NO-REFERENCE QUALITY ASSESSMENT OF DIGITAL IMAGES

ZHANG JING

(M.Eng., Shandong University)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELECTRICAL & COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2010

Acknowledgements

I am grateful to all the people who provided their invaluable assistance for my PhD research.

First of all, I would like to express my gratitude to my supervisors, Prof. Ong Sim Heng and Dr. Le Minh Thinh. In particular, I would like to thank Prof. Ong for his critical assistance. I also appreciate the former supervision of Dr. Stefan Winkler, which prepared me well for my research.

I am thankful to Prof. Wang Xin for his patient advice and warm encouragement.

Further, I wish to thank my dear friends Gao Hanqiao, Wu Yuming, Shao Shiyun, Fan Dongmei, Cao Lingling, Rajesh, Tian Xiaohua, and Leng Yan. It is their encouragement that helped me through difficulties. I also thank all my lab mates at the Biosignal Processing Lab for creating a pleasant working environment.

Furthermore, thanks go to the anonymous reviewers of my papers for their constructive comments.

Last but not least, I would like to express my special gratitude to my beloved

parents and the rest of my family. It is their selfless love that encouraged me to complete this thesis.

Contents

Summary

Objective image quality measures have been developed to quantitatively predict perceived image quality. They are of fundamental importance in numerous applications, such as to benchmark and optimize different image processing systems and algorithms, to monitor and adjust image quality, and to develop perceptual image compression and restoration technologies, etc. As an important approach for objective image quality assessment, no-reference image quality assessment seeks to predict perceived visual quality solely from a distorted image and does not require any knowledge of a reference (distortion-free) image. No-reference image quality measures are desirable in applications where a reference image is expensive to obtain or simply not available. The intrinsic complexity and limited knowledge of the human visual perception pose major difficulties in the development of no-reference image quality measures. The field of no-reference image quality assessment remains largely unexplored and is still far from being a mature research area. Despite its substantial challenges, the development of no-reference image quality measures is a rapidly evolving research direction and allows much room for creative thinking.

The number of new no-reference image quality measures being proposed is growing rapidly in recent years. This thesis focuses on the development of no-reference image quality measures.

One contribution of this thesis is the kurtosis-based no-reference quality measures developed for JPEG2000 compressed images. The proposed no-reference image quality measures are based on either 1-D or 2-D kurtosis in the discrete cosine transform domain of general image blocks. They are simple, they do not need to extract edges/features from an image, and they are parameter free. Comprehensive testing demonstrates their good consistency with subjective quality scores as well as satisfactory performance in comparison with both representative full-reference image quality measures and state-of-the-art no-reference image quality measures.

The second contribution of this thesis is a pixel activity-based no-reference quality measure developed for JPEG2000 compressed images. Based on the basic activity of general pixels, the proposed no-reference quality measure overcomes the limitations imposed by structure/feature extraction of distorted images. The structural content-weighted pooling approach in the proposed image quality measure does not require any parameters and avoids additional procedures and training data for parameter determination. The proposed image quality measure exhibits satisfactory performance with reasonable computation load and easy implementation. It proves a no-reference quality measure of choice for JPEG2000 compressed images.

The third contribution of this thesis is the development of a structural activity-based framework for no-reference image quality assessment. Under the assumption that human visual perception is highly sensitive to the structural information in a scene, such a framework predicts image quality through quantifying the structural

activities of different visual significance. As a specific example, a model named structural activity measure is developed. The model is validated with a variety of distortions including white noise, Gaussian blur, and JPEG and JPEG2000 compression. The effectiveness of the model is demonstrated through the comparison with subjective quality scores as well as representative full-reference image quality measures. The structural activity-based framework proves effective for no-reference image quality assessment.

The work presented in this thesis is not limited to the development of effective techniques for no-reference image quality assessment. It may also contribute to a better understanding of the working mechanisms underlying human visual perception.

List of Tables

List of Figures

Acronyms

bpp	bits per pixel
CC	(Pearson) correlation coefficient
DCT	discrete cosine transform
DMOS	difference mean opinion score
DWT	discrete wavelet transform
FR	full-reference
HVS	human visual system
MC	monotonically-changing
MMF	multistage median filter
MOS	mean opinion score
MSE	mean-squared error
NR	no-reference
OR	outlier ratio
PDF	probability density function
PSNR	peak signal-to-noise ratio

RMSE	root mean square error
ROCC	(Spearman) rank order correlation coefficient
RR	reduced-reference
SA	structural activity
SSIM	structural similarity
VQEG	video quality experts group
ZC	zero-crossing

Chapter 1

Introduction

The goal of objective image quality assessment is to develop computational models to quantitatively predict perceived image quality. *No-reference* (NR) image quality assessment does not require a distortion-free image as reference and predicts image quality solely from a distorted image. NR image quality measures are highly desirable in practical applications where a reference image is expensive to obtain or simply not available. As an open research field with enormous practical potential, NR image quality assessment is a promising direction with many possibilities and is currently an active and rapidly evolving research area.

In this chapter, the motivation for developing NR image quality measures is presented in Section 1.1, the background of image quality assessment is introduced in Section 1.2, the contributions of this thesis are summarized in Section 1.3, and the outline of this thesis is provided in Section 1.4.

1.1 Motivation

As the saying goes, seeing is believing. Human beings rely highly on visual information to perceive the world. As our world becomes increasingly digital, digital images and videos rapidly proliferate. Digital images are the representation of visual information in a discrete form suitable for storage and transmission. They are subject to diverse distortions during acquisition, compression, processing, transmission, and reproduction. It is crucial to recognize and quantify the quality degradation of images. For example, lossy compression techniques, which are widely applied to reduce bandwidth for the storage and transmission of images, produce artifacts in the reconstructed images and may result in decreased visual quality. It is important to evaluate the visibility of compression artifacts so as to optimize the parameter settings of the related systems and applications. As another example, images are subject to errors, loss, or delay when they are distributed in various communication networks. All these transmission impairments may lead to the poorer quality of the received images. It is imperative for the network server to recognize image quality degradation so as to control streaming resources in transmission.

Given that the *human visual system* (HVS) is the ultimate receiver of most visual information resulting from various applications, the most reliable way for quality assessment is to resort to the judgement of human observers. However, such subjective quality assessment is time-consuming, expensive, and impractical in real-world applications especially for real-time applications. It is desirable to develop computational models that are able to quantitatively and automatically predict perceived image quality. This is the basic motivation for developing objective image quality measures. The final goal of objective image quality measures is to

predict quality the way the HVS does.

Objective image quality measures play an important role in a broad range of applications, including:

- benchmarking different image processing techniques and systems;
- optimizing image processing systems and algorithms;
- monitoring and adjusting image quality;
- developing perceptual image compression and restoration technologies.

Besides the distorted images under quality evaluation, three types of knowledge may be employed in objective image quality assessment: knowledge about the original distortion-free image which is assumed to have perfect quality, knowledge about the distortion process, and knowledge about the HVS. In many real-world applications, knowledge of a distortion-free image is not always available. In this situation, image quality can only be predicted from the distorted images themselves. The fact that the HVS can easily perceive image quality without any reference motivates the kind of image quality assessment without referring to a distortion-free image. Thus, both the practical requirements and the working mechanism of the HVS motivate the kind of image quality assessment without reference to a distortion-free image, i.e., NR image quality assessment.

1.2 Background

The standard objective image quality assessment is the *full-reference* (FR) approach in which a reference image of perfect quality (free of distortion) is assumed

to be completely known to compare with the image under assessment. Another type of objective image quality assessment is known as the *reduced-reference* (RR) approach, which assumes that the reference image is only partially available, such as certain features extracted from a reference image, to provide side information for image quality prediction. The third type is the NR approach, which is also referred to as blind or single-ended or univariant image quality assessment in the literature. NR image quality assessment appraises quality solely from a distorted image without any reference to a distortion-free image. NR image quality measures are highly desirable in practical applications where a reference image is expensive to obtain or simply not available. Due to its intrinsic difficulty, the field of NR image quality assessment is still in its preliminary stages and remains largely unexplored to date. So far, the development of NR image quality measures largely lags the advances in the field of FR image quality assessment. More detailed descriptions of image quality assessment can be found in [1–3].

1.3 Thesis Contributions

This thesis focuses on the development of NR image quality measures. Three different kinds of novel NR image quality measures are presented, including kurtosis-based quality measures, a pixel activity-based quality measure, and a structural activity-based framework. As application-specific NR image quality measures, the kurtosis and pixel activity -based quality measures are developed particularly for

JPEG2000 compressed images. General-purpose NR image quality assessment applicable to all kinds of distortions is an extremely difficult task. We seek to approach the general-purpose goal by developing a structural activity-based framework that is applicable to a variety of distortions. The major contributions of this thesis are summarized below.

(a) Kurtosis-Based No-reference Image Quality Measures: JPEG2000

In this study, kurtosis-based quality measures operating in the *discrete cosine transform* (DCT) domain are developed for NR quality assessment of JPEG2000 compressed images. The proposed quality measures are based on either 1-D or 2-D kurtosis of general image blocks. Specifically, three NR image quality measures are developed, which are based, respectively, on frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis. The proposed image quality measures have these advantages: they are simple, they do not need to extract edges/features, they are parameter free, and their quality predictions are shown to be in good agreement with subjective quality scores.

(b) Pixel Activity-Based No-reference Image Quality Measure: JPEG2000

In this study, a pixel activity-based quality measure is developed for NR quality assessment of JPEG2000 compressed images. The proposed image quality measure is designed with reasonable computation expense and easy implementation. Instead of extracting structures/features from an image, the proposed quality measure predicts image quality based on the basic activity of general pixels. Specifically, pixel activity is expressed in terms of the monotonically-changing and the zero-crossing activity. The proposed quality measure thus overcomes the limitations imposed by structure/feature extraction of distorted images, i.e., decreased

extracted structures/features under severe distortion and inconvenience incurred by the associated threshold operation. A pooling approach, which is given the name structural content-weighted pooling, is also proposed. This approach does not require any parameters and avoids additional procedures and training data for parameter determination. The proposed NR image quality measure exhibits consistently close correlation with subjective quality scores when the processing block size is subject to a wide range.

(c) Structural Activity-Based Framework for No-Reference Image Quality Assessment

In this study, a structural activity-based framework is proposed for NR image quality assessment. Based on this framework, a structural activity indicator is developed. Under the assumption that human visual perception is highly sensitive to the structural information in a scene, the structural activity framework estimates image quality by quantifying structural activity in an image. The effectiveness of the structural activity-based framework is validated with a variety of distortions, including white noise, Gaussian blur, and JPEG and JPEG2000 compression.

1.4 Thesis Organization

Since the knowledge of the HVS plays a fundamental role in the design of objective image quality measures, Chapter 2 provides a brief overview of the HVS and focuses on those aspects of the physiological and psychophysical properties that are relevant to the image quality assessment models discussed in this thesis.

Chapter 3 reviews some representative work reported in the fields of FR, RR, and NR image quality assessment, as well as the research effort devoted to the

validation of objective quality measures.

Chapter 4 presents the proposed kurtosis-based NR image quality measures. It includes the calculation of 1-D and 2-D kurtosis in the DCT domain, the demonstration of the working principle of kurtosis in image quality prediction, the approach of kurtosis-based image quality measures, qualitative and quantitative performance evaluations, and outlier analysis.

Chapter 5 presents the proposed pixel activity-based NR image quality measure. It describes the expressions of pixel activity, the structural activity-based pooling approach, qualitative and quantitative performance evaluations, and outlier analysis.

Chapter 6 presents the proposed structural activity-based NR image quality assessment framework together with a model named structural activity indicator. It presents the concept of structural activity, the model of structural activity indicator, qualitative and quantitative performance evaluations, and outlier analysis.

Chapter 7 provides a summary, highlights the contributions of this thesis, and gives the recommendations for the future work.

Chapter 2

Human Visual System

The *human visual system* (HVS) is extremely complex. Numerous psychophysical and physiological studies in the past century have gained considerable knowledge about the HVS. However, due to the intrinsic complexity of the HVS, current knowledge is largely limited to the early vision stage, and many properties and working mechanisms of the later visual pathways and higher-level cognitive processes that occur in the visual cortex are still not well understood. Since the final goal of objective image quality assessment is to emulate or at least perform close to human quality perception, the knowledge of HVS plays a fundamental role in the design of objective image quality measures. This chapter provides a brief overview of the HVS and focuses on those aspects of the physiological and psychophysical properties that are relevant to the image quality assessment models discussed in this thesis. Specifically, the anatomy of the early HVS and its related psychophysical properties are provided in Sections 2.1 and 2.2, respectively. More detailed knowledge of the HVS can be found in [4, 5].

2.1 Anatomy of the Early Human Visual System

A schematic diagram of the early HVS is shown in Figure 2.1. During its extensive exposure to the visual environment in the long evolution, the HVS is well adapted to extracting useful information for visual perception. There are roughly two major stages in the human vision. In the early stages, the eyes capture light and convert the visual stimulus into signals which can be interpreted by the neurons in the human brain. In the later stages, the human brain extracts the higher-level cognitive information for visual perception.

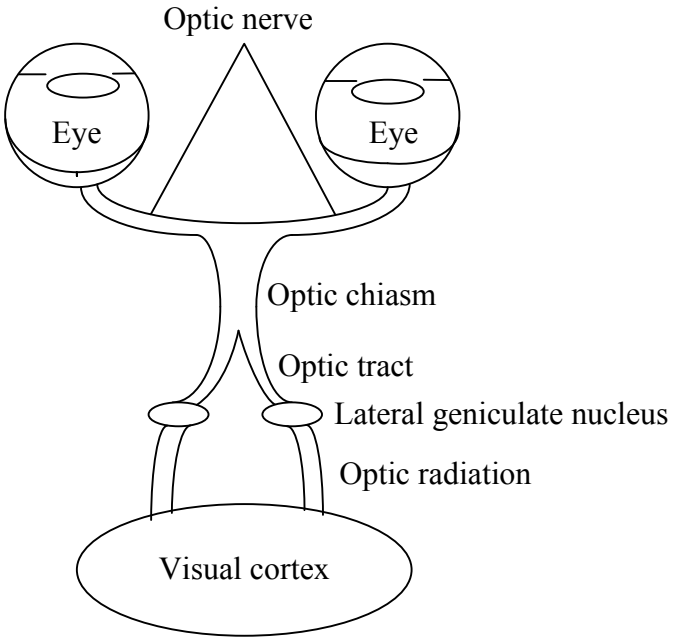


Figure 2.1: Schematic diagram of the early HVS.

