0.7836, N = 0.9970, Q = 0.9413.

Chapter 4

Adaptive Windowing for Optimal Visualization of Medical Images Based on a Structural Fidelity Measure

Medical imaging devices often capture raw data with high precision, producing HDR images. To visualize HDR images on regular displays, an increasing number of tone-mapping algorithms have been developed in recent years. These algorithms convert HDR to LDR images. To visualize HDR medical images, a so-called "windowing" procedure is typically employed, by which the structural details within the intensity region of interest are mapped to the dynamic range of regular displays. Such intervals of interest vary for different imaging modality body parts. These intervals can be defined using two parameters: (i) window width, the range of the interval, to be denoted here as W and (ii) the window center, the center of this interval, to be denoted as C. It follows that the tone-mapping algorithm maps the range of luminance values $C - \frac{1}{2}W \leq l \leq C + \frac{1}{2}W$ to the LDR range $[0, 255]$ using a linear function. The default values for window width and window center are embedded in headers of HDR medical image files. These parameters, however, are not optimized for

the visualization of different body parts. In practice, radiologists often adjust the window width and window center manually so that the details for particular body regions become more visible. Linear mapping is the most straightforward windowing operator, but it may not be the optimal mapping function in terms of structure preservation and visualization.

In this chapter, our goal is to produce tone-mapping operators that are superior to the linear mappings currently employed for the purpose of visualizing HDR medical images. Our proposed approach employs two types of continuous, monotonically increasing tone-mapping functions and tunes their parameters to map the structural information within the window width onto a display's dynamic range in an optimal way. The optimization task is carried out by exploiting the structural fidelity measure introduced in Section 3.1.1. The statistical naturalness measure in Section 3.1.2 is not used because naturalness is not a directly relevant attribute of medical images. Moreover, we focus on X-ray computed tomography (x-ray CT) images as they are one of the most common modalities of medical images. Our experiments confirm that the linear mapping function is not optimal in terms of the fidelity of structural information. In addition, they show that modifying the mapping function to obtain maximal structural fidelity measurement produces CT medical images with higher contrast and more visible details.

4.1 Structural Fidelity Measurement for Medical Images

Since the windowing function reduces the dynamic range of an image, not all the information contained in an HDR medical image can be preserved. Human observers, particularly doctors/radiologists, may not be aware of this loss of information. A tool to measure structural fidelity may, therefore, play an important role in assessing the quality of LDR medical images.

Let x and y be two local image patches extracted from an HDR and an LDR medical

image, respectively. In 3.1.1, we defined a local structural fidelity measure as

$$S_{\text{local}}(x,y) = \frac{2\sigma'_x\sigma'_y + C_1}{{\sigma'_x}^2 + {\sigma'_y}^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2}\,. \tag{4.1}$$

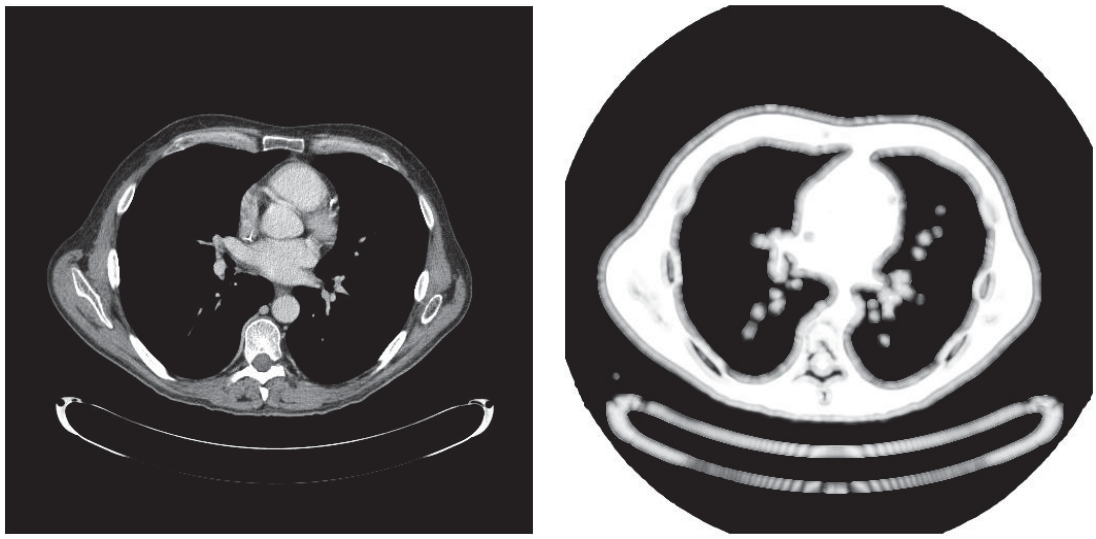
where σ_x, σ_y and σ_{xy} denote the local standard deviations and cross correlation between the two corresponding patches in the HDR and LDR medical images, respectively, and C_1 and C_2 are positive stability constants. The σ' is then used to quantify the significance of local contrast. To do so, we pass the local standard deviation through a nonlinear mapping function, resulting in the σ' value employed in (4.1).

In 3.1.1, Galton's psychometric function was adopted as the nonlinear mapping function and was rewritten in terms of the standard deviation of the signal. As a result, the mapping between σ and σ' is defined as:

$$\sigma' = \frac{1}{\sqrt{2\pi}\theta_\sigma} \int_{-\infty}^{\sigma} \exp\left[-\frac{(t-\tau_\sigma)^2}{2\theta_\sigma^2}\right] dt\,, \tag{4.2}$$

where τ_σ is the contrast threshold and $\theta_\sigma = \tau_\sigma/3$. In 3.1.1, the contrast threshold, τ_σ, was calculated for natural images using a CSF model as well as a contrast sensitivity measurement, assuming a pure sinusoidal stimulus. However, since a judgement about significant and insignificant contrast details in medical images is crucial and the neglect of any important structural information might lead to grave consequences, we prefer here to set the contrast thresholds to be very small. As a result, the structural fidelity method penalizes mappings from non-flat regions to flat regions and vice-versa. In our experiments, we set τ_σ to 1 and 0.5 for HDR and LDR medical images, respectively. In (4.1), σ'_x and σ'_y are the mapped versions of σ_x and σ_y, respectively. They are bounded between 0 and 1, where 0 and 1 represent completely insignificant and completely significant signal strengths, respectively.

The local structural fidelity measure S_{local} is applied to an image using a sliding window that runs across the image, resulting in a map that reflects the variation of structural fidelity across space. Figure 4.1(a) shows a CT image of an abdomen region tone-mapped by a linear function. The quality map produced by the proposed measure is shown in



(a) (b)

Figure 4.1: Medical images compared with the DICOM reference file. (a) Tone mapped image using the linear mapping function S = 0.40. (b) Associated quality map.

Figure 4.1(b). The window width and window center parameters are extracted from the DICOM image header. It is interesting to observe these fidelity maps and examine how they correlate with perceived image fidelity. For example, because of the window width and window center parameters, the structural details in the lung are missing in Figure 4.1(a). In Figure 4.1(b), the quality map in the lung region is black, indicating that some details in the original DICOM image are not mapped into the LDR image. On the other hand, a white region in the boundary illustrates that there is no structural information in the original DICOM image in the corners – therefore, nothing is lost by the linear mapping function. Finally, the components of the quality map are averaged to provide a single score – the overall structural fidelity-based quality measure,

$$S = \frac{1}{N} \sum_{i=1}^{N} S_{\text{local}}(x_i, y_i)\,, \tag{4.3}$$

where x_i and y_i are the i-th patches in the HDR and LDR medical images being compared,

respectively, and N is the number of patches. To implement the proposed algorithm, we set $C_1 = 0.01$, $C_2 = 10$, and employ a Gaussian sliding window of size 11×11 with standard deviation 1.5 to create the quality map.

The main advantage of the structural fidelity measure described above is the ability to compare LDR and HDR medical images without creating an LDR image as a reference. This provides a useful tool for medical imaging since radiologists do not have to produce an LDR reference image each time the window width and window center are adjusted. By contrast, commonly employed quality metrics such as PSNR and SSIM have to compare the test image with an LDR reference image generated by a windowing process. In addition, the quality map indicates the performance of tone mapping or image processing algorithms in the regions of interest. For example, in Figure 4.1, the quality maps reflect the quality of heart and tissue regions regardless of the black background which is of no interest.

4.2 Finding the Optimal Windowing Function

Let x be the original HDR image; l_l and l_u be the lower and upper bounds of the window range, respectively; f be the windowing (or intensity mapping) function lives in the space defined by

$$\mathcal{F}_{[l_l,l_u]} = \{f : [l_l, l_u] \to [0,1] \mid f \text{ continuous \& monotonically increasing}\}\,; \tag{4.4}$$

$T_f(.)$ be the tone-mapping operator that applies the function f pointwise to an image and quantizes the mapped value to the dynamic range of the LDR display; and $S(\cdot,\cdot)$ be the structural fidelity measure defined in the previous section. Our goal is to search for the optimal mapping function f in terms of $S(x, T_f(x))$:

$$f_{opt} = \underset{f \in \mathcal{F}_{[l_l,l_u]}}{\arg\max}\, S(x, T_f(x))\,. \tag{4.5}$$

Here we consider only two subsets of the function space $\mathcal{F}_{[l_l,l_u]}$ (i) piecewise linear functions and (ii) functions spanned by an appropriate family of sine functions.

4.2.1 Windowing function using piecewise linear basis

For simplicity, we consider piecewise linear functions defined by an equipartition of the HDR intensity range $[l_l, l_u]$ into n subintervals $I_k = [l_{k-1}, l_k]$ for $1 \leq k \leq n$ of length $\Delta l = (l_u - l_l)/n$. The partition points are defined by $l_k = l_l + k\Delta l$, $0 \leq k \leq n$, and so $l_l = l_0$ and $l_u = l_n$. The window width and window center are, respectively,

$$W = l_n - l_0 = n\Delta l, \quad C = \frac{1}{2}(l_0 + l_n) = l_0 + \frac{n\Delta l}{2}. \tag{4.6}$$

Every such equipartition piecewise linear function can be expressed as a linear combination of n basis functions. The first basis function is a "ramp" function that corresponds to a direct linear mapping in the full range:

$$\phi_0(l) = \begin{cases} (l - l_0)/W, & l_0 \leq l \leq l_n; \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

The other $n - 1$ basis functions are defined in terms of the standard triangle or "hat" function given by

$$t(l) = \begin{cases} 1 - |l|, & -1 \leq l \leq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{4.8}$$

Specifically, we have

$$\phi_k(l) = t\left(\frac{l - l_k}{\Delta l}\right), \text{ for } k = 1, \cdots, n-1 \tag{4.9}$$

As such, any equipartition piecewise linear function can be expressed as

$$f(l) = \sum_{k=0}^{n-1} c_k \phi_k(l) = \phi_0(l) + \sum_{k=1}^{n-1} c_k \phi_k(l)\,, \tag{4.10}$$

where the value of c_0 is known to be 1. In order for the function to be monotonically increasing, we need $0 \leq \cdots \leq f(l_{k-1}) \leq f(l_k) \leq \cdots \leq 1$, which yields

$$0 \leq \cdots \leq c_{k-1} + \frac{k-1}{n} \leq c_k + \frac{k}{n} \leq \cdots \leq 1\,. \tag{4.11}$$

For example, in the case that $n = 3$, we can derive

$$\begin{cases} c_1 \geq -\frac{1}{3}; \\ c_2 - c_1 \geq -\frac{1}{3}; \\ c_2 \leq \frac{1}{3}. \end{cases} \tag{4.12}$$

4.2.2 Windowing function using family of sine basis

The windowing function may also be expressed using a linear combination of a family of sine basis functions defined by

$$\phi_k(l) = \sin\left(\frac{k\pi(l-l_l)}{W}\right) \text{ for } l_l \leq l \leq l_u \text{ and } k = 1, 2, \cdots \tag{4.13}$$

We then obtain an n-th order approximation of any f in $\mathcal{F}_{[l_l,l_u]}$ using the same expression as in (4.10), with the only difference being that the triangle basis functions are replaced by the sine basis functions.

As a special case, when $n = 3$, we have

$$f(l) = \frac{l-l_l}{W} + c_1 \sin\left(\frac{\pi(l-l_l)}{W}\right) + c_2 \sin\left(\frac{2\pi(l-l_l)}{W}\right)\,, \tag{4.14}$$

To ensure that $f(l)$ is monotonically increasing, its derivative needs to be no less than 0:

$$f'(l) = \frac{1}{W} + \frac{\pi c_1}{W}\cos\left(\frac{\pi(l-l_l)}{W}\right) + \frac{2\pi c_2}{W}\cos\left(\frac{2\pi(l-l_l)}{W}\right) \geq 0\,. \tag{4.15}$$

To find the extrema l^* of $f'(l)$, we set its derivative to 0, which yields

$$f''(l^*) = -\frac{\pi^2 c_1}{W^2}\sin\left(\frac{\pi(l^* - l_l)}{W}\right) - \frac{4\pi^2 c_2}{W^2}\sin\left(\frac{2\pi(l^* - l_l)}{W}\right) = 0\,. \tag{4.16}$$

Expanding the second term, we obtain

$$\sin\left(\frac{\pi(l^* - l_l)}{W}\right)\left[c_1 + 8c_2\cos\left(\frac{\pi(l^* - l_l)}{W}\right)\right] = 0\,, \tag{4.17}$$

for which we have three possible solutions:

$$l^* = l_l\,, \tag{4.18}$$

$$l^* = l_u\,, \tag{4.19}$$

$$\cos\left(\frac{\pi(l^* - l_l)}{W}\right) = \frac{-c_1}{8c_2}\,. \tag{4.20}$$

From (4.20), we have

$$\cos\left(\frac{2\pi(l^* - l_l)}{W}\right) = 2\left(\frac{-c_1}{8c_2}\right)^2 - 1 = \frac{c_1^2}{32c_2^2} - 1\,. \tag{4.21}$$

Substituting (4.18), (4.19) and the pair (4.20) and (4.21) into (4.15), we obtain the three constraints on the solutions of c_1 and c_2:

$$\begin{cases} c_1 + 2c_2 \geq -\frac{1}{\pi} \\ -c_1 + 2c_2 \geq -\frac{1}{\pi} \\ \frac{c_1^2}{16c_2} + 2c_2 \leq \frac{1}{\pi} \end{cases}. \tag{4.22}$$

4.2.3 Finding optimal windowing functions

With the two types of windowing functions defined in the previous subsections, the problem of finding f_{opt} in (4.5) is converted to finding the best set of coefficients c_k's for the basis functions. This can be done by substituting (4.10) into (4.5) and solving it using numerical

optimization tools (e.g., Matlab's *fmincon* function) under appropriate constraints; for example, in the case $n = 3$, the constraints are given by (4.11) and (4.22) for the piecewise linear and sine basis functions, respectively.

To demonstrate the proposed optimization methods, Figure 4.2 (a) shows the result of linear mapping ($S = 0.8853$), where the window width and window center parameters are preset values embedded in the DICOM header. Our optimization algorithm does not change these parameters, but attempts to find the optimal values for c_1 and c_2. Figure 4.2 (b) illustrates the result of optimal piecewise linear mapping, where the best coefficients are given by $c_1 = -0.15$, and $c_2 = -0.01$ and S = 0.9294 is obtained. Enhanced contrast in the image is observed, in which the details in the spine and the lung are more discernable. Using the family of sine bases, Figure 4.2 (c) is obtained for optimal coefficients $c_1 = -0.0001$ and $c_2 = -0.16$ with an even higher structural fidelity measure S = 0.9446, producing an image with higher contrast and more visible details.

As mentioned earlier, the window width and the window center parameters in the HDR file header do not necessarily provide a desirable contrast for specific body parts such as the lungs, bones, soft tissues and brain. In practice, radiologists often change them manually for different body parts, in order to visualize the desired region with appropriate contrast. Figure 4.3 (a) is a tone-mapped image using the DICOM standard windowing procedure with predefined values for bone, where S = 0.7746. The result of our optimization method using piecewise linear windowing is shown in Figure 4.3(b), where $c_1 = 0.56$, $c_2 = 0.28$ and $S = 0.99$. Figure 4.3 (c) shows the image produced by optimal sine basis windowing with coefficients $c_1 = 0.37$ and $c_2 = 0.04$ and quality measure $S = 0.9852$. It can be observed that the performance of the optimization task using either approaches provides images with strong contrast enhancement. The optimal windowing curve in Figure 4.3 (d) reveals that the intensity of CT bone images is concentrated in the middle of the window width. Since piecewise linear functions can model drastically increasing functions, these functions work slightly better than the sine bases.