As an important component of the HVS, the eye plays a role equivalent to a photographic camera. A simplified transverse section of the human left eye is illustrated in Figure 2.2. It is shown that the optical components of the eye are mainly composed of the cornea, the pupil (a circular opening in the center of

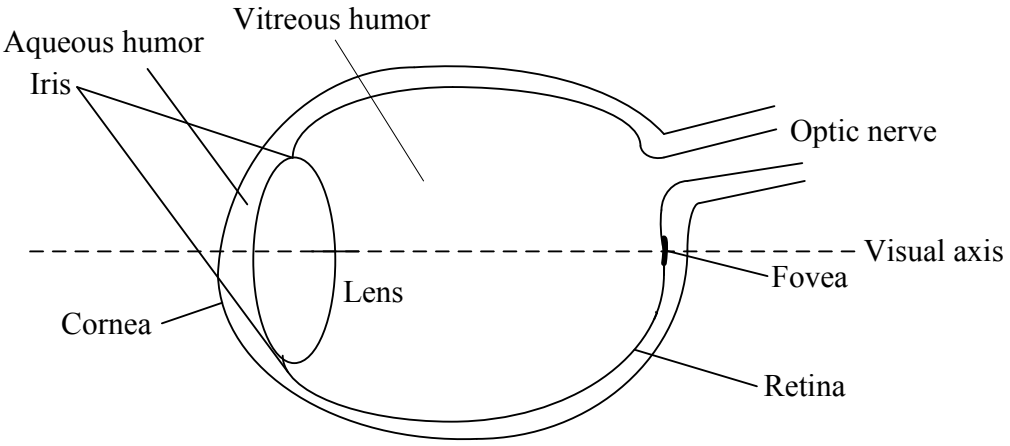


Figure 2.2: The simplified transverse section of the human left eye.

the iris), the lens, and the fluids filling the eye including the aqueous humor and vitreous humor. The visual stimulus in the form of light first encounters the eye at the cornea which provides the major optical power of the HVS. Then the light enters the eye through the pupil. Depending on the exterior light levels, the size of the pupil can be changed under muscular control, and the amount of light entering the eye is controlled. This makes the pupil equivalent to the eye's aperture. After passing through the watery fluid of aqueous humor, the light enters the lens of the eye. An important characteristic of the lens is that its optical power can be altered with accommodation, a process in which the curvature of the lens is modified by the contraction of the muscles attached to it. The process of accommodation enables the HVS to focus objects at different distances onto the back of the eye. The main body of the eyeball is filled with the gelatinous fluid of vitreous humor. After passing through the vitreous humor, the light coming from objects is finally focused on the retina, which is a membrane of neural tissue at the back of the eye. The projection of the visual stimulus onto the retina is a blurred image of the visual field due to the inherent limitations and imperfections of the optical system

in the human eye.

As an extension of the central nervous system, the retina is composed of several layers of neurons. The layer of photoreceptors in the retina are light-sensitive neurons and converts light into signals that can be understood by the human brain. There are two types of photoreceptor cells, namely the rods and the cones. The rods are sensitive to luminance at low light levels and are responsible for vision under very low light conditions, while the cones are sensitive to color at high light levels and are responsible for vision under normal light conditions. The distribution of photoreceptors varies largely over the surface of the retina. Cones are concentrated in the fovea and its density rapidly declines with the distance from the fovea, while rods dominate the region outside the fovea and the central fovea contains no rods at all. The fovea (as shown in Figure 2.2) is a small area at the center of the retina. The concentrated distribution of cones in the fovea results in high-resolution vision only over a small region around the point of fixation (projected onto the fovea) and quickly decreased resolution with distance from the fixation point. The sampled signals from photoreceptors are further processed by several layers of interconnecting retinal neurons and then transmitted to the ganglion cells where the optic nerves carry the output signal of the retina to the brain.

As illustrated in Figure 2.1, the optic nerves carry visual information leaving from the retina via the optic chiasm, the optic tract, the lateral geniculate nucleus, the optic radiation, to the visual cortex in the brain. The visual cortex is responsible for the high-level aspects of human vision. The primary visual cortex is the layer of visual cortex that makes up the largest part of the HVS. It is found that a large number of neurons in the primary visual cortex respond strongly to

certain types of information, such as some specific spatial and temporal frequencies, orientations, phases, colors, velocities, and directions of motions, etc. The receptive fields of neurons could be well described using localized, band-pass, and oriented functions. The visual streams generated in the visual cortex are carried off to other parts of the brain for further processing, such as motion sensing and high-level cognitive understanding. Current knowledge is largely limited to the low-level processes of human vision. The precise functional mechanisms of the high-level processes occurring in the human brain remains an active research area in vision science.

2.2 Psychophysical Properties of the Human Visual System

This subsection describes those psychophysical properties that are relevant to the image quality measures discussed in this thesis.

Light Adaptation - The HVS is able to adapt to a wide range of light intensities from a few photons to bright sunlight. This phenomenon is known as light adaptation, which operates mainly via the mechanical variation of the pupillary aperture to control the quantity of light entering the eye, and the chemical processes in the photoreceptors of rods and cones. The HVS is more sensitive to the relative variation of luminance than to the absolute luminance over a wide range of background luminances. This property allows the HVS to better discriminate the relative intensity variations at each light level. The human visual perception

of luminance can be approximated by the Weber-Fechner law:

$$\frac{\Delta I}{I} = K \tag{2.1}$$

where I is the background luminance, ΔI is the just noticeable incremental luminance perceived by the HVS over the background, and K is a constant. The Weber-Fechner law holds over a wide range of background luminance which covers the luminances in most image processing applications.

Contrast Sensitivity Functions - The contrast sensitivity functions model the variations in the visual sensitivity to different spatial and temporal frequencies in visual stimulus. In the modeling of the HVS, the contrast sensitivity functions are typically implemented as filtering operations or weighted subbands after frequency decomposition. Contrast sensitivity is also a function of temporal frequencies, which has been modeled as temporal filters in video quality assessment.

Masking - Masking refers to the reduction in the visibility of one visual stimulus (called the signal) due to the simultaneous presence of another (called the mask). It is basically due to the limitations in sensitivity of the retinal neurons in relation to the activity of its surrounding neurons. The masking effect is strongest when the mask and the signal have similar characteristics, such as similar spatial locations, frequency components, orientations, and colors, etc. Two typical visual masking effects are luminance masking and texture masking. Luminance masking refers to the effect that the visibility of a visual stimulus is maximum for medium background intensity, and the visibility reduces when the visual stimulus occurs against a very low or very high intensity background. Luminance masking is mainly due to the brightness sensitivity of the HVS. The average brightness of

the surrounding background can alter the visibility threshold of the visual stimulus. Texture masking refers to the effect that a visual stimulus is more visible in homogeneous areas than in textured or detailed areas. In textured image regions, small variations in the texture are masked by the macro properties of genuine high-frequency details, and therefore, are not easily perceived by the HVS.

Pooling - Pooling refers to the task of making a perceptual decision from the visual streams. It is still not well understood how the HVS performs the task of pooling, but high-level cognitive understanding should play an important role. In image quality measures, the Minkowski pooling strategy with the expression given below is usually employed to pool the error signals across spatial locations or different channels (usually in terms of different frequency and orientation components) to obtain a single scalar as image quality score.

$$E(\{e_{l,k}\}) = (\sum_{l} \sum_{k} |e_{l,k}|^{\beta})^{\frac{1}{\beta}} \tag{2.2}$$

where $e_{l,k}$ is the normalized error of the kth coefficient in the lth channel, and β is a constant exponent typically chosen between 1 and 4. Minkowski pooling may be implemented in the spatial space (indexed by k) and then over different channels (indexed by l), or vice versa.

Chapter 3

Literature Review

This chapter reviews the related work in the field of image quality assessment. FR image quality measures are reviewed in Section 3.1, RR image quality measures in Section 3.2, NR image quality measures in Section 3.3, and the validation of image quality measures is detailed in Section 3.4.

3.1 Full-Reference Image Quality Assessment

Assuming full access to a reference image, FR image quality measures predict quality through the comparison between a reference and a distorted image. A reference image is distortion-free and is assumed to have "perfect" quality. As a standard approach of image quality assessment, FR image quality measures have received a great deal of attention over the past decades. A considerable percentage of the literature is devoted to the development of FR image quality measures, or more precisely, image fidelity or similarity measures. However, only in recent years have the relatively "easier" FR approaches been developed to predict image quality in good consistency with perceived visual quality.

The classical FR image quality measures of *mean-squared error* (MSE) and its variant *peak signal-to-noise ratio* (PSNR) have found widespread use due to their simplicity and mathematical convenience. Assume that $\mathbf{I}$ denotes the original image and $\hat{\mathbf{I}}$ denote a distorted version of $\mathbf{I}$, both with n-bit pixel values (i.e., intensities) in the range $[0, 2^n - 1]$ (e.g., $[0, 255]$ for 8-bit images). With the distortion $\mathbf{E} = \mathbf{I} - \hat{\mathbf{I}}$, the MSE between $\mathbf{I}$ and $\hat{\mathbf{I}}$ is expressed by

$$\text{MSE} = \frac{\|\mathbf{E}\|^2}{N} = \frac{1}{N}\sum_{i=1}^{N} E_i^2 \tag{3.1}$$

and the PSNR by

$$\text{PSNR} = 10\log_{10}\left(\frac{(2^n-1)^2}{\text{MSE}}\right) = 20\log_{10}(2^n-1) - 20\log_{10}\left(\frac{\|\mathbf{E}\|}{\sqrt{N}}\right) \tag{3.2}$$

where $\|\cdot\|$ denotes the L_2-norm, E_i denotes the ith distortion value in $\mathbf{E}$, and N denotes the number of pixels. The PSNR is useful when the images being compared have different ranges of pixel values, but it contains no new information relative to the MSE. The definitions (3.1) and (3.2) show that MSE and PSNR operate based on the energy of pixel-wise distortions $\|\mathbf{E}\|$. Despite being widely used over a very long time, MSE and PSNR have been widely criticized for their limited accuracy when estimate perceived visual quality, e.g., [6–10].

With an intention to predict image quality in a similar way as the HVS, the perceptual image quality measures following a "bottom-up" approach have been developed to mathematically model the functional components in the HVS that are relevant to image quality assessment. Although the HVS is extremely complex and its many properties are still not well understood, it is of great interest in the past three decades to deploy the relevant features of the HVS to predict

image quality [11, 12]. Some representative HVS-based image quality measures are reviewed below.

Lubin's (or Sarnoff's) model [13–15] predicts image quality by estimating the probability of the differences between two images being compared. To obtain the probability map, the model filters and resamples an image in a way to simulate eye optics and the retinal photoreceptor sampling, and decompose the image using a Laplacian pyramid [16] followed by the band-limited contrast calculations [17]. Next, the signal is further decomposed using a bank of steerable filters [18] to reflect the orientation selectivity of the HVS, followed by a normalization operation determined by the contrast sensitivity functions and an implementation of point nonlinearity to account for the intra-channel masking of the HVS. The normalized error signal is then convolved with disk-shaped kernels before a Minkowski pooling across scales. The errors across the spatial space after the pooling stage are then converted into a probability-of-detection map.

The Teo-Heeger Model [19, 20] involves two major components: a steerable pyramid transform [21] and contrast normalization. Specifically, a steerable pyramid decomposition is used to accomplish the channel decomposition to account for the observation that a large number of neurons in the primary visual cortex are tuned to visual stimuli with specific spatial locations, frequencies, and orientations, and the normalization scheme is motivated from those models that have been widely used to explain physiology data in early visual systems.

Watson's DCT model [22] was originally designed for JPEG optimization. In this model, an image is divided into distinct blocks, and a visibility threshold is calculated for each DCT coefficient in each block. The visibility threshold is determined by three factors to simulate the properties of the HVS, namely, baseline

contrast sensitivity, luminance masking, and contrast/texture masking. The errors between the reference and distorted images are normalized using the visibility threshold, and are pooled spatially and across frequencies to obtain the final image quality estimation.

Karunasekera *et al.* [23] developed a distortion measure based on the human visual sensitivity to horizontal and vertical edge artifacts resulting from block DCT-based image compression.

Miyahara *et al.* [24] reported a picture quality scale which combines a number of human visual properties for both global features and localized distortions, including light adaptation according to the Weber-Fechner law, contrast sensitivity, and visual masking of the HVS. Winkler [25] developed a perceptual distortion metric for color images based on the following properties of visual perception: color perception and the theory of opponent colors, the response properties of neurons in the primary visual cortex, and contrast sensitivity and contrast masking of the HVS.

Damera-Venkata *et al.* [26] models the degradation as the linear frequency distortion and additive noise injection. Two complementary measures were thus developed to quantify the separate distortions. Specifically, the frequency distortion is quantified based on a model of the frequency response of the HVS over visible frequencies, and the noise distortion is quantified by taking into account the HVS properties including the variation in contrast sensitivity, the variation in the local luminance mean, the contrast interaction between spatial frequencies, and the contrast masking effects of the HVS.

Wang *et al.* [27] developed an image quality index in the wavelet transform domain, namely the foveated wavelet image quality index, which takes into account

the following HVS factors: the space variance of the contrast sensitivity function, the spatial variance of the local visual cut-off frequency, the variance of human visual sensitivity in different wavelet subbands, and the influence of the viewing distance on the display resolution and the HVS features.

Based on the noticeable local contrast changes as perceived by the HVS, Lin *et al.* [28] presented a distortion metric by discriminatively analyzing the impact of pixel differences in visual quality. A scheme for estimating just-noticeable distortion is proposed in [29]. This scheme proposes a new formula for luminance adaptation adjustment and incorporates block classification for contrast masking of the HVS.