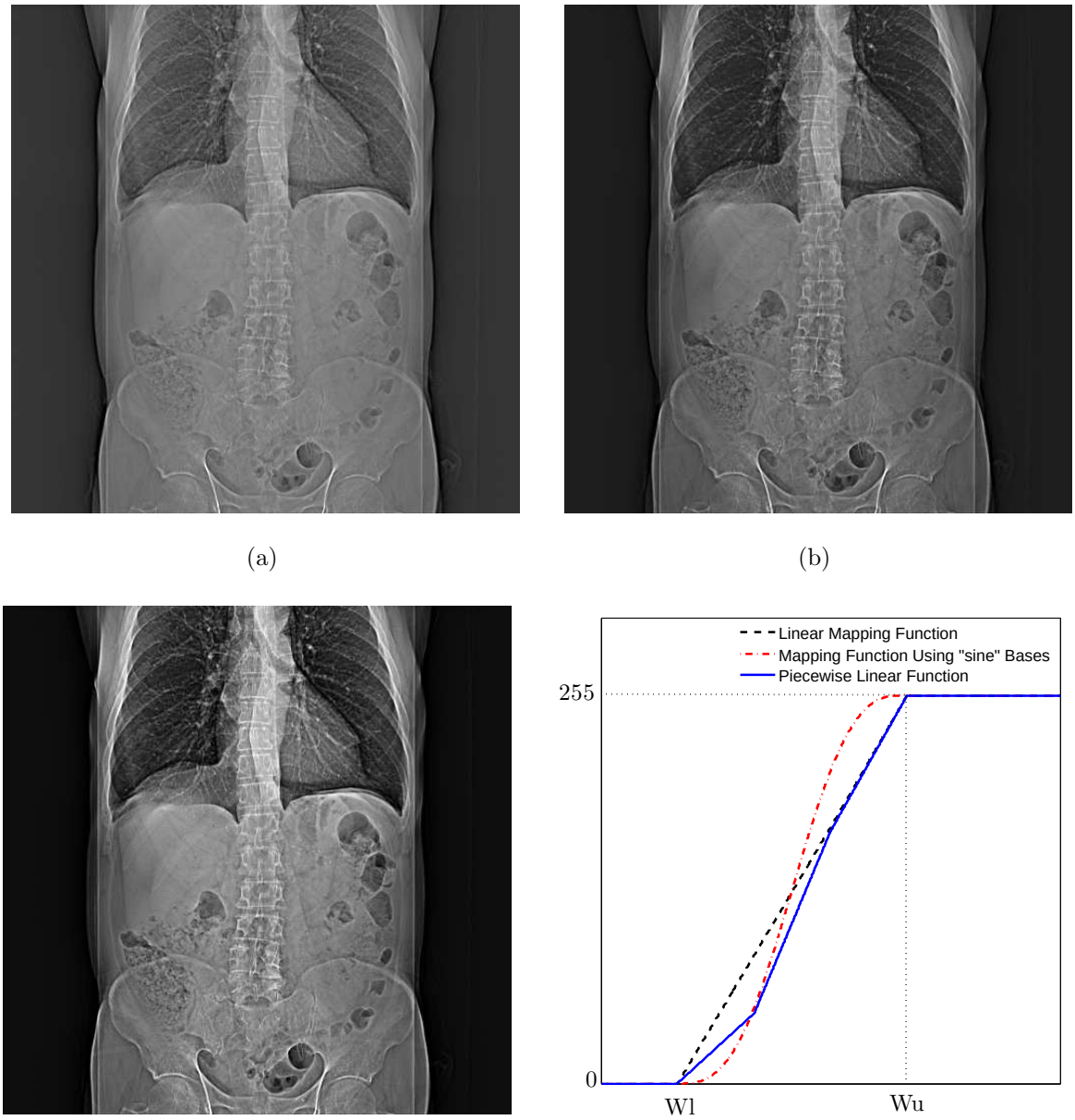


Figure 4.2: Results of the optimization method, (a) where S = 0.8853 is the tone-mapped image using linear mapping function, where the window width and window center are read from DICOM file header. (b) with $S = 0.9292$ and (c) with $S = 0.9446$ are the enhanced images employing functions in (4.10) and (4.14), respectively. Image courtesy of AGFA Healthcare Inc.

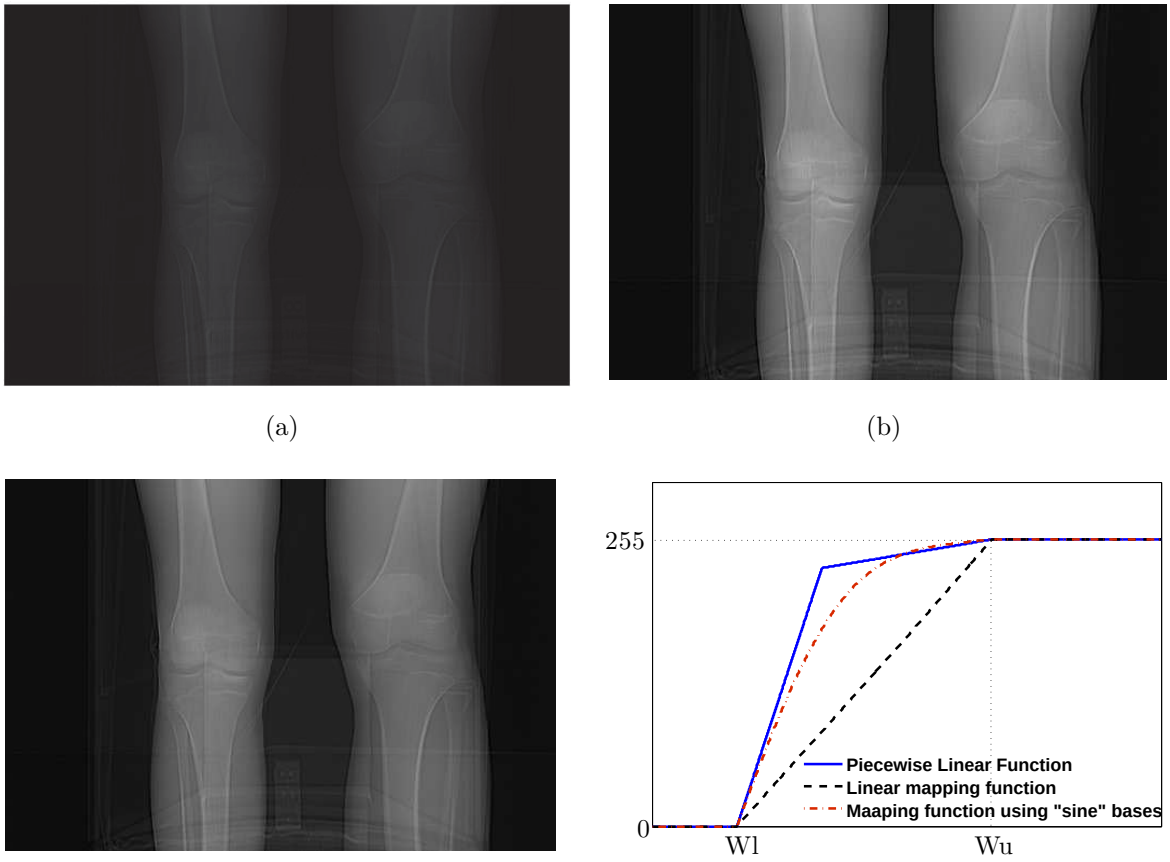


Figure 4.3: Results of the optimization method for bone, (a) where S = 0.7746 is the linearly mapped image using predefined windowing parameters for bone (window width = 2000, window center = 500).(b) with $S = 0.99$ and (c) with $S = 0.9852$ are the enhanced images using functions in (4.10) and (4.14), respectively. Image courtesy of AGFA Healthcare Inc.

Chapter 5

Optimization-Based High Dynamic Range Image Tone-Mapping

Many existing TMOs produce visually appealing images. However, most of them are not able to completely map structural details in HDR images, and usually some structural information is missing in the tone-mapped images. In this chapter, we propose a substantially different tone-mapping approach, whereby instead of explicitly designing a new computational structure for TMO, we search in the space of images to find better-quality images in terms of objective measures described in Chapter 3. Specifically, we explicitly treat tone-mapping as an optimization problem in the image space and propose an iterative search approach that starts from any initial image and moves step-by-step in the image space, towards improving the proposed objective quality measures until a (local) maximal point is reached.

A tone-mapped image quality index (TMQI) was proposed in Chapter 3 and has shown to have good correlations with subjective evaluations of tone-mapped images. The TMQI consists of two components, structural fidelity and statistical naturalness. In this chapter, we first exploit the structural fidelity measure as the objective quality assessment tool. When applied to initial images generated by existing and state-of-the-art TMOs, our algorithm almost always enhances the visibility of image details and improves the structural

fidelity measure. Indeed, it often restores image structures that are missing in the images produced by state-of-the-art TMOs. Although the performance of the proposed approach is quite promising, it sometimes restore only structural details in tone-mapped images. However, the design philosophy behind TMQI asserts that the perceptual quality of tone-mapped images is determined by both structural fidelity and statistical naturalness terms. Therefore, restoring structural information from HDR images is not guaranteed to produce the best tone-mapped images. Therefore, in the second part of this chapter, we extend the proposed optimization framework to develop a novel TMO that utilizes the full TMQI as the optimization goal. This new approach involves an iterative process that alternatively improves the structural fidelity and statistical naturalness of the resulting image, which are the two fundamental building blocks in TMQI. Experiments show that this approach leads to consistent enhancement of the perceptual quality of tone-mapped images, and produces better-quality images upon a wide variety of initial images, including those produced by state-of-the-art TMOs.

5.1 Tone-Mapping by Structural Fidelity Maximization

In Chapter 3, we introduced a structural fidelity measure that not only provides an overall structural fidelity of a tone-mapped image, but also produces a structural fidelity map that indicates how well the local structural details are preserved at each spatial location.

Let $\mathbf{x}$ and $\mathbf{y}$ be two image patches extracted from the HDR and the LDR images, respectively. A local structural fidelity measure is defined as

$$S_{\text{local}}(\mathbf{x}, \mathbf{y}) = \frac{2\tilde{\sigma}_x\tilde{\sigma}_y + C_1}{\tilde{\sigma}_x^2 + \tilde{\sigma}_y^2 + C_1} \cdot \frac{\sigma_{xy} + C_2}{\sigma_x\sigma_y + C_2}, \tag{5.1}$$

where σ_x, σ_y and σ_{xy} denote the local standard deviations and cross correlation between the two corresponding patches in the HDR and LDR images, respectively. C_1 and C_2 are positive constants used to avoid instabilities at low-energy regions. As described in 3.1.1,

to gauge the significance of local contrast, we pass the local standard deviation σ through a nonlinear mapping function, resulting in the $\tilde{\sigma}$ value in (5.1). The nonlinear mapping function is given by

$$\tilde{\sigma} = \frac{1}{\sqrt{2\pi}\theta_\sigma} \int_{-\infty}^{\sigma} \exp\left[-\frac{(t-\tau_\sigma)^2}{2\theta_\sigma^2}\right] dt\,, \tag{5.2}$$

where τ_σ is a contrast threshold and $\theta_\sigma = \tau_\sigma/3$.

The local structural fidelity measure S_{local} is applied using a sliding window that runs across the image, resulting in a map that reflects the variation of structural fidelity across space. Figure 5.1(f) shows an example of such a structural fidelity map computed for a tone mapped "memorial" image Figure 5.1(a). The structural fidelity map is reasonably consistent with visual perception. For example, due to overexposure, the structural details of the brightest window region are missing, as is well indicated in the map. Finally, the quality map is averaged to provide a single overall structural fidelity measure of the image:

$$S(\mathbf{X},\mathbf{Y}) = \frac{1}{M}\sum_{i=1}^{M} S_{\text{local}}(\mathbf{R}_i\mathbf{X},\mathbf{R}_i\mathbf{Y}) = \frac{1}{M}\sum_{i=1}^{M} S_{\text{local}}(\mathbf{x}_i,\mathbf{y}_i)\,, \tag{5.3}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are column vectors representing the HDR and the tone-mapped LDR images, respectively. $\mathbf{R}_i$ denotes the matrix that extracts the i-th patch from the image; $\mathbf{x}_i = \mathbf{R}_i\mathbf{X}$ and $\mathbf{y}_i = \mathbf{R}_i\mathbf{Y}$ are column vectors of length N representing the i-th patches extracted from the HDR and LDR images, respectively; and M is the total number of patches. Following 3.1.1, we set $C_1 = 0.01$, $C_2 = 10$, and employ a Gaussian sliding window of size 11×11 (and thus $N = 121$) with standard deviation 1.5 to create the fidelity map.

It was shown in 3.1.1 that the structural fidelity measure described above is well correlated with subjective quality evaluations of LDR images, and its performance is statistically similar to that of an average human subject.

5.1.1 Tone-Mapping as an Optimization Problem in the Image Space

If the purpose of tone-mapping is to achieve the best structural fidelity, then optimal TMO can be formulated as a maximum structural fidelity (MSF) problem given by

$$\mathbf{Y}_{\text{MSF}} = \arg\max_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})\,. \tag{5.4}$$

This is an optimization problem in high dimension space (the same dimension as the number of pixels in the images), and finding the global optimal is generally difficult. Assuming smooth and regular behavior of the structural fidelity function, here we propose to use a gradient ascent algorithm to search for local optimal solutions. Similar gradient-based approaches for SSIM optimization have been studied previously for the purpose of comparing competing image quality measures but have not been explored in the context of image quality enhancement or high dynamic range imaging [85, 86].

Given an initial guess image $\mathbf{Y}_0$, we use an iterative algorithm to search along the gradient ascent direction. At the k-th iteration, the solution is updated by

$$\mathbf{Y}_k = \mathbf{Y}_{k-1} + \lambda\, \nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})|_{\mathbf{Y}=\mathbf{Y}_{k-1}}\,, \tag{5.5}$$

where $\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})|_{\mathbf{Y}=\mathbf{Y}_{k-1}}$ is the gradient of $S(\mathbf{X}, \mathbf{Y})$ with respect to $\mathbf{Y}$ in the previous solution $\mathbf{Y}_{k-1}$, and λ is a constant that determines the speed of movement along the gradient direction.

To compute the gradient $\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})$, we start from the local structural fidelity and rewrite (3.1) as

$$S_{\text{local}}(\mathbf{x}, \mathbf{y}) = \frac{A_1\, A_2}{B_1\, B_2}\,, \tag{5.6}$$

where

$$A_1 = 2\tilde{\sigma}_x\tilde{\sigma}_y + C_1 \tag{5.7}$$

$$B_1 = \tilde{\sigma}_x^2 + \tilde{\sigma}_y^2 + C_1 \tag{5.8}$$

$$A_2 = \sigma_{xy} + C_2 \tag{5.9}$$

$$B_2 = \sigma_x\sigma_y + C_2\,. \tag{5.10}$$

Since both image patches are represented as column vectors of length N, we have

$$\mu_y = \frac{1}{N}\mathbf{1}^T\,\mathbf{y} \tag{5.11}$$

$$\sigma_y^2 = \frac{1}{N}(\mathbf{y}-\mu_y)^T(\mathbf{y}-\mu_y) \tag{5.12}$$

$$\sigma_{xy} = \frac{1}{N}(\mathbf{x}-\mu_x)^T(\mathbf{y}-\mu_y)\,. \tag{5.13}$$

The gradient of the local structural fidelity measure with respect to $\mathbf{y}$ can then be expressed as

$$\begin{aligned}\nabla_{\mathbf{y}} S_{\text{local}}(\mathbf{x},\mathbf{y}) &= \frac{(A_1'A_2 + A_1A_2')}{B_1B_2} \\ &- \frac{(B_1'B_2 + B_1B_2')A_1A_2}{(B_1B_2)^2}\,,\end{aligned} \tag{5.14}$$

where

$$A_1' = \nabla_{\mathbf{y}}A_1\,,\ B_1' = \nabla_{\mathbf{y}}B_1\,,\ A_2' = \nabla_{\mathbf{y}}A_2\,,\ B_2' = \nabla_{\mathbf{y}}B_2\,. \tag{5.15}$$

Noting that

$$\nabla_{\mathbf{y}}\sigma_y = \frac{1}{N\sigma_y}(\mathbf{y}-\mu_y) \tag{5.16}$$

$$\nabla_{\mathbf{y}}\sigma_{xy} = \frac{1}{N}(\mathbf{x}-\mu_x)\,, \tag{5.17}$$

we have

$$\begin{aligned} A_1' &= 2\tilde{\sigma}_x \nabla_{\mathbf{y}} \tilde{\sigma}_y \\ &= \frac{2\tilde{\sigma}_x}{\sqrt{2\pi}\theta_\sigma} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] \cdot \nabla_{\mathbf{y}} \sigma_y \\ &= \sqrt{\frac{2}{\pi}} \frac{\tilde{\sigma}_x}{N\theta_\sigma \sigma_y} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] (\mathbf{y} - \mu_y)\,, \end{aligned} \tag{5.18}$$

$$\begin{aligned} B_1' &= 2\tilde{\sigma}_y \nabla_{\mathbf{y}} \tilde{\sigma}_y \\ &= \sqrt{\frac{2}{\pi}} \frac{\tilde{\sigma}_y}{N\theta_\sigma \sigma_y} \exp\left[-\frac{(\sigma_y - \tau_\sigma)^2}{2\theta_\sigma^2}\right] (\mathbf{y} - \mu_y)\,, \end{aligned} \tag{5.19}$$

$$A_2' = \frac{1}{N}(\mathbf{x} - \mu_x)\,, \tag{5.20}$$

$$B_2' = \sigma_x \nabla_{\mathbf{y}} \sigma_y = \frac{\sigma_x}{N\sigma_y}(\mathbf{y} - \mu_y)\,. \tag{5.21}$$

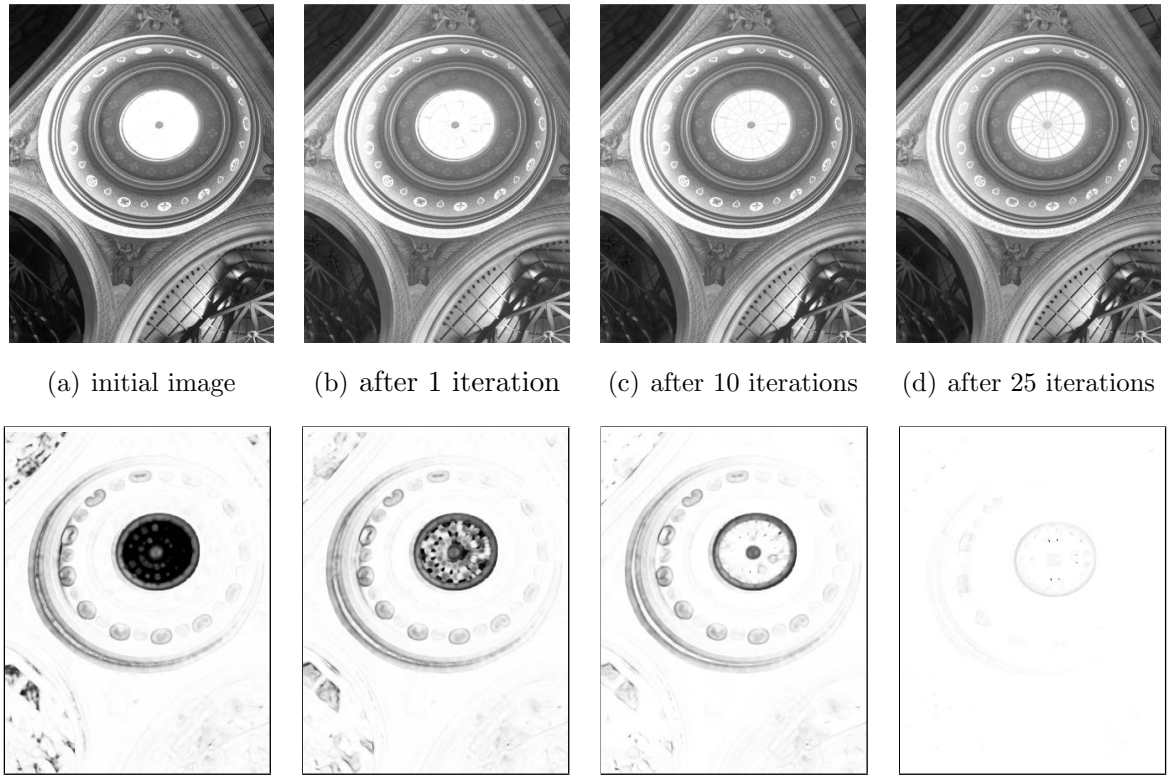
Plugging (5.7), (5.8), (5.9), (5.10), (5.18), (5.19), (5.20) and (5.21) into (5.14), we obtain the gradient of local structural fidelity. Finally, summing all the local gradients together, we can compute the the gradient of the overall structural fidelity measure with respect to the LDR image $\mathbf{Y}$ as

$$\nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{R}_i^T \nabla_{\mathbf{y}} S_{\text{local}}(\mathbf{x}, \mathbf{y})|_{\mathbf{x}=\mathbf{x}_i, \mathbf{y}=\mathbf{y}_i}\,, \tag{5.22}$$

which is subsequently plugged into (5.5) to update the solution that is fed into the next iteration.

5.1.2 Experimental Results

There are only two new parameters that need to be determined in our iterative algorithm: λ, which controls the speed of convergence, and the termination threshold T, which



(a) initial image (b) after 1 iteration (c) after 10 iterations (d) after 25 iterations

(e) $S = 0.9157$ (f) $S = 0.9620$ (g) $S = 0.9895$ (h) $S = 0.9914$

Figure 5.1: Tone mapped "Memorial" images and their structural fidelity maps. The initial image (a) was created by Adobe Photoshop's "Exposure and Gamma" method, and (f) is its structural fidelity map, where brighter indicates higher structural fidelity. The top row also shows the images created after the first (b), the 10-th (c) and the 25-th (d) iterations using the proposed algorithm, and the bottom row shows their corresponding structural fidelity maps (e)-(h). All images are cropped for better visualization.

Table 5.1: Comparison of structural fidelity scores between initial and converged images

Image		Blank image	Linear mapping	Fattal [35]	Durand [12]	Reinhard [13]	Drago [4]
Kitchen	initial image	0.0099	0.7367	0.8748	0.7467	0.8073	0.7728
	converged image	0.8757	0.9737	0.9721	0.9758	0.9840	0.9796
Bristol	initial image	0.0099	0.8726	0.9020	0.7269	0.8013	0.7574
	converged image	0.8754	0.9812	0.9610	0.9493	0.9796	0.9747
Tinterna	initial image	0.0099	0.9572	0.9471	0.8797	0.9473	0.9307
	converged image	0.9052	0.9951	0.9786	0.9818	0.9958	0.9949
Memorial	initial image	0.0099	0.9573	0.9586	0.9559	0.9832	0.9765
	converged image	0.9184	0.9934	0.9863	0.9950	0.9959	0.9972

stops the iteration when the structural fidelity value between two consecutive iterations is smaller than the threshold. Throughout our experiment, we set $\lambda = 150$ and $T = 0.0001$ and empirically find that the behavior of the proposed algorithm is not sensitive to these parameters.

To test the performance of the proposed method, we select a set of widely-used HDR images (partially listed in Table 5.1) as test images. Similar results are obtained for other test images. The initial images for the iterative algorithm are created using different approaches, including blank images (all image pixels are set to 128), linearly mapped images (direct linear scaling from the source HDR images to the dynamic range of [0,255]), and images created by state-of-the-art TMOs [4, 12, 13, 35]. It can be observed in Table 5.1 that starting from simple blank and linearly-scaled initial images, the proposed method successfully produces images with high structural fidelity. In addition, it is also quite effective at improving upon all state-of-the-art TMOs.

Figure 5.1 provides a visual demonstration of the iterative procedure, where the proposed algorithm is applied to an initial tone mapped "Memorial" image created by the "Exposure and Gamma" method in Adobe Photoshop. The structural fidelity map is very effective at detecting the missing details in the tone mapped images. For example, the structural details in the brightest window region in the initial image Figure 5.1(a) are completely lost due to tone-mapping and are clearly indicated by the central dark region in the structural fidelity map Figure 5.1(e). With the progress of iterations, such details become more and more visible until nearly perfectly restored after 25 iterations, as shown in Figure 5.1(d). The evolution of the structural details is very well tracked by the structural fidelity maps computed along with the iterations, which eventually converge to a nearly uniform white picture.

To observe the behavior of the iterations numerically, we plot the structural fidelity measure as a function of iteration for two source images in Figures 5.2 and 5.3, respectively. It appears that the proposed iterative approach is well-behaved and always increases monotonically until it converges to a fixed point in the image space, although the fixed point may only be a local maximum. The structural fidelity values of the converged images suggest that the local maxima obtained from different initial images vary. This finding

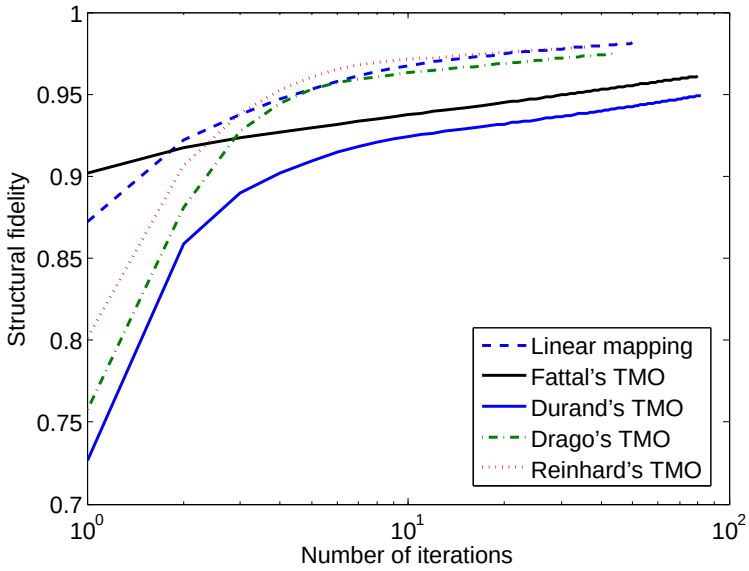


Figure 5.2: Structural fidelity versus iteration of tone mapped "Bristol" images with initial images created by different TMOs.

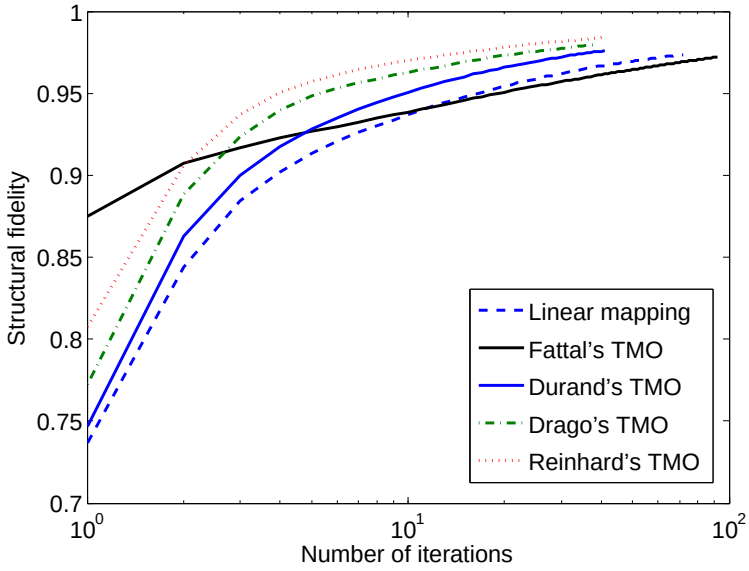


Figure 5.3: Structural fidelity versus iteration of tone mapped "Kitchen" images with initial images created by different TMOs.

indirectly reflects the complication of the search space.

To further demonstrate the performance of the proposed method, Figures 5.4, 5.5 and 5.6 show the results of optimization problems using three different tone-mapped images as initial points. Figure 5.4(a) is the tone-mapped "memorial" image using linear mapping, where the structural fidelity equals 0.88. The converged image is illustrated in Figure 5.4(b) and shows that the structural fidelity has been increased to 0.99. The restored details in the top window as well as on the floor prove the effectiveness of our approach. Figure 5.5 shows another example of our method, where the initial image is generated using one of the state-of-the-art TMOs described in [13]. The initial tone-mapped image looks visually plausible; however, the structural fidelity index is 0.80. Our optimization method results in an LDR image (Figure 5.5(b)) wherein missing structural details such as shadows on the wall and the structure in the ceiling become visible. The goal in our optimization problem is to maximize the structural fidelity measure, an approach that does not take luminance component into account. Therefore, starting from a completely blank image, the proposed method is able to restore the structural details in HDR images. Figure 5.6(a) shows a blank image where all pixel values are 125. The result of the proposed method is demonstrated in Figure 5.6(b), wherein most of the structural details (S = 0.87) are restored.

The computational complexity in each iteration of the proposed algorithm is linear with respect to the number of pixels in the image. Our unoptimized MATLAB implementation on an Intel Quad-Core 2.67GHz computer takes, on average, around 6 seconds per iteration for an image of size 512×512.

5.2 Tone-Mapping by Optimizing TMQI

Assuming TMQI to be the quality criterion of tone-mapped images, the problem of optimal tone-mapping can be formulated as

$$\mathbf{Y}_{\text{opt}} = \arg\max_{\mathbf{Y}} \text{TMQI}(\mathbf{X}, \mathbf{Y}) \,. \tag{5.23}$$

(a) (b)

Figure 5.4: (a) Initial tone-mapped image using linear mapping (S = 0.88), (b) Converged image (S = 0.97)

(a) (b)

Figure 5.5: (a) Initial tone-mapped image using Reinhard's method [13] (S = 0.80), (b) Converged image (S = 0.98).

(a) (b)

Figure 5.6: (a) Blank image as initial image (S = 0.00), (b) Converged image (S = 0.87).

It was shown in Chapter 3 that TMQI consists of two fundamental components, structural fidelity and statistical naturalness. In this section, we propose an iterative algorithm that starts from any initial image $\mathbf{Y}_0$ and searches for the best solution in the image space. Specifically, in each iteration, we first adopt a gradient ascent algorithm to improve the structural fidelity S, and then solve a constrained optimization problem to improve the statistical naturalness N. These two steps are applied alternately until convergence. Details of the algorithm are elaborated as follows.

In the k-th iteration, given the result image $\mathbf{Y}_k$ from the last iteration, a gradient ascent algorithm is first applied to improve the structural fidelity:

$$\hat{\mathbf{Y}}_k = \mathbf{Y}_k + \lambda \, \mathbf{G}_{\mathbf{Y}}|_{\mathbf{Y}=\mathbf{Y}_k} \,, \tag{5.24}$$

where $\mathbf{G}_{\mathbf{Y}} = \nabla_{\mathbf{Y}} S(\mathbf{X}, \mathbf{Y})$ is the gradient of $S(\mathbf{X}, \mathbf{Y})$ with respect to $\mathbf{Y}$ computed in section 5.1.1, and λ controls the updating speed.

After the structural fidelity update step of (5.24), we obtain an intermediate image $\hat{\mathbf{Y}}_k$, which will be further updated to $\mathbf{Y}_{k+1}$ such that the statistical naturalness is improved. This is done by a point-wise intensity transformation through a 3-segment equipartition monotonic piecewise linear function given by

$$y_{k+1}^i = \begin{cases} (3/L)a\hat{y}_k^i & 0 \le \hat{y}_k^i \le L/3 \\ (3/L)(b-a)\hat{y}_k^i + (2a-b) & L/3 < \hat{y}_k^i \le 2L/3 \\ (3/L)(L-b)\hat{y}_k^i + (3b-2L) & 2L/3 < \hat{y}_k^i \le L \end{cases} \tag{5.25}$$

where L is the dynamic range of the tone-mapped images, and the parameters a and b (where $0 \le a \le b \le L$) need to be selected so that the mapped image $\mathbf{Y}_{k+1} = \{y_{k+1}^i \text{ for all } i\}$ has an increased likelihood of mean μ_{k+1} and std σ_{k+1} values based on the statistical naturalness models P_m and P_d described in section 3.1.2. To solve for a and b, we first decide on the desired mean and std values by

$$\begin{aligned} \mu_{k+1}^d &= \hat{\mu}_k + \lambda_m(c_{P_m} - \hat{\mu}_k) \\ \sigma_{k+1}^d &= \hat{\sigma}_k + \lambda_d(c_{P_d} - \hat{\sigma}_k) \,, \end{aligned} \tag{5.26}$$

where $\hat{\mu}_k$ and $\hat{\sigma}_k$ are the global mean and std of $\hat{\mathbf{Y}}_k$, respectively; c_m and c_d are the values corresponding to the peaks in the P_m and P_d models, respectively; and λ_m and λ_d are step sizes that control the updating speed. Thus, finding the parameters a and b can be formulated as an optimization problem:

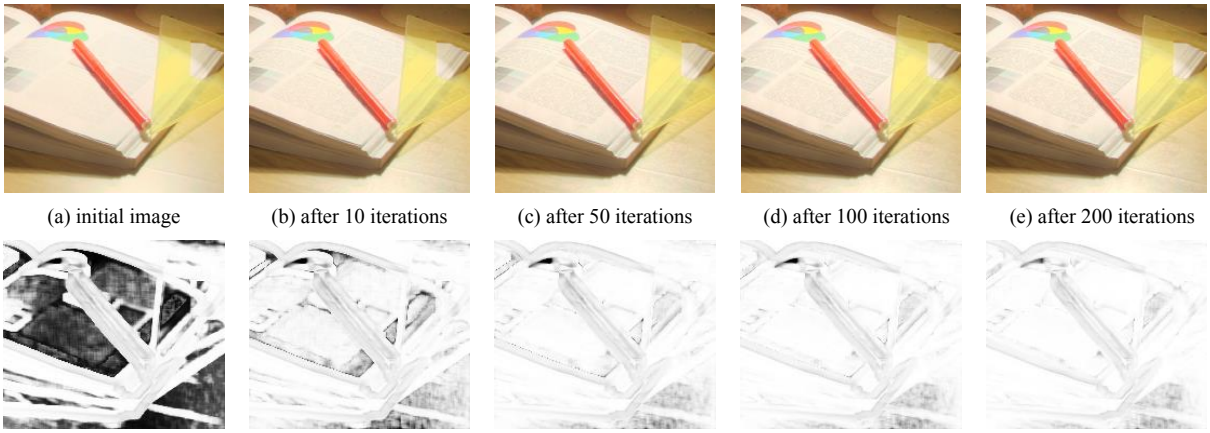
$$\{a, b\}_{opt} = \arg\min_{\{a,b\}} ||\mu_{k+1} - \mu^d_{k+1}||^2 + \eta||\sigma_{k+1} - \sigma^d_{k+1}||^2$$
$$\text{subject to } 0 \leq a \leq b \leq L\,, \tag{5.27}$$

where η controls the relative importance of the mean and std terms. In our implementation, the Matlab function *fmincon*, with its interior-point algorithm, is used to solve this optimization problem. Once the optimal values of a and b are obtained, they are plugged into (5.25) to create the output image $\mathbf{Y}_{k+1}$, which is subsequently employed as the input image in the $(k+1)$-th iteration.