Chandler *et al.* [30, 31] developed a metric named the visual signal-to-noise ratio which quantifies the visual fidelity of natural images based on near-threshold and supra-threshold properties of the human vision. In Chandler's metric, it is first determined that whether the distortions are visible through the comparison with contrast thresholds that are computed via wavelet-based models of visual masking and visual summation. If the distortions are below the threshold of visual detection, the distorted image is deemed to have perfect visual fidelity. If the distortions are supra-threshold, the distortions are quantified based on the low-level visual property of perceived contrast and the mid-level visual property of global precedence.

Different from the "bottom-up" HVS-based approaches, the "top-down" image quality measures treat the HVS as a black box and only the input-output relationship of the HVS is of concern. They are based on the hypotheses regarding the overall functionality of the HVS. A notable feature of the top-down approaches is that they may provide much simplified computational models for image quality

assessment.

A representative top-down approach is the *structural similarity* (SSIM) approach advanced by Wang *et al.* [32,33], which is employed as a benchmark for the performance evaluation of the NR image quality measures developed in this thesis. The algorithm of the SSIM approach is detailed below.

Based on the assumption that the HVS is highly adapted to extract structural information from visual scenes, the basic idea of the SSIM approach is that a measurement of SSIM or structural distortion should provide a good prediction of perceived image quality. The SSIM index performs the task of similarity measurement based on the comparison of three aspects: luminance, contrast, and structure. First, the mean intensity is employed as the estimate of the signal luminance:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.3}$$

where i is the index and N is the total number of signal samples. Second, the standard deviation is employed as the estimate of the signal contrast:

$$\sigma_x = \left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2\right)^{\frac{1}{2}} \tag{3.4}$$

Third, the correlation coefficient between the two signals being compared is employed as the measure to quantify the structural similarity:

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{1}{\sigma_x\sigma_y}\left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)\right) \tag{3.5}$$

Suppose that $\mathbf{x}$ and $\mathbf{y}$ are local image patches taken from the same location of two images being compared. The similarity $l(\mathbf{x},\mathbf{y})$ of the local luminance, the similarity

$c(\mathbf{x}, \mathbf{y})$ of the local contrasts, and the similarity $s(\mathbf{x}, \mathbf{y})$ of the local structures are combined to form the local SSIM $S(\mathbf{x}, \mathbf{y})$:

$$S(\mathbf{x}, \mathbf{y}) = l(\mathbf{x}, \mathbf{y}) \cdot c(\mathbf{x}, \mathbf{y}) \cdot s(\mathbf{x}, \mathbf{y}) = \left(\frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \right)^{\alpha} \cdot \left(\frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \right)^{\beta} \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \right)^{\gamma} \tag{3.6}$$

where μ_x and μ_y are respectively the local mean of $\mathbf{x}$ and $\mathbf{y}$, σ_x and σ_y are respectively the local standard deviations of $\mathbf{x}$ and $\mathbf{y}$, and σ_{xy} is the cross correlation of $\mathbf{x}$ and $\mathbf{y}$. $\alpha > 0$, $\beta > 0$, and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. C_1, C_2, and C_3 are small positive constants used to stabilize each term, so that near-zero means, variances, or correlations do not lead to numerical instability. The choice of $C_1 = C_2 = C_3 = 0$ defines the initial version of SSIM, namely the universal image quality index [34, 35]. The dynamic range of $S(\mathbf{x}, \mathbf{y})$ is $-1 < S(\mathbf{x}, \mathbf{y}) \leq 1$ where a large value indicates greater similarity between a reference and a distorted image. The maximum value $S(\mathbf{x}, \mathbf{y}) = 1$ occurs if and only if $\mathbf{x} = \mathbf{y}$ for all $i = 1, 2, ..., N$. The SSIM index is symmetric with $S(\mathbf{x}, \mathbf{y}) = S(\mathbf{y}, \mathbf{x})$, so that two images being compared give the same index value regardless of their ordering. With $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$, the specific form of the SSIM index in [32] is given by

$$S(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.7}$$

The SSIM index is computed locally within a sliding window that moves pixel-by-pixel across an image, resulting in an SSIM map. The SSIM score of the entire

image is then computed by pooling the SSIM map:

$$Q(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{j=1}^{M} S(\mathbf{x}_j, \mathbf{y}_j) \tag{3.8}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are respectively a reference image and a distorted image; $\mathbf{x}_j$ and $\mathbf{y}_j$ are the local image patches at the jth sliding window corresponding to the reference and distorted images, respectively; and M is the total number of local windows of the images being compared. The SSIM index [32, 33] is the implementation of the SSIM approach at a single scale in the spatial domain. There also exists other forms of the SSIM approach. The implementation of SSIM over multiple scales [36,37] supplies more flexibility than the single-scale method by incorporating the variations of viewing conditions. An image synthesis method was proposed to calibrate the parameters which are used to define the relative importance of different scales. A complex wavelet domain version of SSIM [38] alleviates the sensitivity of SSIM in the spatial domain to geometric distortions, such as relative translations, scalings, and rotations of images.

Based on the statistical modeling of source, channel distortion, and receiver, the visual information fidelity approach developed by Sheikh *et al.* [39–42] seeks to relate image quality to the amount of information that is shared between a reference and a distorted image. Specifically, a reference image (i.e., the source) is modeled by a wavelet-domain Gaussian scale mixture [43], the distortions between a reference and a distorted images (i.e., the channel distortion) are modeled as a combination of a uniform wavelet-domain energy attenuation with the subsequent independent additive noise, and the visual distortion process (i.e., the receiver) is modeled as a stationary, zero-mean, additive white Gaussian noise process in

the wavelet domain. Finally, an image quality estimate is obtained by quantifying the shared information between a reference and a distorted images in terms of the mutual information, a widely used statistical measure of information fidelity in information theory.

Weken *et al.* [44] resort to the similarity measures in the fuzzy set theory for image quality assessment. The image quality estimate is given as the weighted average of the local similarities between the disjoint image patches taken from a reference and a distorted images. Specifically, the local similarity is calculated using one of the thirteen fuzzy similarity measures presented in [44] that were found appropriate for the comparison of images, and the weight is defined as the similarity between the homogeneities of the corresponding image patches where the homogeneity is computed as the similarity between the maximum and minimum intensities of an image patch.

Shnayderman *et al.* [45] presented a singular value decomposition-based image quality measure that can be used as either a graphical or a scalar measure. Han *et al.* [46] proposed to use LU factorization to represent the structural information of an image. Based on the LU factorization of a reference and a distorted image, a two-dimensional distortion map is constructed and an image quality score is computed from the distortion map.

Although the bottom-up and top-down approaches are motivated from substantially different design principles, they are complementary in many respects and have no sharp boundary. On one hand, it is impossible for a bottom-up approach to simulate all the related functional components in the complicated HVS. Many hypotheses have to be made in the bottom-up approaches to simplify the

simulation so as to achieve a tractable solution. On the other hand, a good understanding of the relevant functional components in the HVS is quite useful for a top-down approach to make reasonable hypotheses about the the overall functionality of the HVS. Indeed, it is desirable to find solutions that may combine the principles underlying the top-down and bottom-up approaches in the common task of image quality assessment.

3.2 Reduced-Reference Image Quality Assessment

RR image quality measures do not require full access to a reference image, and instead utilize only partial information such as the features extracted from a reference image to predict quality. The RR approach provides a compromise between the FR and the NR approaches. In concept, RR image quality assessment is easier than the NR approach at the cost of transmitting additional side information. The standard deployment of a RR image quality assessment system is to send the side information through a separate data channel which, however, may be inconvenient or expensive to provide in practical applications. An alternative solution is to send the side information using the same channel as the images being transmitted. In this case, the limited available data rate in practical applications lay some basic requirements for the selected RR features: they should provide efficient information to represent a reference image; they should be sensitive to image distortions; and they should have good perceptual relevance, etc. Based on the features extracted from a reference image together with a distorted image, a final scalar score is computed to describe the quality of a distorted image.

Quite a few RR image quality measures or systems were reported in the literature. Wang *et al.* [47] proposed a RR image quality measure based on a natural image statistics model in the wavelet transform domain. In Wang's quality measure, the Kullback-Leibler distance [48] is applied between the marginal distributions of wavelet coefficients of a reference and a distorted image. With a generalized Gaussian density model [49] to represent the marginal distribution of wavelet coefficients of an image, image quality is described by the fitting error between the wavelet coefficients of a distorted image and the Gaussian distribution of a reference image. Based on this RR image quality measure, a subsequent quality-aware image system was developed in [50], which combines the techniques of information data hiding, robust image communication, information data decoding, and RR image quality assessment. The basic concept of quality-aware image is to embed the features extracted from an original (high-quality) image into the image data as invisible hidden messages. When a distorted version of this image is received, users can decode and use the hidden messages to predict image quality by an RR quality measure.

Gao *et al.* [51] developed a RR framework by incorporating the merits from multiscale geometric analysis, contrast sensitivity functions, and the Weber-Fechner law of just noticeable difference (Section 2.2). In this framework, the multiscale geometric analysis is utilized to decompose images and extract features to mimic the multichannel structure of the HVS. The multiscale geometric analysis also offers a series of transforms to capture different types of image geometric information. To take into account the nonlinearities inherent in the HVS, contrast sensitivity functions are applied to weight the coefficients obtained by the multiscale geometric analysis. And the just noticeable difference is introduced to produce a noticeable

variation in sensory experience.

3.3 No-Reference Image Quality Assessment

In contrast to the FR and the RR image quality measures, which require full and partial access to a distortion-free image, NR image quality measures predict quality solely from a distorted image without any reference. Without any knowledge of a distortion-free image as reference, NR image quality assessment is no doubt a difficult approach to predict quality. Despite their intrinsic difficulty, NR image quality measures are highly desirable in real-world applications where a reference image is expensive to obtain or simply not available.

The development of NR image quality measures is largely motivated by the fact that human beings can easily identify image quality without seeing a distortion-free image, and moreover, different individual observers tend to have similar image quality opinions. However, the HVS is enormously complex and current knowledge of the human visual perception is far from complete. In particular, there exists such factors like learning, visual context, and cognitive relevance in the human visual perception that can hardly be quantified. NR image quality assessment thus turns out to be an extremely difficult task. The field of NR image quality assessment has experienced fast advances in recent years and the number of new approaches that are being proposed is growing rapidly. At present, most existing NR image quality measures are designed with prior knowledge of the distortion. Fortunately, the distortion process is known in many applications, and the task of developing distortion-specific NR quality measures is of practical importance.

The progress made in NR image quality assessment is closely related to the

advances in the technologies of image acquisition, display, compression, transmission, and reproduction. The distortions of blur and noise are usually generated in image acquisition and display systems, so the early NR quality measures were usually designed for blur/sharpness and noise measurement. With the evolution of high-quality image acquisition and display systems as well as the rapid proliferation of digital images, compression techniques are widely used to reduce bandwidth for the storage and transmission of digital images in numerous applications. Lossy image compression contains an inherent irreversible information loss, which produces visible and undesirable artifacts in the reconstructed images and leads to degraded visual quality, especially in the case of compression at low bit rates. In fact, compression artifacts and transmission errors are two major distortions in the experiments conducted by the *Video Quality Experts Group* (VQEG) [52] which aims to provide industrial standards for video quality assessment. These experiments indicate that compression and transmission have become two significant sources of image distortions in practical applications. The scope of this thesis concerns images subject to a single distortion. Considering that the distortion of transmission errors is basically a kind of multiple distortions composed of compression artifacts followed by the distortions during transmission, we do not discuss transmission errors in this thesis. The rest of this section will review those NR image quality measures in the literature that work with the distortions created by two widely used compression technologies, i.e., the block DCT and the wavelet -based image compressions.

In the standard block DCT-based JPEG compression [53], an image is partitioned into non-overlapping 8×8 blocks and the DCT is applied to each block. Then each DCT coefficient in each block is subject to independent quantization

and subsequent entropy coding. Due to the coarse quantization of block-based DCT coefficients at low bit rates, it is noted that blurring occurs within blocks, and blocking artifacts appear across block boundaries in JPEG compressed images [54]. The blurring distortion within coding blocks is due to the loss of high frequency components during the process of quantization and blocking artifacts, characterized by periodic horizontal and vertical discontinuities at the boundaries of the blocks, due to independent quantization within coding blocks. Since blocking artifacts are usually the most annoying distortion in JPEG compressed images, most NR quality measures are designed based on the adjustment over the quantification of blocking artifacts, either in the spatial domain [55–59] or in the frequency domain [60–62]. Some other approaches formulate the NR quality assessment of JPEG compressed images as a machine learning problem in which the functional relationship between image features and subjective quality ratings is learned by a training procedure [63–65].

Wu *et al.* [55] took into account the luminance masking effect and developed a generalized impairment metric at block edges to quantify the blocking artifacts. According to the subjective experiments carried out in [56], it was claimed that the quality of JPEG compressed images could be described based on a single distortion since the subjective data regarding the individual distortions were found to be highly correlated. Relying on this finding, Meesters *et al.* [56] proposed an image quality measure based on the blocking artifacts alone. The Hermite transform [66], a signal decomposition technique in which signals are locally approximated by polynomials within a Gaussian window, was used to detect the low-amplitude edges resulting from blocking artifacts, and edge amplitudes were estimated as the amount of blockiness.

Wang *et al.* [57] proposed to estimate the average luminance differences across block boundaries to indicate blockiness, and the average absolute difference between block image pixels together with the zero-crossing rate within coding blocks to indicate the activity of a JPEG compressed image. All these factors are combined to yield a quality score for a JPEG compressed image.

In [58], it was claimed that simple edge detection around block boundaries should be sufficient for the measurement of blocking artifacts, since the likelihood that a horizontal or vertical edge happens to be located exactly at the block boundary is small in natural images. Li [58] proposed to use the Prewitt operator to detect horizontal and vertical edges, and discard those gradients above a specified threshold since the blocking artifacts are usually weak edges. With the existence of blocking artifacts determined by a majority vote on the significance of local gradients, the effect of blockiness is indicated by the overall edge detection ratio at block boundaries.

In the quality measure reported in [59], blockiness is evaluated via block-by-block analysis based on inter-pixel differences at and near block boundaries, and flatness (blur) is evaluated in terms of the proportion of zero crossings within 8×8 regions near block boundaries. Contrast and spatial masking effects are also incorporated.