The iteration continues until convergence, which is determined by checking the difference between the images of consecutive iterations. Specifically, when $||\mathbf{Y}_{k+1} - \mathbf{Y}_k|| < \epsilon$, the iteration stops. The proposed algorithm involves five parameters in total, which are set empirically to $\epsilon = 0.01$, $\lambda = 0.3$, $\lambda_m = \lambda_d = 2$ and $\eta = 1$ in all of our experiments.

5.2.1 Experimental Results

The proposed algorithm is tested on a database of 15 HDR images, which include various contents such as humans, landscapes, architectures, as well as indoor and night scenes. The effect of adopting structural fidelity measure as an optimization criterion was shown in Figures 5.1, 5.4, 5.5 and 5.6. Figure 5.7 also illustrates how structural fidelity is being improved in different iterations. It can be observed that the structural fidelity map is very effective at detecting the missing structures (e.g., text in the book region, and fine textures on the desk). By contrast, in Figure 5.8, the initial "building" image is created by a gamma correction mapping ($\gamma = 2.2$), and we apply the proposed iterative algorithm but using statistical naturalness updates only. With the iterations, the overall brightness and contrast of the image are significantly improved, leading to a more visually appealing



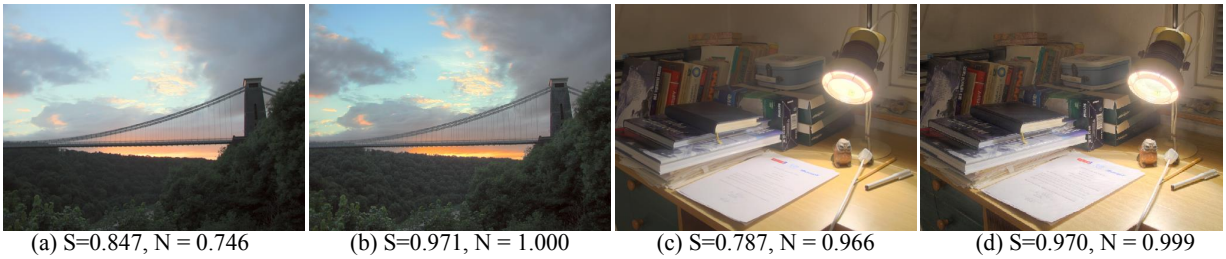
(a) initial image (b) after 10 iterations (c) after 50 iterations (d) after 100 iterations (e) after 200 iterations

(f) initial image, S = 0.689 (g) 10 iterations, S = 0.921 (h) 50 iterations, S = 0.954 (i) 100 iterations, S = 0.961 (j) 200 iterations, S = 0.966

Figure 5.7: Tone-mapped "desk" images and their structural fidelity maps. (a): initial image created by Reinhard's algorithm [13]; (b)-(e): images created using iterative structural fidelity update only; (f)-(j) corresponding structural fidelity maps of (a)-(e), where brighter indicates higher structural fidelity. All images are cropped for better visualization.

(a) initial image, N = 0.000 (b) 10 iterations, N = 0.001 (c) 50 iterations, N = 0.428 (d) 100 iterations, N = 0.868 (e) 200 iterations, N = 0.971

Figure 5.8: Tone-mapped "building" images. (a): initial image created by Gamma correction ($\gamma = 2.2$); (b)-(e): images created using iterative statistical naturalness update only.



(a) S=0.847, N = 0.746 (b) S=0.971, N = 1.000 (c) S=0.787, N = 0.966 (d) S=0.970, N = 0.999

Figure 5.9: Tone-mapped "bridge" and "lamp" images. (a) and (c): initial images created by Reinhard's algorithm [13]. (b) and (d): images after applying the proposed algorithm.

(a) S=0.148, N=0.000 (b) S=0.959, N=0.997 (c) S=0.521, N=0.038 (d) S=0.976, N=0.999

Figure 5.10: Tone-mapped "memorial" and "women" images. (a) and (c): initial images created by Gamma correction. (b) and (d): images after applying the proposed algorithm.

and natural-looking image.

The results shown in Figures. 5.9 and 5.10 are obtained by applying the proposed algorithm. In Figure 5.9, the initial images are created by Reinhard's TMO [13]. Although the initial image of Figure 5.9(a) presents a seemingly reasonable appearance, the fine details of the woods, the brick textures of the tower, and the details of the clouds are fuzzy or invisible. The proposed algorithm recovers these fine details and makes them much sharper, as can be seen in Figure 5.9(b). In addition, the overall appearance is more pleasant due to the statistical naturalness update. Similar results are also observed in Figure 5.9(d), where the details of the wall, scribbling papers and the drawer are well recovered. In Figure 5.10, the initial images are obtained by applying Gamma correction mapping ($\gamma = 2.2$), which creates dark images with missing details. Starting from these images, the proposed iterative algorithm successfully recovers most details in the images and presents a more realistic and pleasant appearance. It is worth mentioning that the proposed method often recovers image details that are unseen in the initial images; for example, the wall and door in the background are missing in Figure 5.10(c) but are clearly visible in Figure 5.10(d).

To further verify the effectiveness and consistency of the proposed algorithm, we conducted a subjective experiment. In particular, we select 15 HDR images that contain various natural scenes and adopt Gamma mapping and log-normal mapping methods to tone-map them to $15 \times 2 = 30$ LDR images. We then use them as initial images of the iterative algorithm and obtain 30 TMQI optimized images. Eventually, we obtain 15 sets of tone mapped images, each of which contains 4 images. 24 naive subjects (9 males and 15 females aged between 22 and 30) were asked to give an integer score between 0 and 10 for the perceptual quality of each tone mapped image, where 0 denotes the worst quality and 10 the best. The final quality score for each individual image is computed as the average of subjective scores, named mean opinion score (MOS), from all subjects. The results are listed in Table 5.2, from which we have several interesting observations. First, using TMQI as the optimization goal, the proposed algorithm leads to consistent perceptual gain for both types of initial images. Note that because the image space is extremely complicated and the proposed algorithm can only guarantee to find a local optimum, better initial

images often lead to better local optima, which correspondingly have better perceptual quality.

Table 5.2: Mean opinion scores of tone mapped images

Image set	Mean opinion scores (MOS)			
	Gamma mapping		log-normal mapping	
	Initial	TMQI optimized	Initial	TMQI optimized
1	1.00	4.71	5.58	4.13
2	1.54	4.92	2.13	4.33
3	0.25	4.50	1.88	3.79
4	3.33	4.63	3.33	4.96
5	0.54	3.88	4.58	4.25
6	0.58	3.29	5.21	3.17
7	1.29	5.67	2.38	5.33
8	1.54	3.79	3.29	3.83
9	0.96	4.38	2.67	3.79
10	6.50	5.21	2.54	5.13
11	0.46	4.83	4.00	2.79
12	3.63	5.21	3.46	3.00
13	5.33	3.04	2.67	3.29
14	2.17	2.75	3.58	2.83
15	4.67	5.13	1.92	3.88
Average	**2.25**	**4.39**	**3.28**	**3.90**

To have a close look at the iterative behavior of the proposed method, Figures 5.11 and 5.12 show the structural fidelity and statistical naturalness measures as functions of iteration using different initial images as the starting point. There are several observations. First, both measures increase monotonically with iterations. Second, the proposed algorithm converges in all cases regardless of whether a simple or sophisticated TMO result is used as the initial image. Third, different initial images may result in different converged

images. From these observations, we conclude that the proposed iterative algorithm is well behaved, but the search space is irregular in the high-dimensional space and contains many local optima, and the proposed algorithm may be trapped in one of the local optimum. This is the major limitation of the current approach that points to the direction for future improvement.

The computational complexity of the proposed algorithm increases linearly with the number of pixels in the image. Our unoptimized Matlab implementation takes around four seconds per iteration for a 341×512 image on an Intel Quad-Core 2.67 GHz computer.

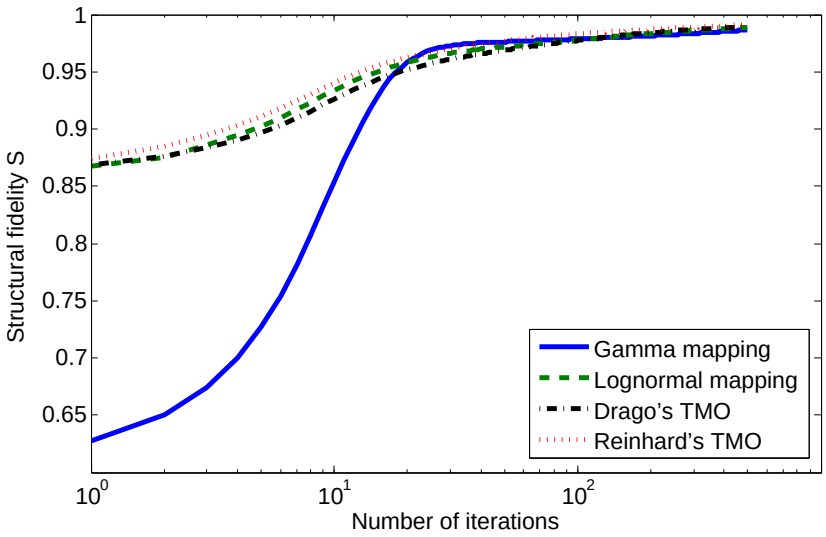


Figure 5.11: Structural fidelity as a function of iteration with initial images created by different TMOs.

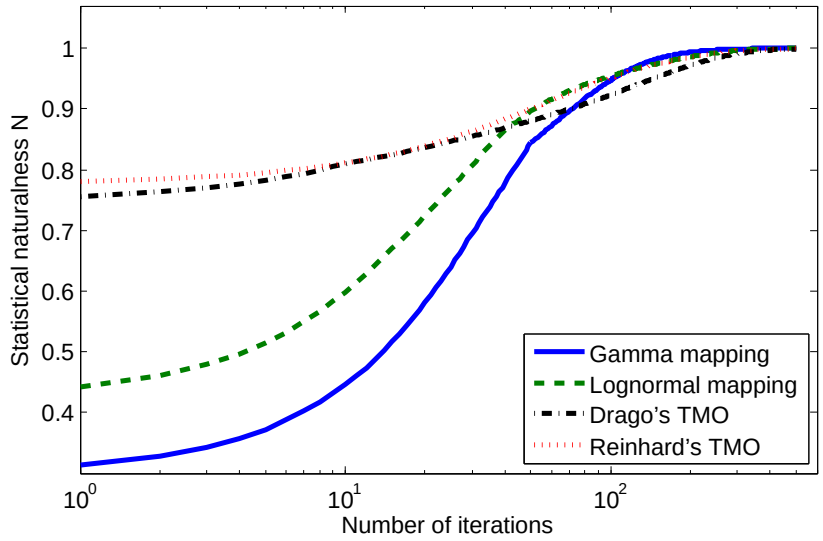


Figure 5.12: Statistical naturalness as a function of iteration with initial images created by different TMOs.

Chapter 6

Cross Spatial Resolution Image Quality Assessment

Image interpolation techniques that can create High-Resolution (HR) from Low-Resolution (LR) images are extensively used in real world applications. An increasing number of interpolation and image super-resolution (SR) algorithms have been proposed recently to create images with higher spatial resolution from low-resolution (LR) images, but how to evaluate the quality of interpolated images is not a well-resolved issue. Subjective assessment methods are useful and reliable, but are also slow, expensive, and difficult to incorporate into the design and optimization of algorithms and systems. In this work, we make one of the first attempts to develop an objective quality assessment method of a given resolution-enhanced image using the available LR image as a reference. We provide a generalized distortion measure that can be applicable to the case that the interpolation factor is an integer number. The proposed Interpolated Natural image Distortion (IND) and Weighted IND (WIND) measures, which incorporate frequency energy fall-off, dominant orientation statistics and spatial continuity statistics features, perform statistically equivalently or sometimes better than an average human subject. Moreover, we demonstrate the potential application of the proposed method in the parameter tuning of image interpolation algorithms.

The quality measures described in this chapter adopt a Natural Scene Statistics (NSS) framework, where image quality degradation is gauged by the deviation of the image's statistical features from NSS models trained upon high quality natural images. Recently, there has been a growing interest in using natural scene statistics (NSS) based approaches for RR and NR IQA [65]. The basic assumption behind NSS approaches is that high-quality images captured from the natural visual environment have strong low-level statistical regularities [71], to which the biological visual apparatus has adapted and evolved over the millennia [65]. Consequently, any departure from such regularities creates "perceptual unnaturalness", which is assumed to be directly related to perceived image quality. In the past decade, NSS based approaches have been successfully used in a number of RR and NR IQA algorithms [87, 88], though they have not been exploited in quality assessment of interpolated images.

6.1 Statistical Features

Given an LR image, an image interpolation algorithm increases the spatial resolution to create an HR image by predicting and inserting new pixels between the existing LR image pixels. In this work, we consider only the case of interpolations by integer-scaling factors, and there is no fractional-pixel shift in the LR image. As a result, the LR image can be viewed as a downsampled version of the HR image, where the pixel intensities remain unchanged at the sampling points. Let α be the interpolation scaling factor. Starting from the HR image, we can create α^2 downsampled sub-images, where one of them is the original LR image.

An illustration of this sub-image extraction process is shown in Figure 6.1, and a corresponding example of an interpolated image is given in Figure 6.2, where $\alpha = 3$ and all the pixels marked with "0" in Figure 6.1 constitute a sub-image that is exactly the same as the original LR image. All other sub-images are composed of pixels generated during the interpolation process. By close observation of the sub-images in Figure 6.2, one can discern the differences between the original LR and the other sub-images. The latter

often exhibits "unnaturally" smoothed edges and structural content. This observation has inspired us to design statistical features based on sub-image statistics and the ways such features differ between high-quality natural HR images and interpolated HR images. Since the LR image is available as one of the sub-images, the design of the statistical features can make use of such information. The first two statistical features employed in the proposed IQA algorithm are developed by following these ideas and are elaborated in the following sections.