With a blocky image modeled as a nonblocky image interfered with by a pure blocky signal, Wang *et al.* [60] proposed to assess the blocking effect by evaluating the energy of the blocky signal in the frequency domain. The power spectra of the absolute difference signals in either the horizontal or vertical direction are estimated using the fast Fourier transform. The effect of blockiness is estimated as the power of the blocky signal which is computed after smoothing the power spectrum by

a median filter. Wang's blockiness measure also incorporates the luminance and texture masking effects.

In [61], an approach working in the DCT domain was developed for the blind measurement of blocking artifacts. The blocking artifacts are modeled as 2-D step functions within the shifted blocks constructed across two adjacent coding blocks, and the amount of blocking artifacts is estimated as the amplitude of the 2-D step function. The luminance and texture masking effects are also integrated into the quality measure.

In [62], the proposed image quality measure is based on natural scene statistics of the DCT coefficients whose distribution is modeled by a Laplace probability density function. The resulting coefficient distributions are used to estimate the local error due to JPEG encoding. The resulting local errors perceptually weighted by Watson's model [22] are pooled together to obtain an image quality score.

In [63], neural networks were used to predict image quality. The principle is to first "teach" a neural network to assess image quality using the features extracted from example images and subjective quality data provided by human observers, and then the "calibrated" neural network is used to assess the quality of images to be evaluated.

In [64], a circular back-propagation model-based neural network is trained to simulate the quality scoring process of human evaluators to perform the task of image quality assessment.

In [65], a NR quality measure was developed based on a sequential learning algorithm for growing and pruning radial basis function network. The network is trained to approximate the functional relationship between the features extracted from an image and subjective quality scores.

Although blocking artifacts are avoided in the subsequent wavelet-based JPEG2000 compression [67], blur and ringing artifacts arise from the coarse quantization of *discrete wavelet transform* (DWT) coefficients and are generally considered as the most significant distortions in JPEG2000 compressed images. The coarse quantization results in the truncation of the high-frequency DWT coefficients. It leads to visible irregularities around edges in the spatial domain, since the high-frequency components basically correspond to the edges in the spatial domain. These irregularities from natural appearance are usually referred to as ringing artifacts. Characterized by ripples or oscillations around sharp edges and contours in an image, the ringing artifacts can range from imperceptible to very annoying in a compressed image, depending on the data source, target bit rate, or underlying compression scheme. In contrast to the blocking artifacts that appear as periodic horizontal and vertical edges at predictable locations (i.e., boundaries of coding blocks), the ringing artifacts are strongly dependent on image content and compression degree without regularity to follow and largely complicate the task of NR quality assessment. Only a handful of NR quality measures in the literature are able to work with JPEG2000 compressed images.

In [68], binary morphological operators are used to isolate the regions of an image where the ringing artifacts are most visually prominent while preserving genuine edges and other fine details present in an image. The effect of ringing artifacts are evaluated by computing the image intensity variance in the vicinity of edges.

In [58], an approach was proposed to measure the ringing effect by examining the noise spectrum filtered out by anisotropic diffusion [69] implemented on an image. The basic idea is that the ringing artifacts would be mostly assimilated

into the noise spectrum since anisotropic diffusion is effective in deringing, and the noise spectrum would thus be colored if an image contains ringing artifacts. So the percentage of energy measured at high frequencies is employed to indicate the ringing effect.

The blur measures reported in [70, 71] are based on the measurement of edge spread which is virtually the number of pixels with monotonically-changing intensities along the gradient orientation at an extracted edge pixel. Both measures are applied to predict quality of JPEG2000 compressed images. The differences between [70] and [71] include: edges are detected by the Sobel operator in [70], and by the Canny operator [72] in [71]; edge spread is measured along the horizontal direction of a vertical edge in [70], and along the gradient orientation at a general edge pixel in [71].

Tong *et al.* [73] proposed the use of principal component analysis to extract local features by viewing all edge points as "distorted" or "un-distorted" in a JPEG2000 compressed image. The relationship between the local features and the local distortion metric is modeled based on the probability of an edge point being "distorted" or "un-distorted".

In [74], the blurring measure is formulated as the ratio of edge activity weighted by the probability of edge occurrence in the middle/low frequencies, while the ringing artifact measure is formulated as the ratio of the activity in the middle low over middle high frequencies in the ringing regions around strong edges. The blurring measure, the blocking artifact measure, and the percentage of strong edges in a compressed image are finally pooled together to yield an image quality score.

Sheikh *et al.* [75, 76] developed a model operating in the wavelet domain which incorporates both a natural image statistic model and an image distortion model

to predict quality of JPEG2000 compressed images. The basic idea is to investigate how the quantization processes in JPEG2000 compression modify the statistics of the image wavelet coefficients. Specifically, a natural image statistic model [77] is utilized to describe what a typical wavelet coefficient distribution of natural images should be like, and a distortion model associated with quantization is employed to quantify the departure from the natural image statistic model.

Sazzad *et al.* [78] proposed a quality measure based on pixel distortions and edge information. The pixel distortions are estimated using the local standard deviation and the absolute difference measure of a central pixel from the second closest neighborhood pixels. Edge information is estimated using the zero-crossing rate and the histogram measure with and without edge preserving filter. Finally, the pixel distortions and edge information are combined to quantify image quality.

To quantify the effect of ringing artifacts, the measure in [79] first uses a ringing region detection method [80] to extract the potential regions which are likely to be impaired by the ringing artifacts. For each individually detected ringing region, a ringing annoyance score is calculated by estimating the local visibility of ringing artifacts and comparing it with the local background activity. An overall ringing annoyance score is obtained by averaging the local annoyance scores over all ringing regions. The ringing region detection method [80] consists of two steps: extraction of edges relevant to ringing artifacts and detection of perceived ringing regions. For the edge extraction, a bilateral filter-based advanced edge detector [81] is adopted to extract the potential edges relevant to ringing regions, and the subsequent Canny edge detector [72] is applied on the filtered image to obtain the most relevant edges.

Most existing NR image quality measures are designed for specific distortions with prior knowledge of distortion type or distortion process, just like the image

quality measures developed for JPEG and JPEG2000 compression. Despite its intrinsic difficulty, several general-purpose NR image quality measures have been reported.

Moorthy *et al.* [82, 83] proposed a framework based on natural scene statistics [84] in the wavelet domain. The framework does not require any knowledge of the distortion process once it has been trained. The framework consists of a classifier based on a support vector machine [85] and a quality measure. When predicting image quality, the classifier gauges the probability of each distortion from a predefined distortion set. The quality measure provides a quality index for each distortion within the distortion set. An image quality score is computed as the probability-weighted summation of the measured probabilities and quality indexes of different distortions. In Moorthy's study, the predefined distortion set consists of JPEG compression, JPEG2000 compression, white noise, Gaussian Blur, and fast fading. With respect to the implementation of the component of quality measure, support vector regression [85] is utilized for the distortions of JPEG2000 compression, white noise, Gaussian blur, and fast fading, and the algorithm presented in [57] is employed for JPEG compression.

Another exploration towards the general-purpose goal is a blind image quality index developed by Saad *et al.* [86, 87]. Based on the statistics of DCT coefficients, the index first extracts features to construct a feature vector to represent an image. The constructed feature vector consists of DCT-based contrast, kurtosis of DCT coefficients histograms, and anisotropy in the DCT domain. Then a probabilistic prediction model, following the multivariate Gaussian distribution and the multivariate Laplacian distribution, is used to perform the task of image quality

prediction. Saad's quality index was tested with distortions of white noise, Gaussian blur, compression, fast fading channel distortions, and JPEG and JPEG2000 compression.

3.4 Validation of Objective Quality Measures

Validation is essential for the successful development of objective image quality measures. Since the goal of objective quality measures is to estimate perceived visual quality, the standard approach of validation is to compare the outputs of objective quality measures with the "ground truth" data of subjective quality scores provided by human observers for an extensive database of test images. In this section, two widely used approaches for subjective quality evaluation are introduced in Section 3.4.1, and the evaluation criteria recommended by the VQEG are reviewed in Section 3.4.2.

3.4.1 Subjective Quality Evaluation

Subjective quality evaluation is determined by complicated experiments which involve many aspects of human psychology and viewing conditions, such as the vision ability of observers, translation of human quality perception into ranking scores, stimulus content, display devices, ambient light levels, etc. Two methods are widely employed for subjective quality evaluation, namely single-stimulus continuous quality evaluation (single-stimulus method) and double-stimulus continuous quality scale (double-stimulus method). These two methods have been adopted as parts of an international standard by the International Telecommunications Union. In the single-stimulus method, subjects continuously indicate their impression of

quality on a linear quality scale that is divided into five equal segments. The five segments are not marked numerically and instead are labeled with adjectives of "Bad", "Poor", "Fair", "Good", and "Excellent" to serve as guides. The subjects are instructed to move a slider to any point on the scale that best reflects their impression of quality. The position of the slider is later converted into a quality score. The double-stimulus method is a form of discrimination-based method, in which a reference and a distorted image are presented one after the other in the same session with a small time segment of a few seconds each. Subjects evaluate the reference and the distorted image using sliders similar to those in the single-stimulus method. The difference between the scores of the reference and the distorted images gives the subjective distortion judgement. With subjective quality evaluation conducted using either the single- or double-stimulus method by multiple subjects, the resultant scores can be further averaged and processed to yield the *mean opinion score* (MOS) or *difference mean opinion score* (DMOS), two widely used forms of subjective quality scores. The standard deviation among individual scores may also be used to measure the consistency among different subjects.

Sheikh *et al.* performed an extensive subjective image quality assessment study and developed the LIVE image database [88] which has been widely used to evaluate the performances of image quality measures in the literature. In the LIVE image database, a total of 779 distorted images was evaluated by about two dozen judges, and the subjective image quality scores in terms of DMOS were obtained from about 25000 individual human quality judgments. The entire image database was derived from twenty-nine color images (mostly 768×512 pixels in size) with diverse image contents, including pictures of faces, people, animals, closeup shots, wide-angle shots, nature scenes, man-made objects, images with

Figure 3.1: Sample source images in the LIVE image database.

distinct foreground/background configurations, and images without any specific object of interest, etc. Figure 3.1 shows some sample source images of the LIVE image database. The distorted images are generated by a wide range of distortion types and distortion levels. Specifically, the image distortion type includes white

noise, Gaussian blur, JPEG and JPEG2000 compression, and transmission errors in the simulated fast fading rayleigh channel. The details of the five distortions are given below:

- White noise: A total of 145 distorted images was generated by adding white Gaussian noise of standard deviation σ_N to the RGB components of an image. The values of σ_N range from 0.012 to 2.0.

- Gaussian blur: A total of 145 distorted images was generated by filtering the RGB components using a circular-symmetric 2D Gaussian kernel of standard deviation σ_B pixels. The values of σ_B range from 0.42 to 15 pixels.

- JPEG compression: A total of 175 distorted images was generated by using JPEG compression at bit rates ranging from 0.15 *bits per pixel* (bpp) to 3.34 bpp.

- JPEG2000 compression: A total of 169 distorted images was generated by using JPEG2000 compression at bit rates ranging from 0.028 bpp to 3.150 bpp.

- Transmission errors: A total of 145 distorted images was generated by bit errors during the transmission of compressed JPEG2000 bitstream over a simulated wireless channel.

In the development of the LIVE image database, the single-stimulus method was employed for the tests in which the reference images were also evaluated in the same experimental session as the test images. After seven such independent sessions, a set of 50 images was collected from the individual sessions for a separate realignment experiment using a double-stimulus method to realign the obtained

quality scales. Based on the raw scores obtained in these tests, DMOS scores on a scale of 0-100 were computed after outlier detection, subject rejection, and other processings, with a larger resultant DMOS indicates poorer visual quality. More details of the LIVE image database can be found in [89].

3.4.2 Performance Evaluation Criteria

The VQEG [52] is a group of experts from various backgrounds and affiliations working in the field of video quality assessment. The group was formed in 1997 to validate and standardize objective measurement methods for video quality. The VQEG has established a systematic way to evaluate the performance of objective quality measures. According to the reports released by VQEG [90–92], the performance of objective quality measures can be evaluated by three criteria:

- Prediction accuracy - the ability to predict the subjective quality ratings with low error.
- Prediction monotonicity - the degree to which the model's predictions agree with the relative magnitudes of subjective quality ratings.
- Prediction consistency - the degree to which the model maintains prediction accuracy over the range of video test sequences, i.e., the model's response is robust with respect to a variety of video impairments.

For the fitting in scales between the objective and the subjective quality data, it is generally acceptable for an objective quality measure to predict subjective quality scores with the compensation by a mapping function. In the testing conducted by the VQEG, a nonlinear mapping between the outputs of objective quality measures, i.e., objective quality scores, and subjective quality scores was allowed. All

the performance evaluation criteria were computed after the compensation by the mapping function. According to the explanation given by the VQEG, the data of subjective quality scores are usually compressed at the ends of the rating scales, and it is not reasonable for objective quality measures to mimic this weakness of the subjective quality data. Different logistic functions have been used in the literature for the mapping. In this thesis, the mapping is constructed via the logistic function:

$$Q_p(Q_o) = \frac{\beta_1}{1 + \exp\left[-\beta_2(Q_o - \beta_3)\right]} + \beta_4 \tag{3.9}$$

where Q_o is the objective quality score calculated by quality measures, Q_p is the compensated objective quality score, which, in other words, is the predicted subjective quality score, and β_k with $k = 1, 2, 3, 4$ are parameters. The parameter β_k is determined through the nonlinear regression analysis [93] over the objective and the subjective quality data.