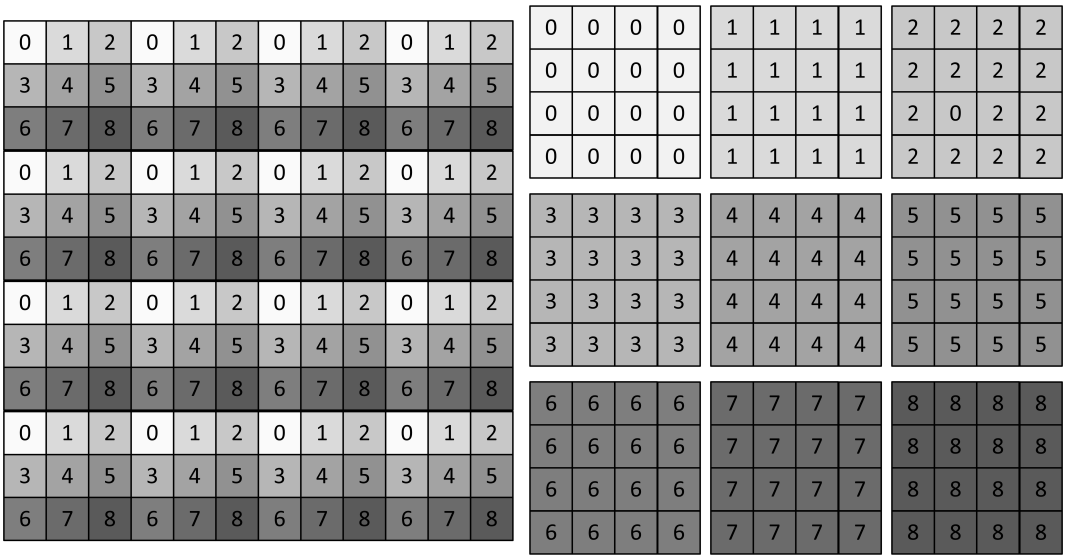


Figure 6.1: Illustration of sub-image extraction from an HR image.

6.1.1 Frequency Energy Fall-off Statistics

It has long been known that the amplitude spectrum of natural images falls with the spatial frequency approximately proportional to $1/f^p$ [89], where f is the spatial frequency and p is an image-dependent constant. This knowledge has helped us helps build a statistical model based on frequency energy fall-off. Specifically, we decompose the sub-images into dyadic scales using a steerable pyramid transform [90] and then compute the energy (the sum of squared transform coefficients) in each scale. For natural images, the energy in

Figure 6.2: An example of sub-images extracted from an interpolated HR image.

each scale, computed as the sum of squared transform coefficients, falls from the coarse to fine scales. Generally, the trends of scale or frequency energy fall-off of natural images are fairly similar [89], while unnatural blurry images tend to have steeper slopes.

Figure 6.3 shows the frequency energy fall-off curves drawn for different sub-images in Figure 6.2. Two useful observations can be made here: First, the fall-offs are approximately straight lines in log scale, which is consistent with the $1/f^p$ model; Second, the slope of energy fall-off reflects the blurriness of the images. In particular, with the increase of blurriness from Sub-image1 to Sub-image3, and to Sub-image9, the slopes of the fall-offs become steeper. The substantial difference in the speed of fall-off between different sub-images observed here is unlikely to happen in high-quality natural images, for which statistics on sub-images are presumably similar. This observation suggests that measuring the statistical differences between sub-image frequency energy fall-off could be a useful way to distinguish interpolated images from high-quality natural images. Specifically, let s_i denote the slope of frequency energy fall-off between the two finest scales in the i-th

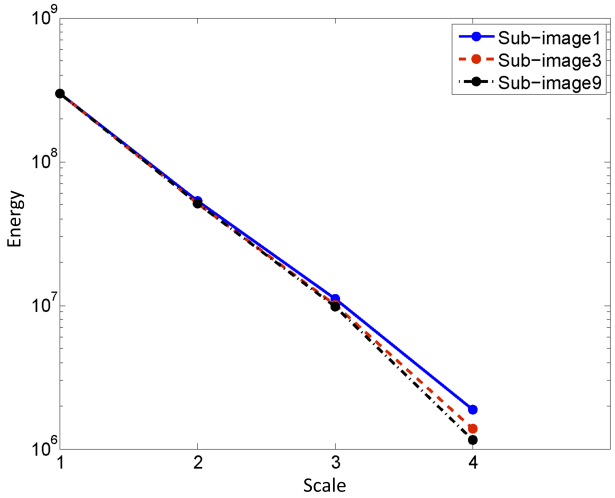


Figure 6.3: Frequency energy fall-offs of sub-images in Figure 6.2.

sub-image. We define a frequency energy fall-off feature as

$$e_f = \frac{\left[\frac{1}{\alpha^2-1}\sum_{k=2}^{\alpha^2}(s_k - s_1)^2\right]^{1/2}}{s_1}, \tag{6.1}$$

where s_1 is computed from the first sub-image (which is the same as the reference LR image) and used as a reference. The deviations of all other sub-images (for $k = 2, ..., \alpha^2$) are averaged and normalized by s_1. For a high-quality HR image, the deviation is expected to be small, leading to a small e_f feature (though unlikely to be zero, as will be shown in Section 6.1.4). By contrast, for interpolated images that create blurry sub-images (as exemplified by Figure 6.2), the e_f statistic is expected to be much larger.

6.1.2 Dominant Orientation Statistics

The frequency energy fall-off feature described above is based on global energy measures, while interpolation processes often result in distortions in local image structures. In the literature, image gradient is widely used to study local structural details, particularly on

edges and orientations [91–93]. Let $\mathbf{I}$ be an $N \times N$ image patch, and the gradient at pixel (x, y) in $\mathbf{I}$ be defined as $\nabla\mathbf{I}(x, y) = [\nabla_x\mathbf{I} \;\; \nabla_y\mathbf{I}]_{(x,y)}$, where ∇_x and ∇_y denote the derivatives in horizontal and vertical directions, respectively. By pooling the gradients of all pixels in the patch, we obtain a gradient matrix given by

$$\nabla\mathbf{I} = \begin{bmatrix} \vdots & \vdots \\ \nabla\mathbf{I}_x(x, y) & \nabla\mathbf{I}_y(x, y) \\ \vdots & \vdots \end{bmatrix}_{N^2\times 2} . \tag{6.2}$$

We follow the well-known compact singular value decomposition (SVD) approach [91] to estimate the dominant orientation in the patch, together with the energy along the dominant and its orthogonal directions. The compact SVD of $\nabla\mathbf{I}$ can be written as

$$\nabla\mathbf{I} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{U}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}[\mathbf{v}_1\, \mathbf{v}_2]^T , \tag{6.3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, where the column vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ indicate the dominant gradient orientation and its orthogonal direction (dominant edge orientation), respectively. The matrix $\mathbf{\Lambda}$ is a 2×2 diagonal matrix, where the singular values $\lambda_1 \geq \lambda_2 \geq 0$ provide energy measures along $\mathbf{v}_1$ and $\mathbf{v}_2$ directions, respectively. A simple energy-independent orientedness measure is given by [91]

$$C = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} . \tag{6.4}$$

If the image patch is fully oriented along one dominant direction, $\lambda_1 \gg \lambda_2 \approx 0$, and thus $C \approx 1$. With the decrease of the gap between λ_1 and λ_2, the C measure declines. At the other extreme, when the energy along the two orthogonal orientations are equally strong or when the image patch is very smooth with little energy, the value of C is close to 0. This orientedness measure has found successful applications in identifying local dominant orientations of textures [91] and has been extended to multi-scale settings [92].

We compute the orientedness measure C over an 11×11 sliding window that runs across

each sub-image. Let l_i be the mean of the C measure of the i-th sub-image. We then define our dominant orientation statistic features as

$$e_l = \frac{\left[\frac{1}{\alpha^2-1}\sum_{k=2}^{\alpha^2}(l_k - l_1)^2\right]^{1/2}}{l_1}\,. \tag{6.5}$$

As in (6.1), here l_1 is computed from Sub-image 1, which is the same as the reference LR image. The value of l_1 is used as a normalization factor to quantify the deviation of orientation strength of all other sub-images (for $k = 2, ..., \alpha^2$). For high-quality HR images, the deviations are expected to be small, resulting in small e_l measures, while interpolated images may generate much larger e_l, which will be shown in Section 6.1.4.

6.1.3 Spatial Continuity Statistics

Many interpolation algorithms create unnatural looking discontinuities in the spatial domain. This fact inspires us to study spatial continuity based statistical models and relate them to the naturalness of images. Let $f(i)$ for $i = 0, \cdots, N-1$ be one row (or column) of pixels extracted from the image, where N is the number of pixels in the row (or column). A straightforward method for examining signal continuity is to compute an absolute differencing signal given by

$$g(i) = |f(i+1) - f(i)| \quad \text{for} \quad 0 \le i \le N-2\,. \tag{6.6}$$

In the case of a high-quality HR natural image, the spatial continuity behavior is presumably close to uniform in a statistical sense, while such behavior in an interpolated image may vary in a certain pattern, depending on the interpolation grid. To capture such variations, for a given interpolation factor α, we first measure the average spatial continuity at every α pixel by

$$k_j = \frac{1}{M}\sum_{i=0}^{M-1} g(\alpha i + j) \quad \text{for} \quad 0 \le j \le \alpha - 1\,. \tag{6.7}$$

where $M = \lfloor (N-1)/\alpha \rfloor$. Doing so results in a length-α vector $\mathbf{k} = \{k_j | j = 0, \cdots, \alpha - 1\}$. We then use the ratio between the standard deviation and the mean of the $\mathbf{k}$ vector as a measure of spatial continuity variation:

$$e_s = \frac{\text{std}(\mathbf{k})}{\mu(\mathbf{k})}. \tag{6.8}$$

This measure is averaged across all rows and columns in the image, resulting in an overall spatial continuity variation feature.

6.1.4 Modeling Statistical Features

At the heart of our approach is the NSS framework, where NSS models obtained from collections of natural images are essential in establishing the base reference used to assess the naturalness and quality of the images being tested. Figures 6.4(a), 6.5(a), and 6.6(a) show the histograms of e_f, e_l and e_s features, respectively, obtained from 1000 high-quality original HR natural images. In addition, in Figures 6.4(b)-(i), 6.5(b)-(i), and 6.6(b)-(i), we show the corresponding histograms of interpolated HR images of scaling factor 2 generated using eight interpolation approaches: bilinear, bicubic, nearest neighbor, NEDI [9], directional filtering and data fusion (DFDF) [14], adaptive autoregression and soft-decision estimation (ARSD) [15], nonlocal autoregressive modeling (NARM) [16], and iterative curvature-based interpolation (ICBI) [17] algorithms. It can be observed that for high quality natural HR images, the histograms of all three features are concentrated near zero but do not peak exactly at zero. By contrast, different interpolation methods introduce different types and levels of changes in e_f, e_l and e_s features. As a result, the statistics of these features deviate from those of natural images. For examples, interpolation algorithms that tend to create overly smooth images (such as bilinear interpolation, bicubic interpolation, NEDI and DFDF) significantly expand the dynamic ranges of the e_f feature. On the other hand, the nearest neighbor method repeats the originally LR pixel values to create HR images, resulting in large peaks at zero in e_f and e_s features. The edge directed NEDI method significantly affects spatial continuity statistics, whereas the bicubic or bilinear interpola-

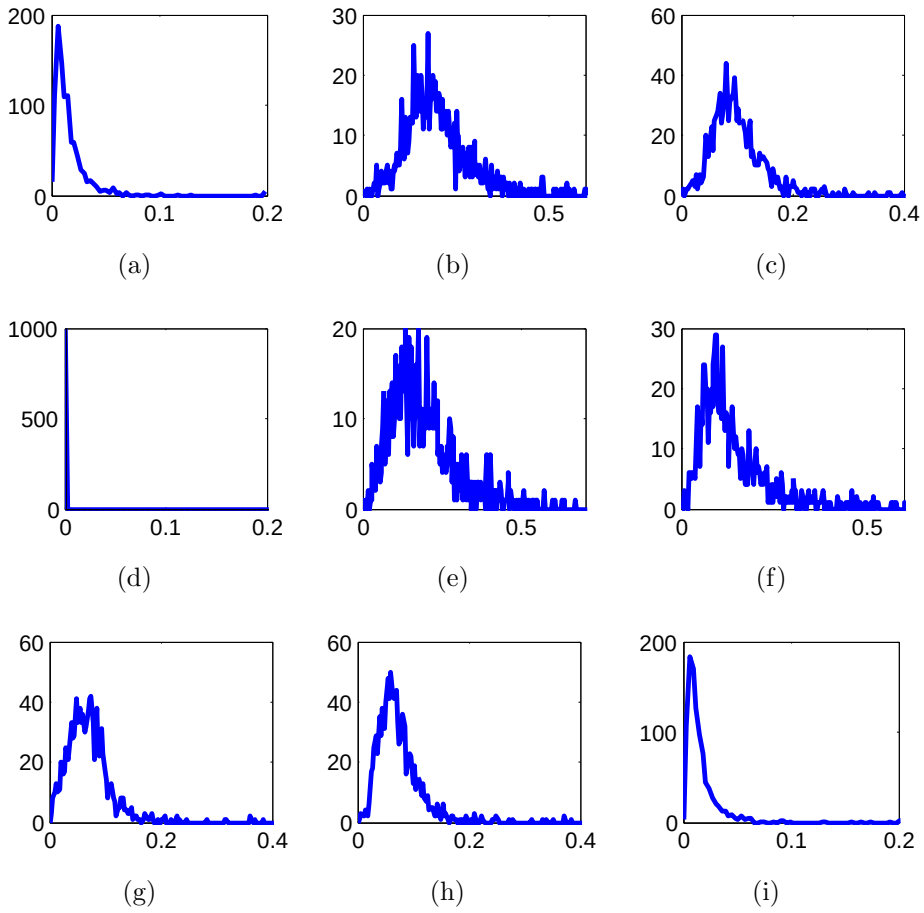


Figure 6.4: Histograms of e_f feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [9], (f) (DFDF) [14], (g) ARSD [15], (h) NARM [16] and (i) ICBI [17] algorithms.

tion methods may strengthen spatial continuity. It is also interesting to observe that the feature histograms created by the most- advanced algorithms, such as ICBI, exhibit the closest statistics to those of the natural images. All the above observations are intuitively sensible and demonstrate the potential usefulness of the proposed features, but in order to convert them to a quantitative image quality/distortion measure, we would first need to build probability density models of these features.

All three features are non-negative by definition. In our study, we find it useful to

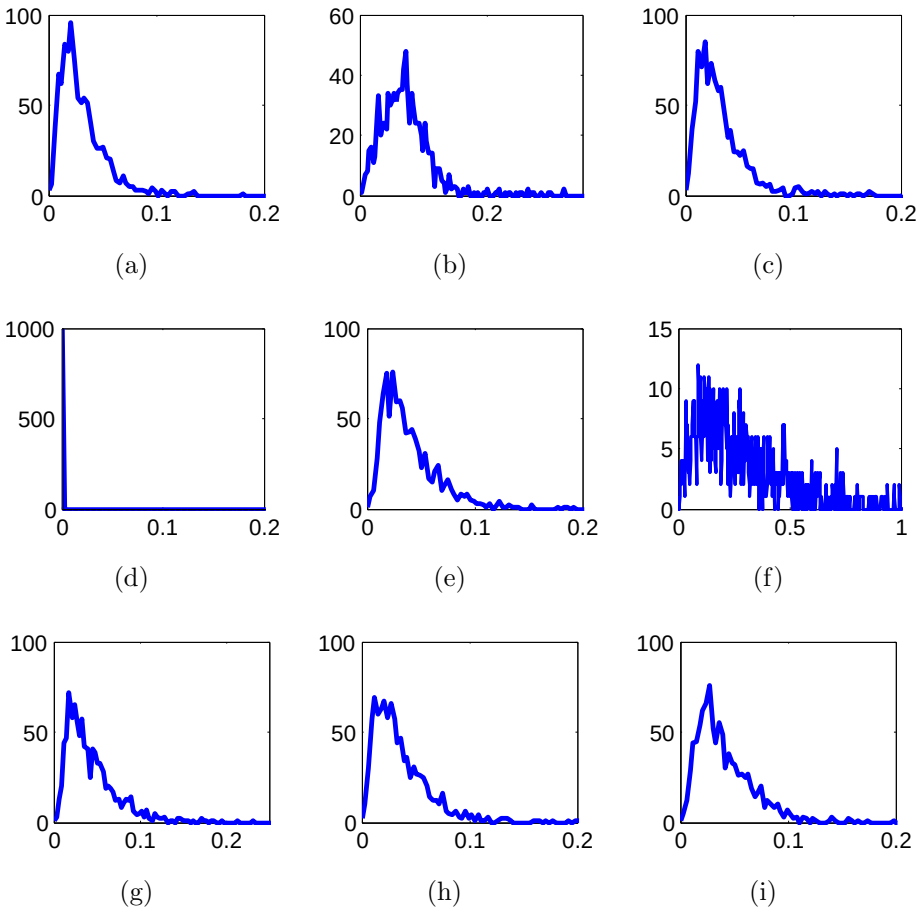


Figure 6.5: Histograms of e_l feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [9], (f) (DFDF) [14], (g) ARSD [15], (h) NARM [16] and (i) ICBI [17] algorithms.

observe them in the logarithm domain. Figures 6.7(a), 6.7(b) and 6.7(c), respectively, show the histograms of $\ln e_f$, $\ln e_l$ and $\ln e_s$ features for $\alpha = 2$ drawn from high-quality HR natural images. It is interesting to observe that all three histograms can be well fitted using Gaussian functions, which are also shown in the corresponding figures. This capability allows us to model the probability density of these features using simple models.