After the compensation of objective quality scores, different performance evaluation criteria can be calculated. The Pearson *correlation coefficient* (CC) on a scale of -1 to 1 is employed as a measure for the prediction accuracy criterion:

$$\mathrm{CC} = \frac{\sum_{i=1}^{N} (Q_s(i) - \bar{Q}_s)(Q_p(i) - \bar{Q}_p)}{\sqrt{\sum_{i=1}^{N} (Q_s(i) - \bar{Q}_s)^2} \sqrt{\sum_{i=1}^{N} (Q_p(i) - \bar{Q}_p)^2}} \tag{3.10}$$

where i denotes the index of image/video samples, $Q_s(i)$ denotes the subjective quality score, $Q_p(i)$ denotes the the predicted subjective quality score as computed in equation (3.9), $\bar{Q}_s$ and $\bar{Q}_p$ respectively denotes the average of $Q_s(i)$ and $Q_p(i)$,

and N denotes the total number of images/video samples under tests. The Spearman *rank order correlation coefficient* (ROCC) is employed as a measure for the prediction monotonicity criterion. To compute ROCC, $Q_s(i)$ and $Q_p(i)$ in equation (3.10) are converted to ranks and ROCC is calculated with the data of ranks. The computation of ROCC can be independent of the compensation by a mapping function since it is calculated based on the ranks of quality data. With each image/video sample deemed as a point, the criterion of prediction consistency can be measured by the *outlier ratio* (OR) which represents the ratio of outlier points to total points N:

$$\text{OR} = \frac{\text{total number of outliers}}{N} \tag{3.11}$$

An outlier is a point whose prediction error, as defined in equation (3.12), exceeds a threshold. Usually, the threshold is chosen as twice the standard deviation of individual subjective quality scores assigned to an image/video sample. The identification of an outlier point can be mathematically described by

$$\text{Perror}(i) = Q_s(i) - Q_p(i) \tag{3.12}$$

$$|\text{Perror}(i)| > 2\sigma_{\text{DMOS}}(i) \tag{3.13}$$

Besides, *root mean square error* (RMSE) also provides an informative performance measure to judge the relative performance between the quality measures under comparison. Based on the prediction error in equation (3.12), RMSE can be calculated by

$$\text{RMSE} = \sqrt{\frac{1}{N-d}\sum_{i=1}^{N}(\text{Perror}(i))^2} \tag{3.14}$$

where N denotes the total number of image/video samples, and d is the number of degrees of freedom of the mapping function. In the case of the mapping function (3.9), $d = 4$ since there are 4 coefficients.

Chapter 4

Kurtosis-Based No-Reference Image Quality Measures: JPEG2000

Kurtosis-based NR quality measures for JPEG2000 compressed images are presented in this chapter. The proposed image quality measures operate in the DCT domain of an image. Their implementation comprises two stages. In the first stage, kurtosis is computed with DCT coefficients of non-overlapping image blocks. In the second stage, a pooling approach is employed to combine all the kurtosis values into a single image quality score. This chapter is organized as follows. Section 4.1 provides the motivation of this study. Section 4.2 introduces the definition of kurtosis. The proposed computation of kurtosis in the DCT domain is detailed in Section 4.3, which includes the frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis. The working principle of kurtosis in image quality prediction is described in Section 4.4. Section 4.5 presents the kurtosis-based image quality measures. The effectiveness and limitations of the proposed NR image quality measures are discussed in Section 4.6. This chapter ends with a

summary in Section 4.7.

4.1 Motivation

Most existing NR image quality measures are developed with prior knowledge of the distortion. Fortunately, the distortion process is known in many applications, and the task of developing distortion-specific NR image quality measures is of practical importance. Compression techniques are widely used nowadays to reduce bandwidth for the storage and transmission of digital images. Due to the coarse quantization of DWT coefficients, blur and ringing artifacts are generally considered as the most significant distortions in JPEG2000 compressed images. These distortions are heavily dependent on image content and make NR quality assessment of JPEG2000 compressed images complicated. This chapter focuses on NR quality assessment of JPEG2000 compressed images.

As reviewed in Section 3.3, only a handful of reported NR quality measures are able to work with JPEG2000 compressed images. Although the researchers claimed good results with their respective techniques, these reported image quality measures have their limitations. The image quality measures presented in [70,71,73,74,79] are heavily dependent on reliable edge information. Given the decreased number of detected edges with increased blurring, the effectiveness of these image quality measures would inevitably decline when images are subject to severe distortion. A sophisticated edge detection method, such as a bilateral filter-based advanced edge detector used in [79] and the Canny edge detector in [71, 79], may help to alleviate the problem but is definitely not a solution. The implementation issue is another concern with respect to existing NR image quality measures. The

threshold operation associated with the image quality measures [70,71,73,74,79] for edge extraction is usually experience/image-dependent and would be inconvenient in practical applications. Similarly, the choice of projection axes in [73] is made in an *ad hoc* manner, and the parameters in [79] are determined empirically. Besides, all the measures in [73–75, 78, 79, 82, 86] require additional procedures and training data for model parameter determination. This imposes a constraint when data are limited and incurs inconvenience in practical applications. Additionally, an obvious shortcoming of the image quality measure in [58] is the absence of adequate performance validation.

The feasibility of kurtosis for NR image quality assessment has already been explored in the literature for sharpness/blur quantification. Kurtosis-based sharpness measures were developed either in the Fourier transform domain over the entire image [94] or in the DCT domain over the edge regions of an image [95]. Based on multivariate kurtosis, the sharpness measure reported in [94] fails to provide local sharpness information and solely assigns a score for the entire image. With image quality assessment still an open question, an effective quality measure is expected to provide not only global but also local quality information. Based on 2-D kurtosis of edge regions, the sharpness measure developed in [95] is subject to the limitations imposed by edge extraction as well, i.e., decreased extracted edges under severe distortion and inconvenience incurred by the associated threshold operation.

In this study, kurtosis is utilized in the DCT domain to blindly appraise the quality of JPEG2000 compressed images. Specifically, three NR image quality measures are proposed, which are based, respectively, on frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis. The proposed NR

image quality measures have these advantages: they are simple, they do not need to extract edges/features, they are parameter free, and their quality predictions are shown to be in good agreement with subjective quality scores.

4.2 Kurtosis

The statistical measure of kurtosis describes the departure of a probability distribution from the normal (Gaussian) distribution in shape. A brief definition of 1-D and 2-D kurtosis in the discrete case is provided below.

For a univariate random variable X with mean μ_X and finite moments up to at least the fourth, 1-D kurtosis is defined as:

$$K1 = \frac{m_4}{m_2^2} \tag{4.1}$$

where

$$m_k = E[(X - \mu_X)^k] = \sum\nolimits_i (x_i - \mu_X)^k P(x_i) \quad k = 2, 4 \tag{4.2}$$

$$\mu_X = E[X] = \sum\nolimits_i x_i P(x_i) \tag{4.3}$$

m_k is the kth central moment, $E[\cdot]$ denotes the expectation operator, and x_i is the sample of X whose *probability density function* (PDF) is represented by $P(X)$.

For a bivariate random vector $(X \quad Y)$ with PDF $P(X, Y)$, 2-D kurtosis can be calculated by

$$K2 = \frac{C_{4,0} + C_{0,4} + 2C_{2,2} + 4C_{1,1}(C_{1,1}C_{2,2} - C_{1,3} - C_{3,1})}{(1 - C_{1,1}^2)^2} \tag{4.4}$$

where

$$\begin{aligned} C_{k,l} &= \frac{E[(X-\mu_X)^k(Y-\mu_Y)^l]}{(E[(X-\mu_X)^2])^{k/2}(E[(Y-\mu_Y)^2])^{l/2}} \\ &= \frac{\sum_i \sum_j (x_i-\mu_X)^k (y_j-\mu_Y)^l P(x_i,y_j)}{(\sum_i (x_i-\mu_X)^2 P(x_i))^{k/2}(\sum_j (y_j-\mu_Y)^2 P(y_j))^{l/2}} \\ & \quad k,l = 0,1,...,4 \end{aligned} \tag{4.5}$$

$$P(x_i) = \sum\nolimits_j P(x_i, y_j) \tag{4.6}$$

$$P(y_j) = \sum\nolimits_i P(x_i, y_j) \tag{4.7}$$

Compared with the shape of the normal distribution, a distribution with excess kurtosis indicates a higher peak and/or smaller shoulder and/or heavier tail, while a distribution with less kurtosis indicates a flatter top and/or larger shoulder and/or lighter tail. More information pertaining to kurtosis can be found in [96].

4.3 Kurtosis in the Discrete Cosine Transform Domain

Suppose that $D(u,v)$ denotes the DCT coefficient of an $M \times M$ image block with $u,v = 0,1,...,M-1$. The following calculation of kurtosis is performed within the DCT coefficient matrices of non-overlapping $M \times M$ blocks over the entire image and exclude the DC coefficient $D(0,0)$. The DC coefficient is discarded in computation because its magnitude is usually much larger than the higher-frequency DCT coefficients and largely influences kurtosis to reflect the shape change of the distribution of the higher-frequency components. The basic principle of the following kurtosis calculations is to construct a 1-D or 2-D frequency distribution

in order of increasing frequency, and then compute 1-D or 2-D kurtosis with the normalized frequency distribution treated as a univariate or bivariate PDF.

4.3.1 Frequency Band-Based 1-D Kurtosis

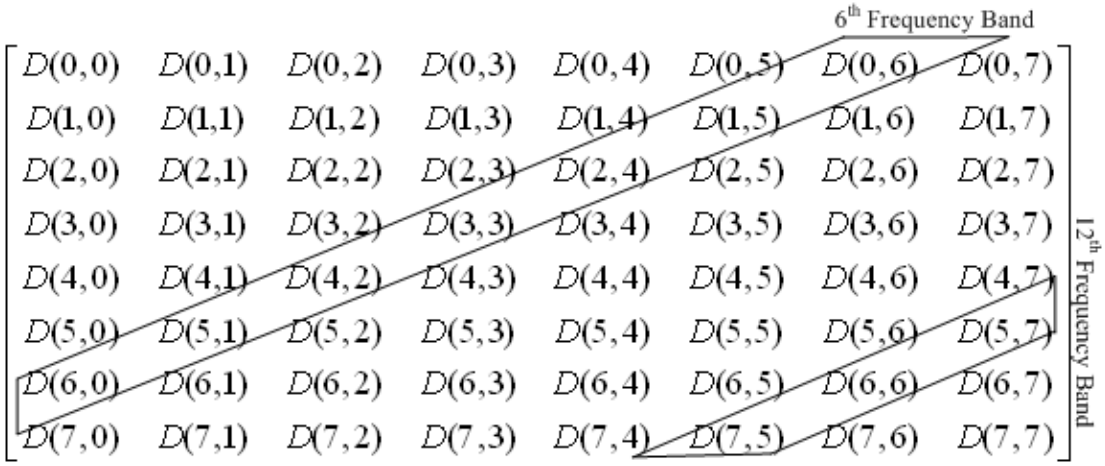


Figure 4.1: Example frequency bands in an 8×8 DCT coefficient matrix.

The DCT coefficients of an $M \times M$ image block are classified into different frequency bands. The DCT coefficients with the same frequency index $n = u + v$ are grouped into the nth band, which is composed of approximately equal radial frequency components [97]. In Figure 4.1, two frequency bands of an 8×8 DCT coefficient matrix are indicated as examples. With the increased frequency index n, the frequency band corresponds to the higher frequency content of the image block. Generalizing the definition given in [97], the average amplitude of the nth frequency band is represented by

$$B(n) = \frac{1}{N} \sum\nolimits_{u+v=n} |D(u,v)| \quad n = 1, 2, ..., 2M-2 \tag{4.8}$$

where

$$N = \begin{cases} n+1 & 1 \leq n < M \\ 2M-1-n & M \leq n \leq 2M-2 \end{cases} \tag{4.9}$$

A 1-D frequency distribution is thus constructed in the form of FB $= \{B(n) \quad n = 1, 2, ..., 2M-2\}$ where the frequency components are arranged in the sequence of increasing frequencies according to the increased frequency index n. Regarding the normalized FB as a univariate PDF $P(X)$, the frequency band-based 1-D kurtosis with the representation of $K1_{\mathrm{FB}}$ can be calculated according to equations (4.1)-(4.3).

4.3.2 Basis Function-Based 1-D Kurtosis

With an 8×8 image block as an example, a 1-D frequency distribution is constructed in equation (4.10) according to the following principles: (a) following a zigzag pattern, the DCT coefficients are basically reordered in the fashion of increasing frequency; and (b) the pair of DCT coefficients which are symmetrical with respect to the diagonal is grouped and represented by the maximum coefficient of the pair. The second principle is based on DCT basis functions, which are basically frequency components. The DCT transforms the input signal into a linear combination of weighted DCT basis functions. These weight values are referred to as DCT coefficients. With reference to the DCT basis functions shown in Figure 4.2, it can be observed that each pair of basis functions that are symmetrical with respect to the diagonal is basically the same frequency pattern distributed in the horizontal and vertical orientations. In fact, each DCT coefficient indicates how much the corresponding basis function is present in an image. It is therefore

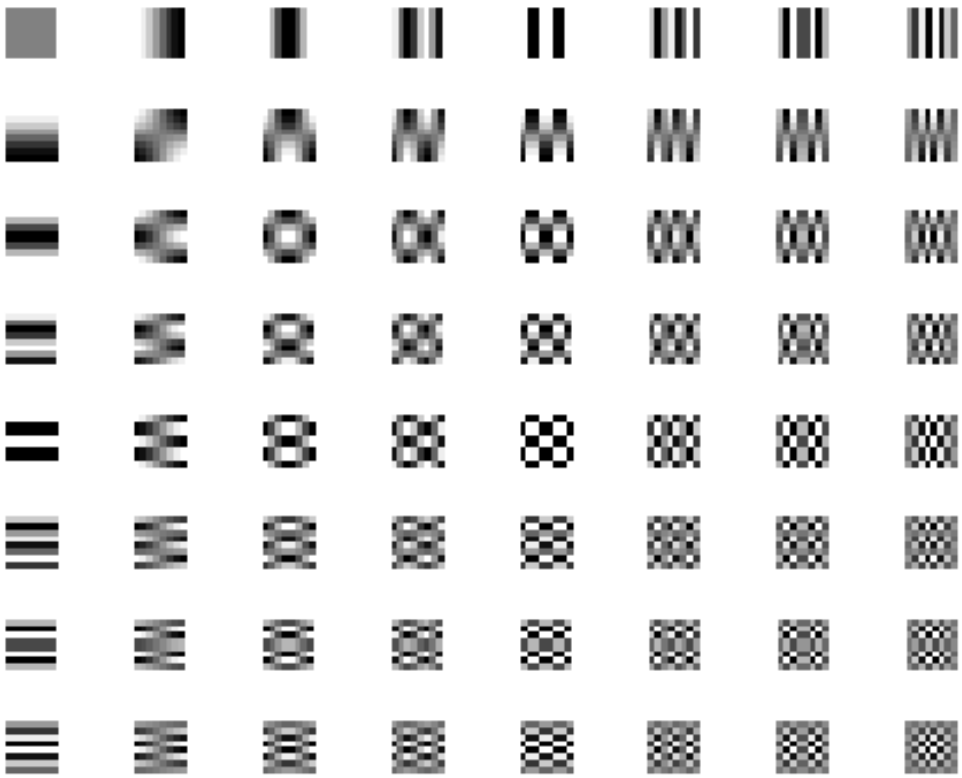


Figure 4.2: Visualization of DCT 8×8 basis functions.

concluded that the pair of DCT coefficients that are symmetrical about the diagonal have essentially the same frequency significance. Only the magnitude of the frequency components is taken into consideration here and the orientation information is not utilized in image quality prediction. Hence, each pair of symmetrical DCT coefficients is grouped together and the maximum coefficient of the pair is

employed in the construction of the frequency distribution BF:

$$
\begin{aligned}
\mathrm{BF} = \{ & \\
& \max(|D(0,1)|, |D(1,0)|), \max(|D(2,0)|, |D(0,2)|), |D(1,1)|, \\
& \max(|D(0,3)|, |D(3,0)|), \max(|D(1,2)|, |D(2,1)|), \max(|D(4,0)|, |D(0,4)|), \\
& \max(|D(3,1)|, |D(1,3)|), |D(2,2)|, \max(|D(0,5)|, |D(5,0)|), \\
& \max(|D(1,4)|, |D(4,1)|), \max(|D(2,3)|, |D(3,2)|), \max(|D(6,0)|, |D(0,6)|), \\
& \max(|D(5,1)|, |D(1,5)|), \max(|D(4,2)|, |D(2,4)|), |D(3,3)|, \\
& \max(|D(0,7)|, |D(7,0)|), \max(|D(1,6)|, |D(6,1)|), \max(|D(2,5)|, |D(5,2)|), \\
& \max(|D(3,4)|, |D(4,3)|), \max(|D(7,1)|, |D(1,7)|), \max(|D(6,2)|, |D(2,6)|), \\
& \max(|D(5,3)|, |D(3,5)|), |D(4,4)|, \max(|D(2,7)|, |D(7,2)|), \\
& \max(|D(3,6)|, |D(6,3)|), \max(|D(4,5)|, |D(5,4)|), \max(|D(7,3)|, |D(3,7)|), \\
& \max(|D(6,4)|, |D(4,6)|), |D(5,5)|, \max(|D(4,7)|, |D(7,4)|), \\
& \max(|D(5,6)|, |D(6,5)|), \max(|D(7,5)|, |D(5,7)|), |D(6,6)|, \\
& \max(|D(6,7)|, |D(7,6)|), |D(7,7)| \\
& \}
\end{aligned}
\tag{4.10}
$$

where $\max(\cdot)$ denotes the operation for the maximum. The maximum operation is employed because it is simple and preserves the principal frequency component. The validation of the use of the maximum operation is given in Section 4.6.2. With 35 frequency components constructed from an 8×8 DCT coefficient matrix, the 1-D frequency distribution BF in equation (4.10) is normalized and treated as a univariate PDF, $P(X)$. Then the basis function-based 1-D kurtosis with the representation of $K1_{\mathrm{BF}}$ can be calculated according to equations (4.1)-(4.3). The

computation of $K1_{\mathrm{BF}}$ can be easily applied to different block sizes.

4.3.3 2-D Kurtosis

The computation of 2-D kurtosis follows the algorithm presented in [95]. A 2-D frequency distribution of an $M \times M$ image block is constructed based on the magnitude of its DCT coefficient matrix with the representation of $\{|D(u,v)| \quad u = 0, ..., (M-1), \quad v = 0, ..., (M-1)\}$. With the normalized 2-D frequency distribution treated as a bivariate PDF, $P(X,Y)$, 2-D kurtosis with the representation of $K2$ can be calculated according to equations (4.4)-(4.7).

4.4 Working Principle of Kurtosis in Image Quality Prediction

Blur distortion occurs due to the attenuation of high spatial frequencies in an image. With the normalized frequency distribution treated as a PDF, the PDF of a more severely blurred image would have a higher peak, smaller shoulder, and heavier tail. In other words, kurtosis tends to increase with increased blurring. It is the basic principle underlying the image quality measures reported in [94, 95]. For a better understanding of kurtosis, we discuss how kurtosis provides a measure of blur in this section. The kurtosis under study includes the frequency-band based 1-D kurtosis $K1_{\mathrm{FB}}$, basis function-based 1-D kurtosis $K1_{\mathrm{BF}}$, and 2-D kurtosis $K2$.

Some illustrative examples are provided in Figure 4.3, where the original "Buildings" image (Figure 4.3a) has been subject to JPEG2000 compression (Figures

Table 4.1: $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ of the image blocks shown in Figures 4.3f-4.3h. $K1_{\mathrm{FB}}$: frequency-band based 1-D kurtosis. $K1_{\mathrm{BF}}$: basis function-based 1-D kurtosis. $K2$: 2-D kurtosis. BR (bpp): bit rate in JPEG2000 compression.

Kurtosis	Figure 4.3f BR = 0.85	Figure 4.3g BR = 0.40	Figure 4.3h BR = 0.20
$K1_{\mathrm{FB}}$	4.36	4.04	6.77
$K1_{\mathrm{BF}}$	2.57	2.56	5.82
$K2$	10.94	11.08	14.45

4.3b-4.3d) at different bit rates BR (in bpp). We divide the image into non-overlapping 8×8 blocks and take a closer look at a specific block selected from the edge of a roof (Figure 4.3e). The corresponding distorted blocks extracted from the distorted images in Figures 4.3b-4.3d are shown in Figures 4.3f-4.3h, respectively. $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ are computed for the 8×8 edge blocks shown in Figures 4.3f-4.3h and the results are listed in Table 4.1. The data in Table 4.1 suggest that $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ generally increase as blurring increases. It is noted that the two JPEG2000 compressed edge blocks shown in Figures 4.3f and 4.3g have similar kurtosis, be it $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, or $K2$, which actually coincides with the similar blur levels observed in these two blocks. With the normalized frequency distribution treated as a PDF, it is further investigated how kurtosis reflects the shape change of PDF as blurring increases in Figures 4.3f-4.3h. Here we only take a look at $K1_{\mathrm{FB}}$ and $K1_{\mathrm{BF}}$, whose PDFs are plotted in Figures 4.3i and 4.3j, respectively. The PDF tails are considered as those consecutive sample points whose probabilities are close to zero. It can be seen that the PDF with the largest $K1_{\mathrm{FB}} = 6.77$ in Figure 4.3i and the PDF with the largest $K1_{\mathrm{BF}} = 5.82$ in Figure 4.3j both have the highest peaks (both at the frequency index of 1), heaviest tails (the PDF in Figure 4.3i starts from the frequency index of 8 and the PDF in Figure 4.3j from 20), and

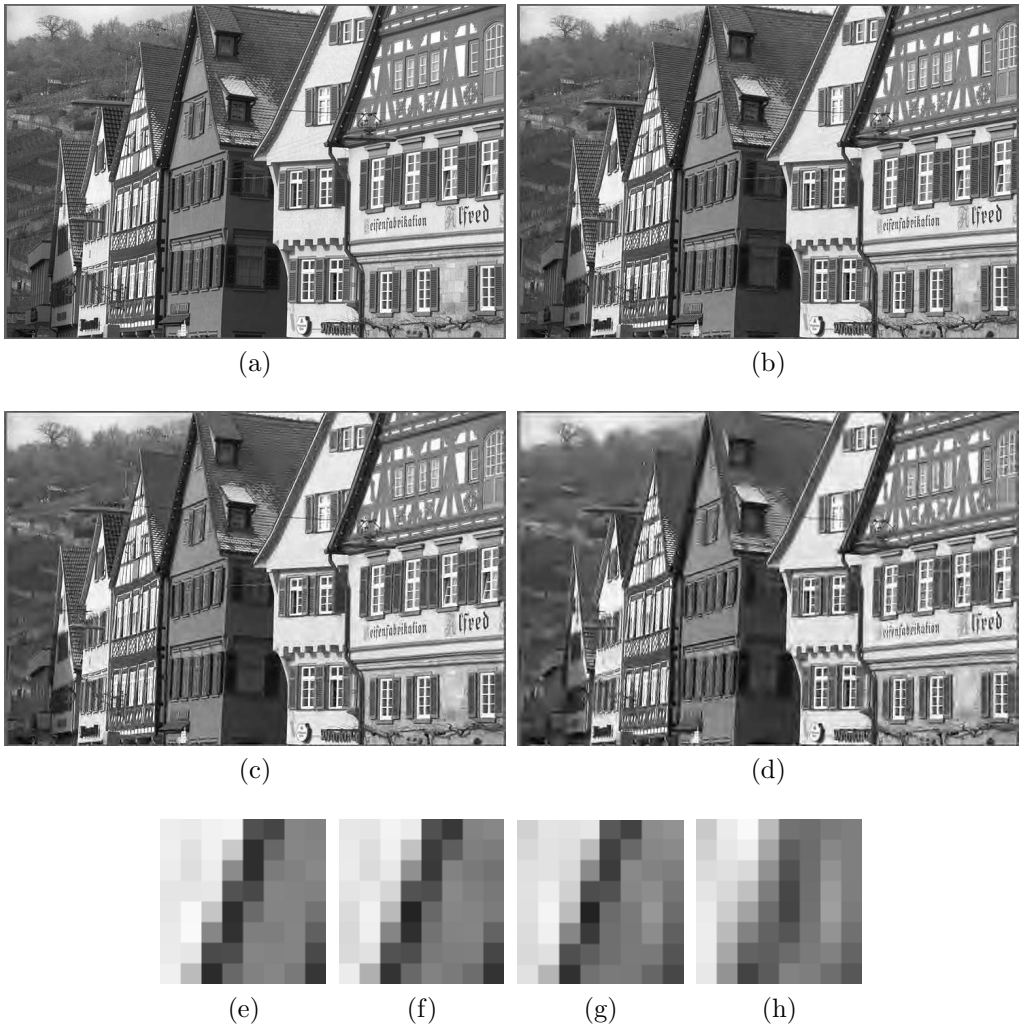
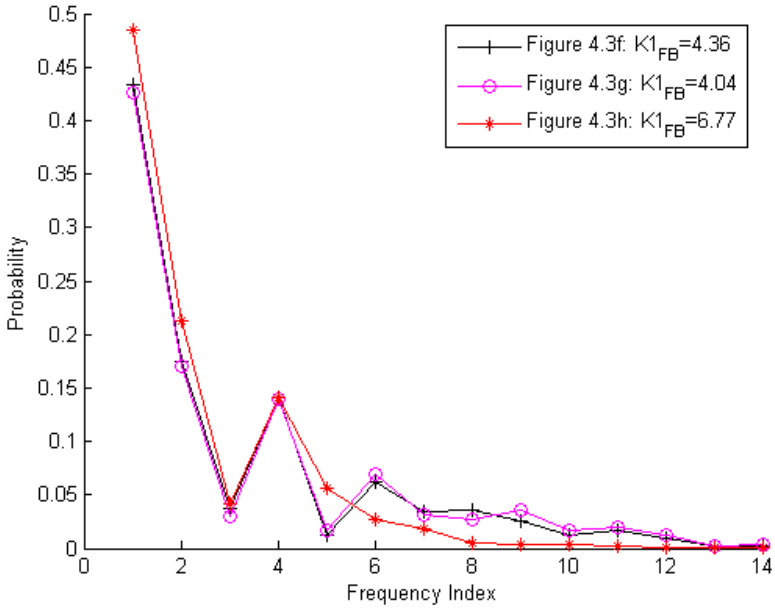
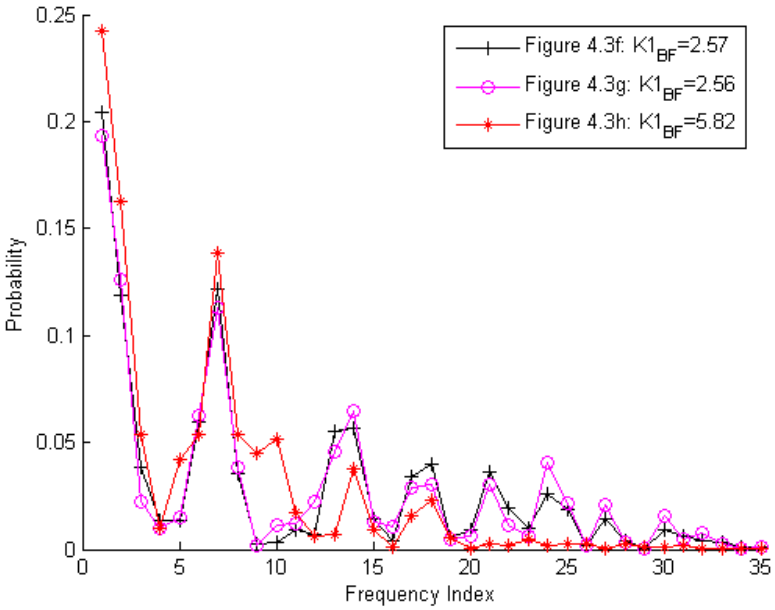


Figure 4.3: Illustration of the working principle of kurtosis in image quality prediction. (a) The original "Buildings" image. (b)-(d) JPEG2000 compressed images with bit rate BR $= 0.85, 0.40, 0.20$ bpp, respectively. (e) An 8×8 block selected from the edge of a roof. (f)-(h) The corresponding distorted blocks extracted from (b)-(d), respectively. (i) Frequency band-based PDF for (f)-(h). (j) Basis function-based PDF for (f)-(h).

smallest shoulders between peak and tail in their respective plots. As indicated by their similar kurtosis values, the PDF pair with $K1_{\mathrm{FB}} = 4.36$ and 4.04 in Figure 4.3i and the PDF pair with $K1_{\mathrm{BF}} = 2.57$ and 2.56 in Figure 4.3j both exhibit similar distributions in shape in their respective plots. Compared with the PDF with



(i)



(j)

Figure 4.3: Illustration of the working principle of kurtosis in image quality prediction. (a) The original "Buildings" image. (b)-(d) JPEG2000 compressed images with bit rate BR $= 0.85, 0.40, 0.20$ bpp, respectively. (e) An 8×8 block selected from the edge of a roof. (f)-(h) The corresponding distorted blocks extracted from (b)-(d), respectively. (i) Frequency band-based PDF for (f)-(h). (j) Basis function-based PDF for (f)-(h).