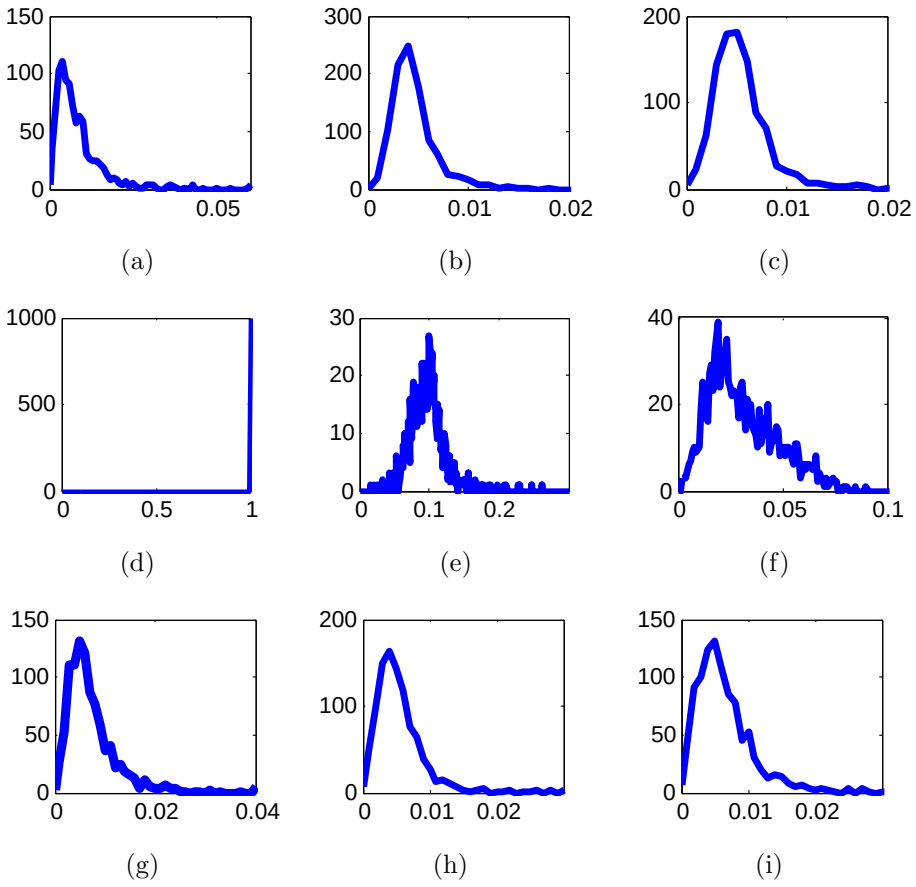
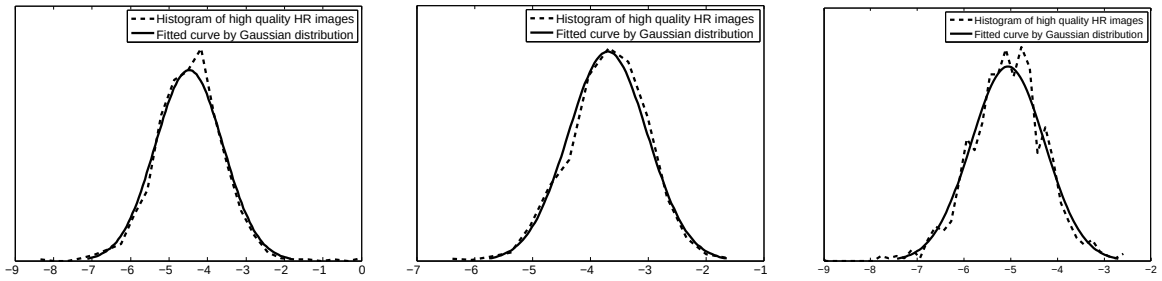


Figure 6.6: Histograms of e_s feature obtained from 1000 (a) HR natural images, and interpolated images using (b) bilinear, (c) bicubic, (d) nearest neighbor, (e) NEDI [9], (f) (DFDF) [14], (g) ARSD [15], (h) NARM [16] and (i) ICBI [17] algorithms.

In particular, for the e_f feature, we have

$$p_{e_f}(e_f) = \frac{1}{Z_f} \exp\left[-\left(\frac{\ln e_f - \mu_f}{\sqrt{2}\sigma_f}\right)^2\right], \tag{6.9}$$

where Z_f is a constant normalization factor that ensures that the density model integrates to 1. μ_f and σ_f are the logarithm domain mean and standard deviation parameters, respectively, for which the optimal values are obtained by maximal likelihood estimation.



(a) (b) (c)

Figure 6.7: Histograms of (a) $\ln e_f$, (b) $\ln e_l$ and (c) $\ln e_s$ features for $\alpha = 2$ drawn from original high-quality HR natural images, along with their corresponding Gaussian fitting functions.

We have found that the same probability density model provides good fittings to the feature histograms drawn from high-quality HR natural images, regardless of the values of α, although the optimal model parameters may change with α. By computing the optimal model parameters for different α values between 2 and 8, we find the following equations well summarize the optimal parameters as a function of α.

$$\begin{aligned} \mu_f &= -6.017\alpha^{-0.40}\,, \\ \sigma_f &= 0.72\,, \end{aligned} \tag{6.10}$$

Similarly, for the e_l feature, we obtain the following model for high-quality HR natural images

$$p_{e_l}(e_l) = \frac{1}{Z_l} \exp\left[-\left(\frac{\ln e_l - \mu_l}{\sqrt{2}\sigma_l} \right)^2 \right], \tag{6.11}$$

where Z_l is a normalization factor and the following equation provides good predictions of the maximal likelihood estimation results of the model parameters μ_l and σ_l as functions of α:

$$\begin{aligned} \mu_l &= -5.5\alpha^{-0.58}\,, \\ \sigma_l &= 0.62\,, \end{aligned} \tag{6.12}$$

For the e_s feature, the model for high-quality HR natural images is given by

$$p_{e_s}(e_s) = \frac{1}{Z_s} \exp\left[-\left(\frac{\ln e_s - \mu_s}{\sqrt{2}\sigma_s}\right)^2\right], \tag{6.13}$$

where Z_s is a normalization factor, and the maximum likelihood estimation of the model parameters μ_s, σ_s is summarized by

$$\begin{aligned} \mu_s &= -6.28\alpha^{-0.31}, \\ \sigma_s &= 1.1\alpha^{-2.2} + 0.53, \end{aligned} \tag{6.14}$$

Note that all the parameters introduced so far are determined by purely statistics of high-quality natural images, without involving distorted images or data from human subjective study.

6.2 Quality Assessment Model

We use the p_{e_f}, p_{e_l} and p_{e_s} models built upon statistics of natural images in Section 6.1.4 as the basis to assess the naturalness of test images. A high-quality natural image is expected to have larger values of p_{e_f}, p_{e_l} and p_{e_s} than distorted unnatural images. Assuming statistical independence between e_f, e_l and e_s features, one can define a naturalness measure based on a joint probability model:

$$p_n = \frac{1}{K} p_{e_f}(e_f) p_{e_l}(e_l) p_{e_s}(e_s), \tag{6.15}$$

where a normalization factor $K = 1/(Z_f Z_l Z_s)$ is added such that the value of p_n is upper-bounded by 1. In information theory, self-information or "surprisal" is often employed as a measure of the information content associated with the outcome of a random variable. We adopt this approach and convert the probability-based measure in (6.15) into an

interpolated natural image distortion (IND) measure given by

$$\text{IND} = -\ln p_n \,. \tag{6.16}$$

Plugging (6.9), (6.11), (6.13) and (6.15) into (6.16), we have

$$\begin{aligned} \text{IND} &= -\ln p_{e_f}(e_f) \times Z_f - \ln p_{e_l}(e_l) \times Z_l - \ln p_{e_s}(e_s) \times Z_s \\ &= \left(\frac{\ln e_f - \mu_f}{\sqrt{2}\sigma_f}\right)^2 + \left(\frac{\ln e_l - \mu_l}{\sqrt{2}\sigma_l}\right)^2 + \left(\frac{\ln e_s - \mu_s}{\sqrt{2}\sigma_s}\right)^2 \\ &\equiv D_f + D_l + D_s \,, \end{aligned} \tag{6.17}$$

where we define the first term, denoted by D_f, as the distortion of frequency energy fall-off feature, the second term, denoted by D_l, as the distortion of dominant orientation statistical feature, and the third term, denoted by D_s, as the distortion of spatial continuity feature.

Although IND provides a simple and parameter-free (no training using distorted images or subjective testing data is involved) measure for the statistical unnaturalness of the test images, it lacks the flexibility to account for the variations in perceptual annoyance to different types of distortions. A natural extension of this approach is to assign different importance to different distortion features by linearly weighting the three distortion components. This results in a weighted IND measure (WIND), given by

$$\text{WIND} = w_f D_f + w_l D_l + w_s D_s \,, \tag{6.18}$$

Without loss of generality, we fix $w_l = 1$, and the remaining weighting parameters w_f and w_s are determined based on subjective evaluation. The details are given in Section 6.3.

6.3 Validation

To validate the proposed quality model, we built a database of interpolated images and carried out subjective quality assessment experiments. The database contains thirteen high-quality natural HR source images, representing different types of structural content,

including indoor and outdoor scenes, humans, animals, natural scenes, and man-made architectures. All source images are 512×512 in image size. By directly downsampling the images by factors of 2, 4 and 8, we created 39 LR images with sizes of 256×256, 128×128 and 64×64, respectively. For each downsampled image, eight interpolation algorithms were employed to create interpolated HR images by scaling factors of 2, 4 and 8, respectively. The interpolation algorithms include classical and widely used bilinear, bicubic and nearest neighboor interpolation methods, as well as state-of-the-art algorithms such as NEDI [9], DFDF [14], ARSD [15], NARM [16], and ICBI [17]. Most of them interpolate an image by a scaling factor of 2, and were iteratively applied 2 and 3 times to achieve scaling factors of 4 and 8, respectively. Eventually, a total of 312 interpolated HR images were created. These images were then divided into 3 scaling factor levels and totally 39 image sets, each with 8 interpolated HR images.

Thirty subjects, 17 males and 13 females, aged between 20 and 30, participated in the experiments. The subjects were either naive or had only general knowledge about image processing, but no prior knowledge about the specific research work being carried out in this study. An HP ZR30w 30-inch monitor was used for the subjective tests, with a display spatial resolution of 2560×1600. This setting allows us to display a full set (out of the 39 sets) of 8 interpolated images, together with the original source HR image and the LR image, on the same screen. The viewing distance was adjusted to be approximately 32 pixels per degree of visual angle. A brief introduction and training session was conducted before the test. For each of the 39 image sets, the subjects were asked to use the source HR image and the LR image as references, and to score each of the 8 interpolated images shown on the screen with a quality scale between 1 and 10. After the subjective test, a statistical analysis was performed, and one subject was identified to be an outlier and the corresponding scores were removed. The remaining 29 subjective scores for each image were averaged to a mean opinion score (MOS). It is worth noting that we used an absolute scale rating approach in the experiment, as opposed to paired comparisons or a direct ranking approach. We find that for this particular experiment, this approach gives the most reliable results as compared to paired comparison method (which is slow and may cause transition problems) or a direct ranking approach (which often leads to large variations between

subjective evaluations when the quality of two or more images is close).

In the subsequent analysis, we treat the MOS value obtained for each image as the "ground truth", which is used to compare against any other quality prediction method. Spearman's rank-order correlation (SRCC) and Pearson's linear correlation coefficient (PLCC) after monotonic linear mapping are calculated to quantify the level of agreement between MOS and the quality prediction method being tested. Both evaluation criteria are upper-bounded by 1, which corresponds to perfect agreement. Higher values represent higher levels of agreement.

Before applying the SRCC and PLCC tests to assess objective quality models, we first evaluate how an average subject would perform in such tests. This is done by computing the SRCC and PLCC values between MOSs and the scores given by any particular individual subject. When this is done for all 29 subjects, we calculate the mean and standard deviation of the SRCC and PLCC values across all subjects. These average subject performance measures give useful baseline reference points on how an objective model behaves relative to a typical human subject. The average subject performance is provided in Tables 6.1, 6.2 and 6.3 for scaling factors 2, 4 and 8, respectively. In general, an average subject's SRCC and PLCC values are only moderately correlated with MOSs and are typically between the range of 0.5 and 0.8. This finding suggests that although subjects generally agree with each other on the quality of interpolated images, there exist significant variations between subject opinions. In the case of scaling factor 2, most interpolation algorithms perform quite well, making it difficult for the subjects to assign quality scores and to assess the relative quality of the interpolated images. This difficulty is reflected in the relatively low mean and high std of SRCC and PLCC values (Tables 6.1). With the increase of the scaling factor, the differences between the interpolation algorithms can be more easily discerned, making it a relatively easy task for the subjects. This leads to improved agreement between subjects, resulting in higher mean and std of SRCC and PLCC values (Tables 6.2 and 6.3).

Determining the optimal weighting parameters w_f and w_s in (6.18) based on subjective data is a straightforward linear regression problem. In addition to finding the optimal parameters, it is also important to test the robustness of this process. To do this, for each scaling factor, we employ a leave-one-out procedure where the 13 image sets are divided

Table 6.1: Performance evaluation on interpolated images with scaling factor 2

Image	SRCC						PLCC					
	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)
Lena	0.66/0.32	0.50	0.52	0.45	0.52	0.57	0.77/0.20	0.67	0.66	0.37	0.75	0.69
Butterfly	0.65/0.37	0.43	0.52	0.64	0.38	0.69	0.73/0.24	0.70	0.74	0.73	0.60	0.70
Barbara	0.45/0.34	0.43	0.47	0.23	0.38	0.38	0.63/0.27	0.50	0.50	0.35	0.48	0.51
Boat	0.61/0.28	0.74	0.85	0.30	0.33	0.40	0.67/0.27	0.72	0.86	0.54	0.63	0.63
Einstein1	0.35/0.31	0.80	0.70	0.35	-0.01	0.04	0.56/0.40	0.93	0.92	0.59	0.81	0.86
Statue	0.61/0.28	0.26	0.74	0.23	0.40	0.52	0.74/0.27	0.87	0.86	0.29	0.76	0.87
Lighthouse	0.34/0.34	0.55	0.59	0.38	0.20	0.24	0.44/0.37	0.77	0.89	0.61	0.69	0.69
Museum	0.54/0.32	0.47	0.43	0.59	0.01	0.14	0.66/0.25	0.82	0.82	0.76	0.73	0.75
Peppers	0.56/0.38	0.26	0.43	0.47	0.3	0.23	0.57/0.39	0.78	0.82	0.63	0.69	0.69
Goldhill	0.32/0.30	0.54	0.71	-0.38	0.21	0.21	0.52/0.42	0.95	0.95	0.17	0.92	0.94
Fruit	0.37/0.36	0.49	0.57	0.19	0.28	0.10	0.55/0.45	0.93	0.93	0.10	0.87	0.90
Baboon	0.31/0.32	0.54	0.70	0.63	0.21	0.16	0.47/0.32	0.77	0.77	0.62	0.77	0.76
Einstein2	0.57/0.26	0.28	0.69	-0.21	0.69	0.64	0.65/0.26	0.66	0.71	0.31	0.75	0.79
Average	**0.50/0.32**	**0.48**	**0.61**	**0.30**	**0.30**	**0.32**	**0.61/0.32**	**0.77**	**0.80**	**0.47**	**0.73**	**0.75**

Table 6.2: Performance evaluation on interpolated images with scaling factor 4

Image Set	SRCC						PLCC					
	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)
Lena	0.73/0.19	0.93	0.81	0.42	0.47	0.42	0.86/0.07	0.90	0.93	0.67	0.76	0.76
Butterfly	0.76/0.20	0.95	0.90	0.59	0.76	0.73	0.84/0.12	0.93	0.92	0.69	0.81	0.88
Barbara	0.71/0.20	0.47	0.59	0.23	0.16	0.10	0.82/0.13	0.70	0.83	0.58	0.70	0.70
Boat	0.58/0.35	0.19	0.57	0.16	0.38	0.23	0.72/0.20	0.67	0.71	0.65	0.67	0.67
Einstein1	0.66/0.25	0.64	0.43	0.19	0.61	0.66	0.79/0.14	0.63	0.63	-0.02	0.71	0.73
Statue	0.64/0.18	0.76	0.71	0.92	0.38	0.38	0.84/0.10	0.92	0.93	0.80	0.88	0.87
Lighthouse	0.51/0.31	0.62	0.78	0.19	-0.02	-0.04	0.68/0.25	0.66	0.73	0.08	0.66	0.66
Museum	0.51/0.30	0.62	0.69	0.30	0.40	0.45	0.64/0.25	0.45	0.63	0.57	0.59	0.70
Peppers	0.75/0.17	0.57	0.83	0.23	0.16	0.02	0.79/0.15	0.74	0.90	0.46	0.65	0.65
Goldhill	0.79/0.16	0.54	0.64	-0.33	0.88	0.90	0.83/0.13	0.67	0.60	0.35	0.80	0.87
Fruit	0.70/0.22	0.64	0.36	0.42	0.35	0.33	0.76/0.20	0.67	0.73	-0.15	0.73	0.73
Baboon	0.64/0.30	0.83	0.78	0.52	-0.16	-0.11	0.71/0.28	0.85	0.84	0.82	0.60	0.60
Einstein2	0.77/0.17	0.74	0.88	0.66	0.11	0.30	0.83/0.15	0.77	0.91	0.80	0.51	0.51
Average	**0.67/0.23**	**0.65**	**0.69**	**0.35**	**0.34**	**0.33**	**0.78/0.17**	**0.73**	**0.79**	**0.48**	**0.70**	**0.72**

Table 6.3: Performance evaluation on interpolated images with scaling factor 8

Image Set	SRCC						PLCC					
	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)	mean/std subject performance	IND	WIND	NIQE (NR)	SSIM (FR)	PSNR (FR)
Lena	0.77/0.19	0.30	0.71	0.07	0.61	0.85	0.82/0.16	0.52	0.75	0.59	0.78	0.90
Butterfly	0.74/0.22	0.77	0.73	0.74	0.62	0.34	0.79/0.18	0.56	0.86	0.77	0.78	0.67
Barbara	0.63/0.25	0.24	0.33	-0.23	0.42	0.92	0.75/0.28	0.80	0.79	0.75	0.85	0.87
Boat	0.71/0.24	0.35	0.38	0.40	0.59	0.33	0.74/0.24	0.75	0.70	0.53	0.52	0.51
Einstein1	0.73/0.26	0.0	0.12	0.40	0.42	0.47	0.80/0.15	0.56	0.60	0.66	0.81	0.81
Statue	0.61/0.28	0.40	0.50	0.19	0.33	0.35	0.74/0.18	0.72	0.70	0.40	0.72	0.72
Lighthouse	0.66/0.29	0.38	0.74	0.64	-0.26	-0.16	0.76/0.29	0.43	0.53	0.46	0.46	0.46
Museum	0.69/0.20	0.19	0.38	-0.09	0.83	0.78	0.78/0.20	0.70	0.73	0.54	0.72	0.75
Peppers	0.58/0.24	0.65	0.62	0.68	0.37	0.26	0.69/0.20	0.68	0.78	0.64	0.64	0.64
Goldhill	0.73/0.24	0.38	0.55	-0.01	0.88	0.70	0.76/0.24	0.66	0.88	0.46	0.87	0.82
Fruit	0.72/0.19	0.35	0.31	0.04	0.38	0.54	0.79/0.18	0.40	0.63	0.37	0.81	0.60
Baboon	0.62/0.30	0.50	0.70	0.49	0.04	-0.29	0.69/0.28	0.65	0.82	0.38	0.38	0.38
Einstein2	0.73/0.23	0.35	0.50	-0.19	0.54	0.45	0.82/0.22	0.66	0.78	0.50	0.55	0.50
Average	**0.69/0.24**	**0.37**	**0.50**	**0.24**	**0.44**	**0.43**	**0.76/0.22**	**0.62**	**0.73**	**0.54**	**0.68**	**0.66**

into 12 training sets and 1 testing set. The weights are then obtained by linear regression using the training sets and then tested on the testing set. The same process is repeated 13 times, each with a different division between training and testing sets. The mean and std of the w_f and w_s values obtained for scaling factors 2, 4 and 8, each obtained from 13 trials, are given in Table 6.4, together with the corresponding mean SRCC and PLCC performance test results over 13 trails. It appears that the weights obtained in all 13 trails are fairly close to each other, with low std values. In addition, the mean SRCC and PLCC results of the 13 leave-one-out trials are also close to those obtained using a fixed set of the average weighting parameters applied to all 13 image sets (which are shown in the WIND performance in the last rows in Tables 6.1-6.3). All these results suggest that the weights obtained through this procedure are robust. The mean w_f and w_s values for each scaling factor α given in Table 6.4 are the final values in all the subsequent tests.

In Table 6.4, we also observe that as the scaling factor increases, the impact of frequency energy fall-off and dominant orientation statistics increases (as compared to spatial continuity statistics). This is not surprising, because with larger scaling factors, it becomes more difficult for an interpolation algorithm to maintain the original energy distributions at high frequencies as well as the orientations in local image structures. These are also the major factors that affect the perceptual quality of the interpolated images.

Table 6.4: Weighting factors w_f and w_s, along with SRCC and PLCC performance, obtained from leave-one-out test for different scaling factor α

α	mean/std of w_f	mean/std of w_s	SRCC	PLCC
2	1.17/0.07	0.09/0.02	0.612	0.802
4	1.26/0.04	0.16/0.01	0.700	0.800
8	3.20/0.10	0.40/0.05	0.511	0.732

Table 6.4 provides only the weighting parameters for scaling factors of 2, 4 and 8 only. A natural way to extend the weight selection to other integer scaling factors is to interpolate

along the scaling factor axis. We found the following functions are useful for this purpose:

$$w_f = 0.0002\alpha^{4.43} + 1.16,, \tag{6.19}$$

$$w_s = 0.008\alpha^{1.7} + 0.06\,, \tag{6.20}$$

To the best of our knowledge, no other objective IQA method is directly applicable to the scenario we are interested in (where an LR image is used as a reference to assess the quality of an HR image). Therefore, in addition to comparing the proposed IND and WIND measures with an average human subject (as described earlier), we compare them with well-known FR-IQA measures, including PSNR and SSIM [20]. We have also included a state-of-the-art NR-IQA method named natural image quality evaluator (NIQE) [94] in the comparison. It has shown promising performance when tested using a number of widely-used image databases [94], but it has never been tested on interpolated images. The performance evaluation results are shown in Tables 6.1, 6.2 and 6.3 for the cases of scaling factors 2, 4 and 8, respectively. It can be observed that for almost all image sets, the SRCC and PLCC values of the proposed IND and WIND measures are well within the range of ± 1 standard deviation from the SRCC and PLCC values of average human subjects. This indicates that the proposed methods behave quite similarly to an average subject. Unsurprisingly, between IND and WIND, WIND performs consistently better and often outperforms an average human subject. The FR SSIM and PSNR methods are general-purpose approaches, without the need for any prior knowledge or specific considerations of image interpolation application; thus, they achieve only moderate correlations with subjective evaluations and are inferior to the proposed methods, which use novel features particularly designed to capture distortions in interpolated images. Similarly, the general-purpose NR NIQE method does not take specific consideration of image interpolation, and does not deliver competitive performance.

6.4 Application: Parameter Tuning in Interpolation Algorithms

The application domain of objective IQA measures is beyond comparing interpolated images and selecting the best interpolation algorithms. Many interpolation algorithms contain one or more parameters. Different selections of these parameters may result in interpolated HR images of very different perceptual quality, and the optimal parameters are often image-dependant. Without human interference, it is challenging to choose these parameters. An objective quality measure provides a useful tool to pick these parameters automatically. To demonstrate this fact, here we use the ICBI algorithm [17] as an example.

The ICBI method is a state-of-the-art "edge-directed" interpolation algorithm that upscales the image by keeping the original pixels in an enlarged grid and then estimates the missing pixels. The estimation is done using weighted averaging of neighboring pixels, with the weights determined based on local edge analysis. The ICBI algorithm uses some initial values for missing pixels, and then tries to define an energy term for each interpolated pixel. The ultimate goal is to minimize the energy term by making small changes in the second order derivatives. In [17], the energy term is a weighted sum of three components: curvature continuity U_c, curvature enhancement U_e and isophote smoothing U_i. The energy of each interpolated pixel at (i, j) is given by

$$U(i,j) = w_c U_c(i,j) + w_e U_e(i,j) + w_i U_i(i,j)\,, \tag{6.21}$$

According to [17], the U_c term is effective in removing artifacts but creates blurry images while the U_e term helps produce sharper edges. As a result, the relative values between w_c and w_e determine a tradeoff between edge sharpness and artifact removal. It was found that the U_i component adds only a slight improvement to perceived image quality, and thus w_i has relatively little influence on the performance of ICBI [17].

Since the perceptual quality of the interpolated image varies significantly with w_c and w_e, these parameters are typically chosen with trial and error [17]. To visualize this process, in Figure 6.8, we plot the WIND measure as a 2D function of w_c and w_e for a test image

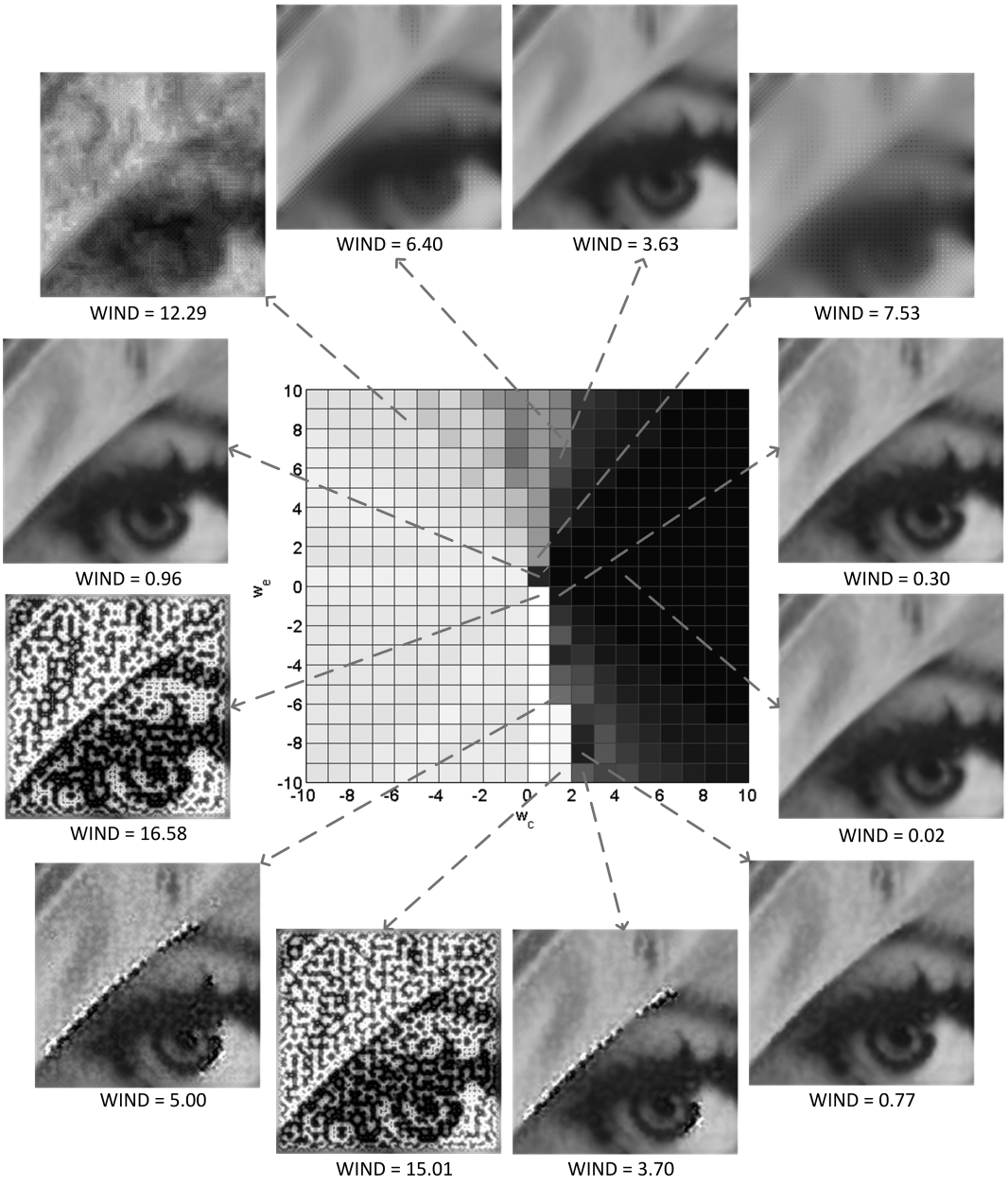


Figure 6.8: ICBI [17] interpolated images over a wide range of w_c and w_e selections. Darker shade indicates lower WIND value or higher image quality.

interpolated by ICBI at scaling factor 4, where a darker shade in the plot indicates a lower value of WIND or higher quality of the interpolated image. Sample interpolated images corresponding to different options of (w_c, w_e) values are also given. It is worth noting that the visual quality as a function of (w_c, w_e) is not smooth everywhere. Indeed, it can be quite sensitive, such that in certain places, small parameter changes can lead to dramatic variations in perceived quality of the interpolated images, making it difficult to manually decide on the right parameters to use. Careful inspection and comparisons of the interpolated images as well as their corresponding WIND values suggest that WIND is a good perceptual quality indicator and provides a useful tool for automatically choose the best values of (w_c, w_e).

Chapter 7

Conclusion and Future Work

This thesis tackles several emerging challenges in the field of visual signal quality assessment and enhancement. The contributions are on two general topics. We have developed an objective measure to assess the perceptual quality of two images with different dynamic range, and have used the objective quality measures to enhance tone-mapping operators. We have also presented a novel distortion measure to compare the visual quality of two images with different spatial resolutions and demonstrated its potential applications. In this chapter, we will summarize the major contributions of the thesis. Then, we will discuss different areas for future research. Related publications are listed at the end of the chapter.

7.1 Conclusion

In Chapter 3, we have developed an objective model to assess the quality of tone-mapped images by combining a multi-scale structural fidelity measure and a statistical naturalness measure. The measure not only provides an overall quality score of an image, but also creates multi-scale quality maps that reflect the structural fidelity variations across scale and space. In addition, we conducted a subjective test and collected subjective data to evaluate the proposed quality measure. Our experiments show that TMQI correlates reasonably well with subjective evaluations of image quality. Moreover, we have demonstrated

the usefulness of TMQI in automatic parameter tuning of tone-mapping algorithms and in fusing multiple tone-mapped images.

Chapter 4 demonstrates the usefulness of objective quality measures in the field of medical imaging. We have shown that commonly used linear functions for displaying HDR medical images on standard LDR displays are unable to map structural information accurately, and have proposed an optimization framework to construct adaptive windowing functions that enhance contrast. The optimization task seeks to maximize the structural fidelity measure in the tone-mapped LDR image using the HDR medical image as a reference. Our experiments have demonstrated very promising results.

A completely new paradigm in the field of tone-mapping operators was presented in Chapter 5 where designing the tone-mapping operators was formulated as an optimization problem in the space of images. The chapter described two main frameworks. First, a gradient ascent algorithm was proposed to maximize the quality of tone-mapped images in terms of structural fidelity measure. Second, we developed a novel approach for designing TMOs by navigating in the space of images to find the optimal image in terms of TMQI. The navigation is based on an iterative approach that alternates between improving the structural fidelity preservation and enhancing the statistical naturalness of the image. Experimental results show that both frameworks improve the overall quality of tone-mapped images significantly. Our experiments also showed that the proposed method is well behaved, and effectively enhances the image quality from a wide variety of initial images, including those created from state-of-the-art TMOs.

In Chapter 6, we designed an NSS-based objective model to automatically assess the quality of interpolated natural HR images using LR images as references. Three statistical features were employed in the proposed approach, including sub-image frequency energy fall-off statistics, sub-image local dominant orientation statistics and spatial continuity statistics. Statistical models were established using statistics based on high-quality natural HR images, and departures from such statistics were used as the key indicators of perceptual unnaturalness, which is assumed to be closely connected to perceived image quality. We built an image database of interpolated natural images and carried out subjective tests. Our experiments showed that the proposed quality measure agrees well with

the mean subjective opinions of interpolated image quality and often outperforms an average human subject. Furthermore, we demonstrated the extended potential applications of the proposed measure by applying it to automatic parameter tuning of the state-of-the-art ICBI interpolation algorithm.

7.2 Future Work

The research work described in this thesis represents initial attempts in several promising directions, and can be further improved in many aspects. Potential directions to continue this research work are summarized as follows.