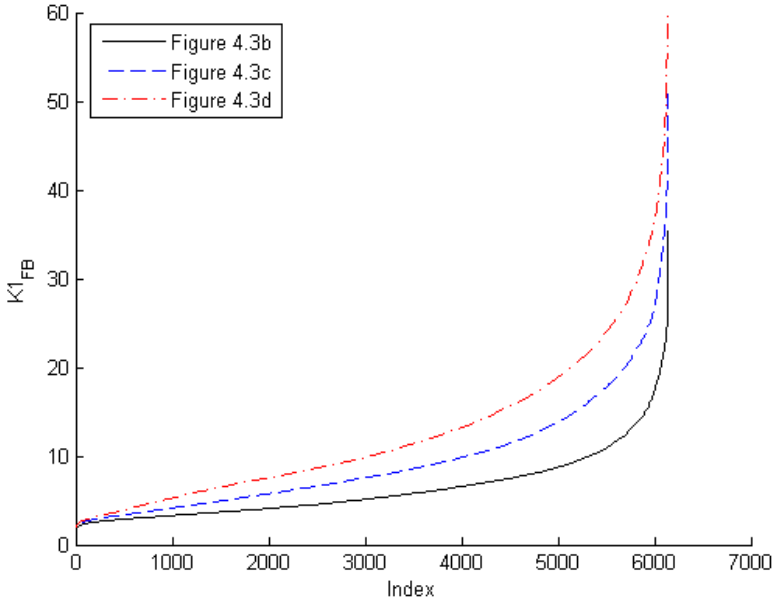
the largest kurtosis, both PDF pairs exhibit lower peaks (both at the frequency index of 1), lighter tails (the PDF pair in Figure 4.3i starts from the frequency index of 13 and the PDF pair in Figure 4.3j from 33), and larger shoulders in their respective plots.

4.5 Kurtosis-Based Image Quality Measure

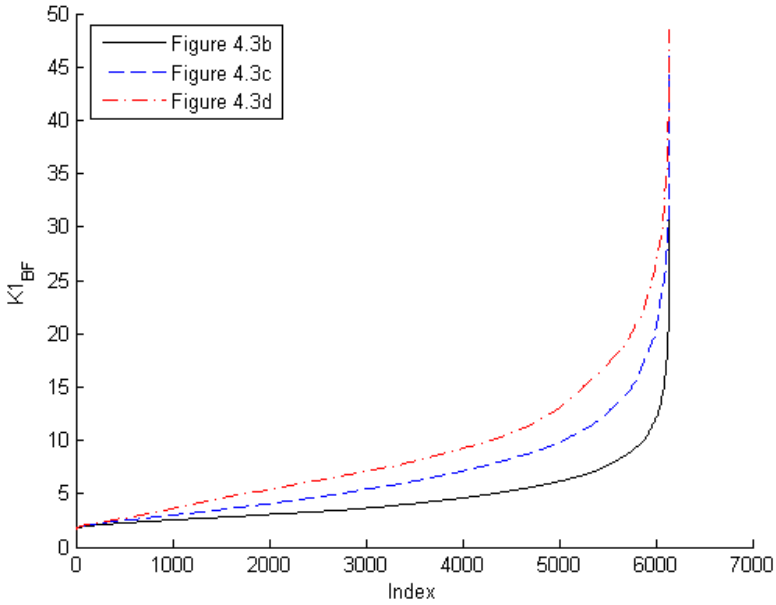
With kurtosis calculated within non-overlapping blocks over the entire image, the image quality score is computed in equation (4.11) as the average of local kurtosis after the difference operation with the median value.

$$Q_o = \frac{\sum_{t=1}^{N} |(q)_t - (q)_{\text{med}}|}{N} \tag{4.11}$$

where Q_o is the image quality score, t is the block index, N represents the total number of image blocks, $(q)_t$ indicates local kurtosis with $q = \{K1_{\text{FB}}, K1_{\text{BF}}, K2\}$, and $(q)_{\text{med}}$ denotes the median of $\{(q)_t \quad t = 1, 2, ..., N\}$. The motivation for the difference operation with the median in equation (4.11) is to reduce the over-sensitivity of kurtosis to the blurring in JPEG2000 compressed images. To illustrate this point, we sort the kurtosis values of the sample images shown in Figures 4.3b-4.3d in ascending order and plot them in Figure 4.4. The images shown in Figures 4.3b-4.3d are 768×512 pixels in size, with each image generating $96 \times 64 = 6144$ blocks after dividing into 8×8 blocks. The sorted kurtosis, be it $K1_{\text{FB}}$, $K1_{\text{BF}}$, or $K2$, exhibits gentle slopes at the beginning and middle portions of the plots, which comprise the major bulk (80-90%) of the kurtosis in quantity, while the steep slopes at the far end of the plots comprise only a small portion

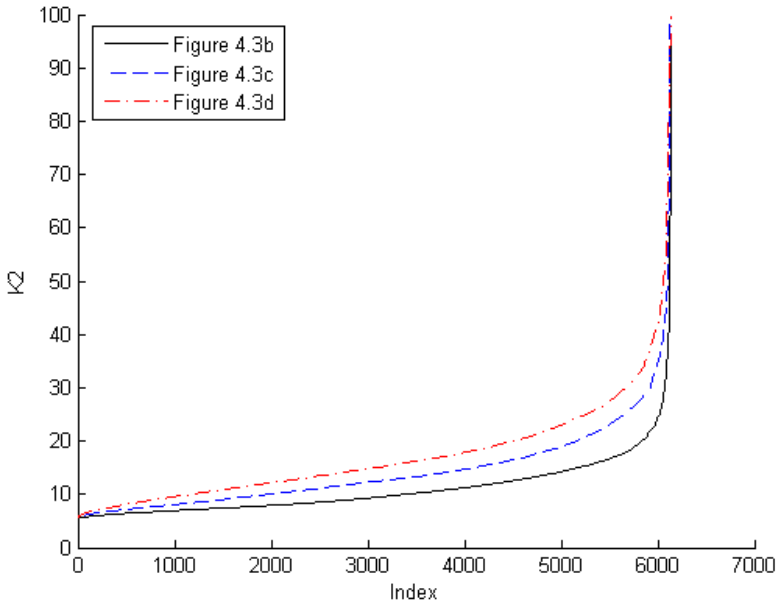


(a)



(b)

Figure 4.4: Plots of sorted kurtosis in ascending order for the images shown in Figures 4.3b-4.3d. (a) The frequency band-based 1-D kurtosis $K1_{\mathrm{FB}}$. (b) The basis function-based 1-D kurtosis $K1_{\mathrm{BF}}$. (c) 2-D kurtosis $K2$.



(c)

Figure 4.4: Plots of sorted kurtosis in ascending order for the images shown in Figures 4.3b-4.3d. (a) The frequency band-based 1-D kurtosis $K1_{\mathrm{FB}}$. (b) The basis function-based 1-D kurtosis $K1_{\mathrm{BF}}$. (c) 2-D kurtosis $K2$.

(<20%) in quantity. This observation is not limited to the sample images but occurs with most of the test images. If objective quality scores are simply computed as the spatial average of local kurtosis, we find those kurtosis values corresponding to steep slopes do not correlate well with the subjective quality scores provided by human observers. Therefore, the difference operation with the median is employed to reduce the effect of those over-sensitive kurtosis values. The pooling strategy expressed in equation (4.11) proves effective in the tests in Section 4.6.2 as well as the image quality measures reported in [45,46]. In fact, it is also possible to assign weights to local kurtosis for the computation of image quality scores depending on different applications. For instance, it is observed that different image regions may attract visual attention at different degrees. Based on visual importance, different weights therefore can be assigned to local kurtosis to determine their relative

contribution in the computation of image quality score [98, 99]. Since the major objective in this study is to design simple and effective image quality measures, we choose uniform weights.

4.6 Results and Discussion

The performances of the proposed image quality measures are evaluated with the JPEG2000 compressed images in the LIVE image database [88, 89], which has been widely employed for the performance evaluation of image quality measures. A total of 169 JPEG2000 compressed images was generated by the application of JPEG2000 compression at bit rates ranging from 0.028 to 3.150 bpp. The subjective quality scores are given in terms of the realigned DMOS on a scale of 0 to 100, with a larger DMOS indicating poorer visual quality. More details of the LIVE image database can be found in Section 3.4.1. In the following content, the performance of the proposed image quality measures is investigated through the kurtosis visualization, quantitative performance evaluation, and outlier analysis.

4.6.1 Visualization of Kurtosis

A set of intuitive examples is seen in Figure 4.5 where $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ are shown for the JPEG2000 compressed images of Figures 4.3b-4.3d. Here $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ are computed using a block size of 8. In comparison, DMOS together with $Q_{K1_{\mathrm{FB}}}$, $Q_{K1_{\mathrm{BF}}}$, and Q_{K2}, which represent the image quality score computed based on $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ respectively, are also given in Table 4.2. With the normalization operation performed for the same type of kurtosis, the brightness of the plots in each column of Figure 4.5 represents the relative magnitude, with

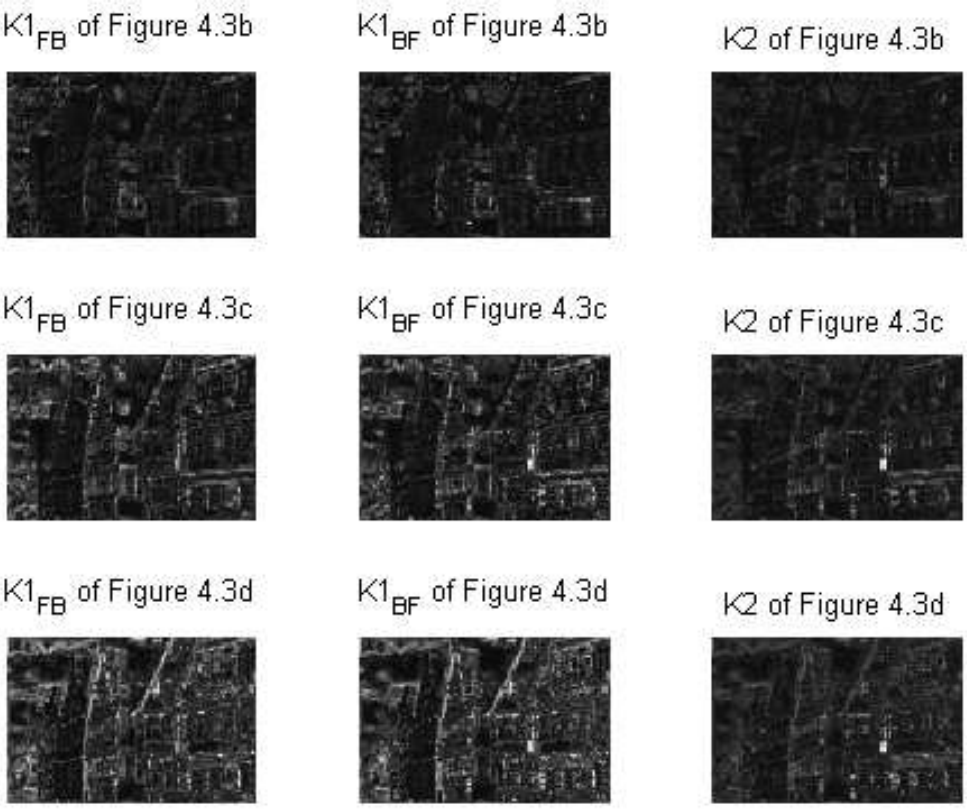


Figure 4.5: Visualization of kurtosis over the JPEG2000 compressed images shown in Figures 4.3b-4.3d. The plots in each row corresponds to one of the images shown in Figures 4.3b-4.3d, with the 1st row corresponding to Figure 4.3b, 2nd row to Figure 4.3c, and 3rd row to Figure 4.3d. The plots in the left column correspond to the frequency band-based 1-D kurtosis $K1_{\mathrm{FB}}$, the middle column to the basis function-based 1-D kurtosis $K1_{\mathrm{BF}}$, and the right column to 2-D kurtosis $K2$. The brightness of the plots obtained by the same type of kurtosis (in each column of the figure) represents the relative magnitude: a brighter pixel indicates a larger local kurtosis.

a brighter pixel indicating a larger local kurtosis. It is seen that the increased $K1_{\mathrm{FB}}$, $K1_{\mathrm{BF}}$, and $K2$ all reflect the increased blurring in JPEG2000 compressed images, as confirmed by the increased brightness together with increased image quality scores listed in Table 4.2. Furthermore, comparing the plots in each row of Figure 4.5, we can observe that (a) $K1_{\mathrm{FB}}$ and $K1_{\mathrm{BF}}$ appear more sensitive to

the blurring of JPEG2000 compressed images compared to $K2$, which is confirmed by the generally brighter plots supplied by $K1_{\mathrm{FB}}$ and $K1_{\mathrm{BF}}$; and (b) there is no visible difference between the plots provided by $K1_{\mathrm{FB}}$ and $K1_{\mathrm{BF}}$.

Table 4.2: $Q_{K1_{\mathrm{FB}}}$, $Q_{K1_{\mathrm{BF}}}$, and Q_{K2} together with the realigned DMOS of the images shown in Figures 4.3b-4.3d. $Q_{K1_{\mathrm{FB}}}$, $Q_{K1_{\mathrm{BF}}}$, and Q_{K2}: the image quality score computed based on the frequency-band based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis, respectively. BR (bpp): bit rate in JPEG2000 compression.