The proposed TMQI method has several limitations that may be resolved or improved in the future. First, TMQI is designed to evaluate grayscale images only, but most HDR images of natural scenes are captured in color. One simple method to evaluate tone mapped color images is to apply the TMQI to each color channel independently and then combine them. Color fidelity and color naturalness measures may be developed to improve the quality measure. Second, simple averaging is used in the current pooling method of the structural fidelity map. Advanced pooling method that incorporate visual attention models may be employed to improve the quality prediction performance. Third, the current statistical naturalness measure is based on intensity statistics only. There is a rich literature on natural image statistics and advanced statistical models (that reflects the structural regularities in space, scale and orientation in natural images) may be included to improve the statistical naturalness measure. Fourth, using TMQI as a new optimization goal, many existing TMOs may be redesigned to achieve better image quality. Novel TMOs may also be developed by taking advantage of the construction of the proposed quality assessment approach. Finally, the current method is applied and tested using natural images only. The application scope of HDR images and TMOs is beyond natural images. For example, modern medical imaging devices often capture HDR medical images that need to be tone-mapped before visualization. The TMQI and optimization methods may be adapted to these extended applications.

Exploiting structural fidelity measure for the purpose of adaptive windowing of medical images provided promising results. However, there are some areas to be further explored in the future. Specifically, other families of continuous and monotonically increasing functions should be examined. Clinically reliable subjective tests by radiologists should be carried out to evaluate and calibrate the structural fidelity measure.

The current work on exploring optimum TMOs opens the door to a new class of TMO approaches. Many topics are worth further investigations. First, as is the case for any algorithm operating in complex high-dimensional space, the current approach only finds local optima. Careful studies on the search space are desirable. Second, the current implementation is computationally costly and requires a large number of iterations to converge. Fast search algorithms are necessary to accelerate the iterations. Third, the current statistical naturalness model is rather crude. Incorporating advanced models of image naturalness into the proposed framework has great potentials in creating more natural-looking tone-mapped images.

Our works on comparing two images with different spatial resolution can be further improved in the future. First, the current method is applicable to the case of interpolations by integer factors only. In practice, users may enlarge an image by a non-integer factor or scale it down to a fractional size. Not all the statistical features used in the current algorithm can be directly applied and more feature extraction and statistical modeling work is necessary. Second, since the current model are built upon natural scene statistics, it may not properly generalize to the case of artificial or graphical images. How to develop new meaningful features for these images is a topic worth further investigating. Finally, many recent image super-resolution algorithms take one or multiple LR images as the input to create HR images, where the positions of the LR image pixels may be shifted by fractional factors from the integer pixel grid. This poses new challenges to IQA research and opens up new space for future exploration.

Publications

Journal Papers

Kede Ma, **Hojatollah Yeganeh**, Kai Zeng and Zhou Wang, "High dynamic range image compression by optimizing tone-mapped image quality index (TMQI)", *Submitted to IEEE Transactions on Image Processing, June 2014.*

Hojatollah Yeganeh, Mohammad Rostami and Zhou Wang, "Objective quality assessment of interpolated natural images", *Submitted to IEEE Transactions on Image Processing, September 2013.*

Hojatollah Yeganeh, and Zhou Wang. "Objective quality assessment of tone-mapped images." *IEEE Transactions on Image Processing, Vol. 22, No. 2, February 2013.*

Conference Papers

Kede Ma, **Hojatollah Yeganeh**, Kai Zeng and Zhou Wang, "High dynamic range image tone mapping by maximizing tone mapped image quality index", *IEEE International Conference on Multimedia and Expo (ICME), China 2014.*

Nima Nikvand, **Hojatollah Yeganeh**, and Zhou Wang, "Adaptive windowing for optimal visualization of medical images based on normalized information distance", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Italy 2014.*

Hojatollah Yeganeh, Zhou Wang, "High dynamic range image tone mapping by maximizing a structural fidelity measure", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Canada 2013.*

Hojatollah Yeganeh, and Zhou Wang, Edward R. Vrscay, "Adaptive windowing for optimal visualization of medical images based on a structural fidelity measure", *International Conference on Image Analysis and Recognition (ICIAR), Portgal 2012.*

Hojatollah Yeganeh, Mohammad Rostami and Zhou Wang, "Objective quality assessment for image super-resolution: a natural scene statistics approach", *IEEE International Conference on Image Processing (ICIP), USA 2012.*

Hojatollah Yeganeh, Zhou Wang, "Structural fidelity vs. naturalness - objective assessment of tone-mapped images", *International Conference on Image Analysis and Recognition (ICIAR), Canada 2011.*

Hojatollah Yeganeh, Zhou Wang, "Objective assessment of tone-mapping algorithms", *IEEE International Conference on Image Processing (ICIP), China 2010.*

References

[1] M. Song, D. Tao, C. Chen, J. Bu, J. Luo, and C. Zhang, "Probabilistic exposure fusion," *IEEE Transactions on Image Processing*, vol. 21, pp. 341–357, Jan. 2012.

[2] M. Čadík, M. Wimmer, L. Neumann, and A. Artusi, "Image attributes and quality for evaluation of tone mapping operators," in *Proceedings of the 14th Pacific Conference on Computer Graphics and Applications*, (Taipei, Taiwan), pp. 35– 44, National Taiwan University Press, 2006.

[3] M. Cadik *et al.*, "Evaluation of tone mapping operators," http://www.cgg.cvut.cz/members/cadikm/tmo.

[4] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, vol. 22, no. 3, pp. 419–426, 2003.

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

[6] J. E. Dowling, *An Approachable Part of the Brain*. Cambridge: Belknap Press, 1987.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Mean squared error: love it or leave it? - a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, pp. 98–117, Jan. 2009.

[8] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H. Seidel, "Dynamic range independent image quality assessment," in *SIGGRAPH'08: International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, 2008.

[9] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Trans. Image Process.*, vol. 10, p. 15211527, 2001.

[10] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. IT-20, no. 4, pp. 525–536, 1974.

[11] R. Mantiuk, K. Myszkowski, and H. . Seidel, "A perceptual framework for contrast processing of high dynamic range images," in *Proceedings - APGV 2005: 2nd Symposium on Applied Perception in Graphics and Visualization*, pp. 87–94, 2005.

[12] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM Transactions on Graphics*, vol. 21, pp. 257–266, 2002.

[13] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *in Proc. of 29th annual Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, vol. 21, pp. 267–276, 2002.

[14] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2226–2238, 2006.

[15] X. Zhang and X. Wu, "Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation," *IEEE Transactions on Image Processing*, vol. 17, no. 6, pp. 887–896, 2008.

[16] W. Dong, L. Zhang, R. Lukac, and G. Shi, "Sparse representation based image interpolation with nonlocal autoregressive modeling," *IEEE Transactions on Image Processing*, vol. 22, no. 4, 2013.

[17] A. Giachetti and N. Asuni, "Real time artifact-free image upscaling," *Image Processing, IEEE Transactions on*, vol. 20, pp. 2760–2768, October 2011.

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

[19] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, pp. 35–44, 2004.

[21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *in Proc. of 37th Asilomar Conf. Signals, Systems and Computers*, 2003.

[22] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced high dynamic range imaging: theory and practice*. CRC Press, 2011.

[23] T. M. Lillesand, R. W. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation*. New York: John Wiley & Sons, 1994.

[24] J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Mller, "Practical hdr texture compression," *Computer Graphics Forum*, vol. 27, no. 6, pp. 1664–1676, 2008.

[25] J. Tumblin and H. Rushmeier, "Tone reproduction for realistic images," *IEEE Trans. on Computer Graphics and Applications*, vol. 13, no. 6, pp. 42–48, 1993.

[26] C. Schlick, *Quantization Techniques for the Visualization of High Dynamic Range Pictures*. New York: Springer-Verlag, 1994.

[27] J. Tumblin, J. K. Hdgins, and B. K. Guenter, "Two methods for display of high contrast images," *ACM Transactions on Graphics*, vol. 18, no. 1, pp. 56–94, 1999.

[28] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pp. 47–54, 2000.

[29] S. Pattanaik and H. Yee, "Adaptive gain control for high dynamic range image display," in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pp. 83–87, 2002.

[30] E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 13–24, 2005.

[31] G. Ward, *A Contrast-Based Scalefactor for Luminance Display*. Boston: Academic Press, 1994.

[32] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pp. 249–258, ACM, 1996.

[33] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.

[34] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *Proceedings - Pacific Conference on Computer Graphics and Applications*, pp. 382–390, 2007.

[35] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02*, pp. 249–256, 2002.

[36] Z. Wang and A. C. Bovik., *Modern Image Quality Assessment*. Morgan & Claypool Publisher.

[37] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–99, 2012.

[38] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.

[39] F. Drago, W. L. Martens, K. Myszkowski, and H. P. Seidel, "Perceptual evaluation of tone mapping operators," *In Proc. Of the SIGGRAPH Conf. Sketches and Applications*, 2003.

[40] J. Kuang, H. Yamaguchi, G. M. Johnson, and M. D. Fairchild, "Testing hdr image rendering algorithms," in *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, pp. 315–320, 2004.

[41] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," in *ACM Transactions on Graphics*, vol. 24, pp. 640–648, 2005.

[42] A. Yoshida, V. Blanz, K. Myszkowski, and H. Seidel, "Perceptual evaluation of tone mapping operators with real-world scenes," *SPIE Human Vision and Electronic Imagin*, vol. 5666, pp. 192–203, 2005.

[43] M. Barkowsky and P. L. Callet, "On the perceptual similarity of realistic looking tone mapped high dynamic range images," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, 2010.

[44] R. Mantiuk, S. Daly, K. Myszkowski, and Seidel, "Predicting visible differences in high dynamic range images - model and its calibration," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5666, pp. 204–214, 2005.

[45] P. G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*. Washington: SPIE Optical Engineering Press, 1999.

[46] S. C. Park, M. K. Park, and M. G. Kang;, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

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

[48] K. P. Hong, J. K. Paik, H. J. Kim, and C. H. Lee, "An edge-preserving image interpolation system for a digital camcorder," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 279–284, 1996.

[49] M. J. Chen, C. H. Huang, and W. L. Lee, "A fast edge-oriented algorithm for image interpolation," *Image and Vision Computing*, vol. 23, no. 9, pp. 791–798, 2005.

[50] V. Patel and K. Mistree, "A review on different image interpolation techniques for image enhancement," *Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 12, 2013.

[51] A. R. Reibman and T. Schaper, "Subjective performance evaluation for super-resolution image enhancement," in *Second Int. Wkshp on Video Proc. and Qual. Metrics (VPQM'06)*, 2006.

[52] A. R. Reibman, R. M. Bell, and S. Gray, "Quality assessment for super-resolution image enhancement," in *In Proc. of International Conference on Image Processing (ICIP)*, pp. 2017–2020, 2006.

[53] J. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, 2011.

[54] J. Lee, L. Goldmann, and T. Ebrahimi, "Paired comparison-based subjective quality assessment of stereoscopic images," *Multimedia Tools and Applications*, vol. 67, no. 1, pp. 31–48, 2013.

[55] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks," *Psychological review*, vol. 113, no. 4, pp. 700–765, 2006.

[56] S. H. Bae, T. N. Pappas, and B. H. Juang, "Subjective evaluation of spatial resolution and quantization noise tradeoffs," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 495–508, 2009.

[57] Y. Xue, Y. Ou, Z. Ma, and Y. Wang, "Perceptual video quality assessment on a mobile platform considering both spatial resolution and quantization artifacts," in *18th International Packet Video Workshop (PV)*, pp. 201–208, IEEE, 2010.

[58] Y. Ou, Y. Xue, and Y. W. Z. Ma, "A perceptual video quality model for mobile platform considering impact of spatial, temporal, and amplitude resolutions," in *IEEE IVMSP Workshop*, pp. 117–122, IEEE, 2011.

[59] G. Cermak, M. Pinson, and S. Wolf, "The relationship among video quality, screen resolution, and bit rate," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 258–262, 2011.

[60] D. Wang, F. Speranza, T. M. A. Vincent, and P. Blanchfield, "Toward optimal rate control: a study of the impact of spatial resolution, frame rate, and quantization on subjective video quality and bit rate," in *Visual Communications and Image Processing 2003*, pp. 198–209, International Society for Optics and Photonics, 2003.

[61] A. M. Demirtas, A. R. Reibman, and H. Jafarkhani, "Full-reference quality estimation for images with different spatial resolutions," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2069–2080, 2014.

[62] E. C. Reed and J. S. Lim, "Optimal multidimensional bit-rate control for video communication," *IEEE Transactions on Image Processing*, vol. 11, no. 8, pp. 873–885, 2002.

[63] E. Akyol, A. M. Tekalp, and M. R. Civanlar, "Content-aware scalability-type selection for rate adaptation of scalable video," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 214–214, 2007.

[64] K. N. Ngan, K. S. Leong, and H. Singh, "Adaptive cosine transform coding of images in perceptual domain," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 11, pp. 1743–1750, 1989.

[65] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment: The natural scene statistic model approach," *IEEE Signal Processing Magazine*, vol. 28, pp. 29–40, 2011.

[66] J. P. Guilford, *Psychometric methods*. New York: McGraw-Hill, 1954.

[67] Y. L. Grand, *Light, colour and vision*. London: Chapman and Hall, 1968.

[68] W. J. Crozier, "On the variability of critical illumination for flicker fusion and intensity discrimination," *Journal of General Physiology*, vol. 19, pp. 503–522, 1935.

[69] D. Kelly, "Effect of sharp edges on the visibility of sinusoidal gratings," *J Opt Soc Amer*, vol. 60, no. 1, pp. 98–103, 1970.

[70] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. on Communications*, vol. 31, pp. 532–540, Apr. 1983.

[71] E. P. Simoncelli and B. A. Olshausen, *Natural image statistics and neural representation*, vol. 24 of *Annual Review of Neuroscience*. 2001.

[72] M. Čadík and P. Slavík, "The naturalness of reproduced high dynamic range images," in *IV '05: Proceedings of the Ninth International Conference on Information Visualisation*, (Washington, DC, USA), pp. 920–925, IEEE Computer Society, 2005.

[73] http://www-2.cs.cmu.edu/afs/cs/project/cil/www/v-images.html.

[74] http://www-staff.lboro.ac.uk/ cogs/datasets/UCID/ucid.html.

[75] V. Mante, R. Frazor, V. Bonin, W. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nature Neuroscience*, vol. 8, no. 12, pp. 1690–1697, 2005.

[76] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, 2010.

[77] Open Source Community, "Qtpfsgui Project Homepage," http://qtpfsgui.sourceforge.net/index.php.

[78] E. Reinhard's high dynamic range data http://www.cs.utah.edu/ reinhard/cdrom/hdr/.

[79] G. Ward's high dynamic range data http://www.anyhere.com/gward/pixformat/tiffluvimg.html.

[80] P. Debevecs high dynamic range data, http://www.debevec.org/Research/HDR/.

[81] E. Adelson, C. Anderson, J. Bergen, P. Burt, and J. Ogden, "Pyramid methods in image processing," *RCA Engineering*, vol. 29, pp. 33–41, 1984.

[82] P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition.," in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 575, pp. 173–181, 1985.

[83] M. Smith and J. Hather, "Review of image fusion technology in 2005," *in Proc. of SPIE*, vol. 575, pp. 173–181, 1985.

[84] R. S. Blum and Z. Liu, *Multi-sensor image fusion and its applications*. Taylor & Francis, 2006.

[85] Z. Wang and E. P. Simoncelli, "Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics," in *Human Vision and Electronic Imaging IX, Proc. SPIE*, vol. 5292, pp. 99–108, Jan. 2004.

[86] Z. Wang and E. P. Simoncelli, "Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities," *Journal of Vision*, vol. 8, pp. 1–13, Sept. 2008.

[87] Z.Wang, G.Wu, H. R. Sheikh, E. P. Simoncelli, E.-H. Yang, , and A. C. Bovik, "Quality-aware images," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1680–1689, 2006.

[88] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[89] D. J. Field and N. Brady, "Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes," *Vision Research*, vol. 37, no. 23, pp. 3367–83, 1997.

[90] E. P. Simoncelli and W. T. Freeman, "Steerable pyramid: a flexible architecture for multi-scale derivative computation," in *IEEE International Conference on Image Processing*, vol. 3, pp. 444–447, 1995.

[91] J. Bigun, G. H. Granlund, and J. Wiklund, "Multidimensional orientation estimation with applications to texture analysis and optical flow," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 8, pp. 775–790, 1991.

[92] X. Feng and P. Milanfar, "Multiscale principal components analysis for image local orientation estimation," in *Conference Record of the Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 478–482, 2002.

[93] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3116–3132, 2010.

[94] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal processing Letters*, vol. 22, no. 3, pp. 209–212, 2013.

[95] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, pp. 2861–2873, 2010.

[96] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No reference perceptual quality assessment of JPEG compressed images," in *IEEE International Conference on Image Processing*, vol. 1, pp. I/477–I/480, 2002.