Image quality score	Figure 4.3b BR = 0.85 DMOS=31.28	Figure 4.3c BR = 0.40 DMOS=46.68	Figure 4.3d BR = 0.20 DMOS=63.64
$Q_{K1_{\mathrm{FB}}}$	2.43	4.17	5.38
$Q_{K1_{\mathrm{BF}}}$	1.66	3.09	3.88
Q_{K2}	3.13	4.61	5.36

4.6.2 Quantitative Performance Evaluation

Hereafter, K1-FB, K1-BF and K2-QM are used to denote the image quality measures developed using the frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis, respectively. Based on the ground truth data of the realigned DMOS provided by the LIVE image database, two representative FR image quality measures of PSNR and SSIM index [32] (introduced in Section 3.1) are employed as the benchmarks to study the performances of the proposed NR image quality measures. Some state-of-the-art NR image quality measures are also included for comparison. The performance evaluation of the proposed image quality measures follows the procedures recommended by the VQEG as detailed in Section 3.4.2.

4.6.2.1 Performances with Different Image Block Sizes

Table 4.3: Performances of K1-FB, K1-BF, and K2-QM with different image block sizes.

	K1-FB			
Block size	CC	RMSE	OR	ROCC
5	0.920	9.991	0.598	0.913
6	0.908	10.690	0.639	0.898
7	0.903	10.951	0.645	0.895
8	0.912	10.457	0.604	0.906
9	0.888	11.720	0.657	0.880
10	0.879	12.192	0.657	0.873
11	0.869	12.634	0.645	0.864
12	0.876	12.315	0.657	0.868
	K1-BF			
Block size	CC	RMSE	OR	ROCC
5	0.926	9.609	0.621	0.921
6	0.923	9.829	0.627	0.916
7	0.923	9.834	0.633	0.915
8	0.922	9.904	0.621	0.915
9	0.915	10.307	0.645	0.908
10	0.912	10.447	0.621	0.904
11	0.909	10.623	0.633	0.901
12	0.912	10.491	0.621	0.903
	K2-QM			
Block size	CC	RMSE	OR	ROCC
5	0.887	11.781	0.651	0.882
6	0.889	11.692	0.657	0.883
7	0.901	11.088	0.633	0.891
8	0.920	10.007	0.627	0.911
9	0.898	11.223	0.657	0.890
10	0.891	11.605	0.645	0.886
11	0.885	11.885	0.639	0.877
12	0.898	11.214	0.663	0.887

The performances of K1-FB, K1-BF, and K2-QM are studied with different image block sizes. The results with evaluation criteria are summarized in Table

4.3 where the block size ranges from 5 to 12 pixels. It is observed that all the proposed image quality measures achieve consistently close correlation with subjective quality scores when the block size varies across a large range. It is noted that the overall performance of K1-BF is the best for most evaluation criteria (higher CC and ROCC, and lower OR and RMSE), K1-FB is more advantageous than K2-QM when the block size is below 8 pixels, while K2-QM is generally superior to K1-FB when the block size is 8 pixels and above.

Table 4.4: Performances of K1-BF-mean with different image block sizes.

	K1-BF-mean			
Block size	CC	RMSE	OR	ROCC
5	0.923	9.834	0.580	0.918
6	0.919	10.067	0.615	0.912
7	0.919	10.091	0.615	0.911
8	0.918	10.138	0.592	0.911
9	0.912	10.501	0.633	0.902
10	0.908	10.713	0.639	0.898
11	0.905	10.848	0.627	0.894
12	0.908	10.709	0.615	0.897

The maximum operation in equation (4.10) in Section 4.3.2 merits further validation. For comparison purposes, the mean operation is employed to replace the maximum operation in K1-BF and the resultant tentative implementation is referred to as K1-BF-mean. Table 4.4 lists the results with evaluation criteria when K1-BF-mean is implemented using a block size ranging from 5 to 12 pixels. Two important observations may be noted. First, K1-BF-mean provides consistently good correlation coefficients not lower than 0.900 with subjective quality scores for different block sizes. Comparing Table 4.3 with Table 4.4, it can be seen that K1-BF-mean outperforms K1-FB and K2-QM for most evaluation criteria. The

effectiveness of K1-BF-mean further validates K1-BF since the only difference between K1-BF and K1-BF-mean is the different way of representing the symmetrical DCT coefficient pairs. Second, the comparison between Table 4.3 and Table 4.4 suggests that K1-BF outperforms K1-BF-mean for most evaluation criteria when their implementations are subject to a large range of block sizes. The better overall performance of K1-BF thus validates the maximum operation in equation (4.10). Apart from the simple mean and maximum operations, other more complicated operations may exist for the representation of symmetrical DCT coefficient pairs. The maximum operation is chosen in the design of K1-BF with the intention to achieve low complexity without sacrificing effectiveness.

4.6.2.2 Performance Comparisons of Image Quality Measures

Table 4.5 lists the results of evaluation criteria pertaining to all the image quality measures employed in the performance tests. The proposed image quality measures are implemented using a block size of 8 in the following tests since this block size is widely used. The negative sign of ROCC in Table 4.5 indicates an inversely proportional relationship between the subjective and objective quality scores. This

Table 4.5: Performance evaluation of the proposed NR image quality measures with the FR quality measures of PSNR and SSIM index as benchmarks. The proposed image quality measures are implemented using a block size of 8.

Quality measure	Type	CC	RMSE	OR	ROCC
PSNR	FR	0.900	11.151	0.615	-0.895
SSIM index	FR	0.967	6.551	0.473	-0.961
K1-FB	NR	0.912	10.457	0.604	0.906
K1-BF	NR	0.922	9.904	0.621	0.915
K2-QM	NR	0.920	10.007	0.627	0.911

does not occur with CC because of the compensation by nonlinear mapping (refer to Section 3.4.2). It means the magnitude of ROCC when ROCC is referred to in the following content. From the results in Table 4.5, PSNR proves an effective FR quality measure for JPEG2000 compressed images, and the FR quality measure of SSIM index performs better than the other quality measures being tested in terms of lowest RMSE, best prediction accuracy (highest CC), best prediction monotonicity (highest ROCC), and best prediction consistency (lowest OR). It is noted that K1-FB performs better than PSNR for all the evaluation criteria (higher CC and ROCC, and lower OR and RMSE); K1-BF and K2-QM perform better than PSNR in terms of lower RMSE, better prediction accuracy (higher CC), and better prediction monotonicity (higher ROCC), while PSNR is better than K1-BF and K2-QM in terms of better prediction consistency (lower OR).

Furthermore, the proposed image quality measures are compared with the state-of-the-art NR image quality measures. The performance tests for the image quality measures in [70,73–75,78,82,86] used JPEG2000 compressed images from the LIVE image database. All the results reported in [70,73–75,78,82,86] together with those obtained by the proposed image quality measures are listed in Table 4.6. Due to the different versions of the subjective quality data, different test procedures, and also limited information and results released in [70,73–75,78,82,86], the results in Table 4.6 provide an intuitive but instructive comparison. With respect to the proposed image quality measures, we would like to highlight the additional advantages that have not been revealed in Table 4.6: (a) the proposed image measures are simple; (b) they do not require any specific edge/feature extraction; (c) they are not subject to the limitations imposed by decreased edges/features under severe distortion; and (d) they are parameter free and avoid additional procedures and training

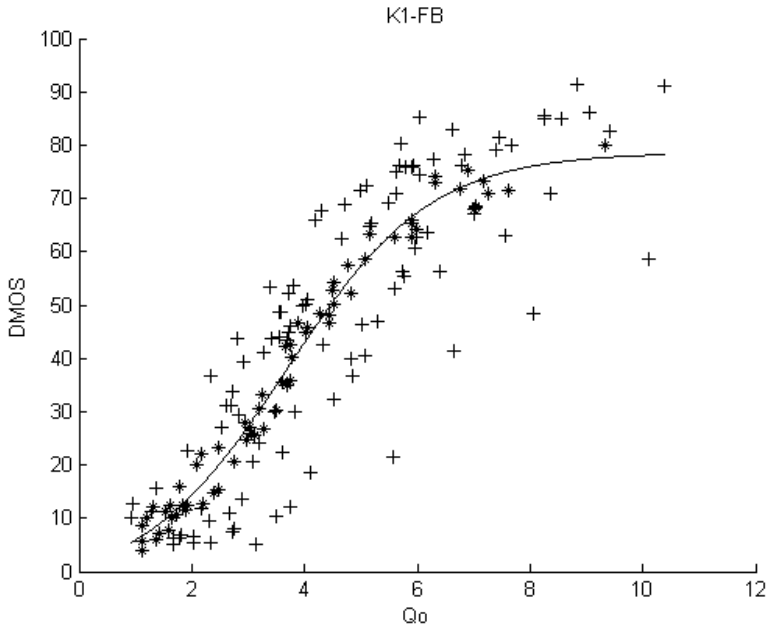
data for parameter determination. Based on these advantages together with the results listed in Table 4.6, it may be concluded that the overall performances of the proposed image quality measures are comparable to the state-of-the-art NR image quality measures.

Table 4.6: Performance comparison between the proposed NR image quality measures and the state-of-the-art NR image quality measures. The proposed image quality measures are implemented using a block size of 8.

Quality measure	CC	RMSE	OR	ROCC
[70]	0.86	N/A	N/A	0.84
[73]	0.857	N/A	N/A	N/A
[74]	0.896	N/A	N/A	0.872
[75]	0.93	N/A	N/A	N/A
[78]	0.93	9.20	0.0396	0.99
[82]	0.8086	14.8427	N/A	0.7995
[86]	N/A	N/A	N/A	0.9219
K1-FB	0.912	10.457	0.604	0.906
K1-BF	0.922	9.904	0.621	0.915
K2-QM	0.920	10.007	0.627	0.911

4.6.3 Outlier Analysis

The scatter plots of DMOS versus image quality scores computed by K1-FB, K1-BF, and K2-QM are shown in Figure 4.6. Each point in the figure, marked by asterisk or "+", represents one test image. With most points close to the fitted logistic curve (corresponding to equation (3.9)) in the scatter plots, the proposed image quality measures are observed to provide satisfactory prediction of DMOS for most test images. As described in Section 3.4.2, the outliers (denoted by "+" in Figure 4.6) are those data points whose prediction error exceeds twice the standard deviation of individual subjective quality scores. We take a close look at some

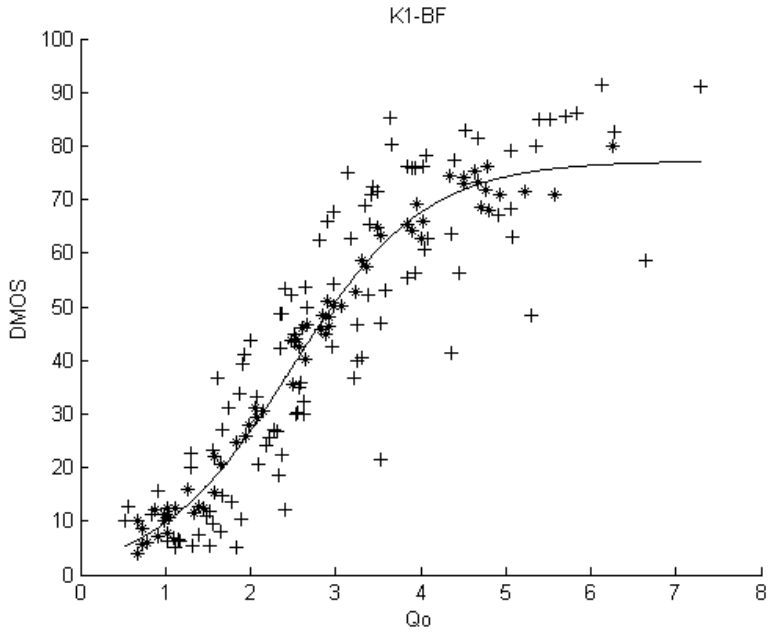


(a)

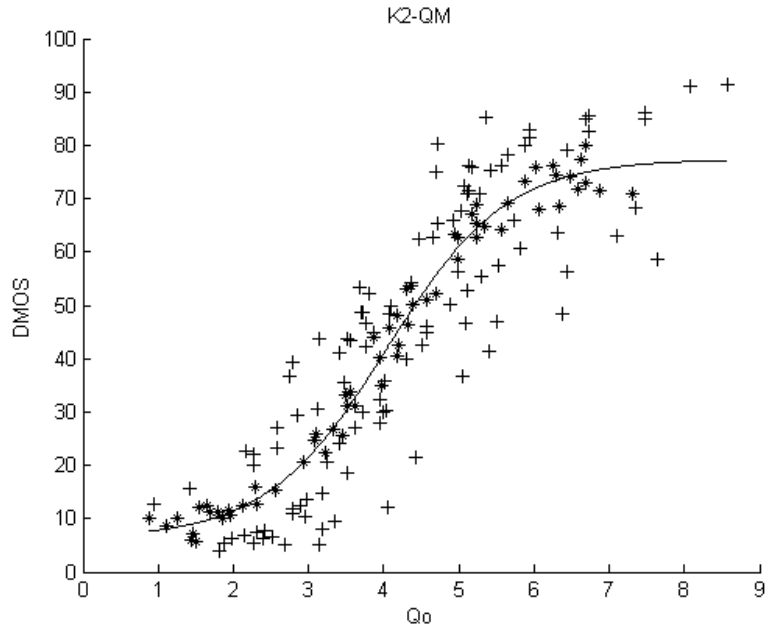
Figure 4.6: Scatter plot of DMOS versus image quality scores computed by K1-FB, K1-BF, and K2-QM. Each point, marked by asterisk or "+", represents one test image with "+" denoting outliers. The curve corresponds to the logistic function (3.9) with parameters fitted over dataset. (a) K1-FB. (b) K1-BF. (c) K2-QM.

representative outliers in this section.

It is found that the most significant outliers correspond to the compressed "Monarch" images, a sample image of which is shown in Figure 4.7a. It is evident that the out-of-focus blurred flowers in the background are due to a macro shot rather than distortion. Unfortunately, the proposed image quality measures are not able to recognize macro shots and under-predict image quality in this scenario. A similar situation arises from a close-up shot in the compressed "Parrots" images of which a sample image is shown in Figure 4.7b. Likewise, the blurred background leads to quality-under-prediction of the proposed image quality measures, which fail to realize that the blur in the background is actually not quality degradation. In fact, the issue uncovered here is not limited to the proposed image quality

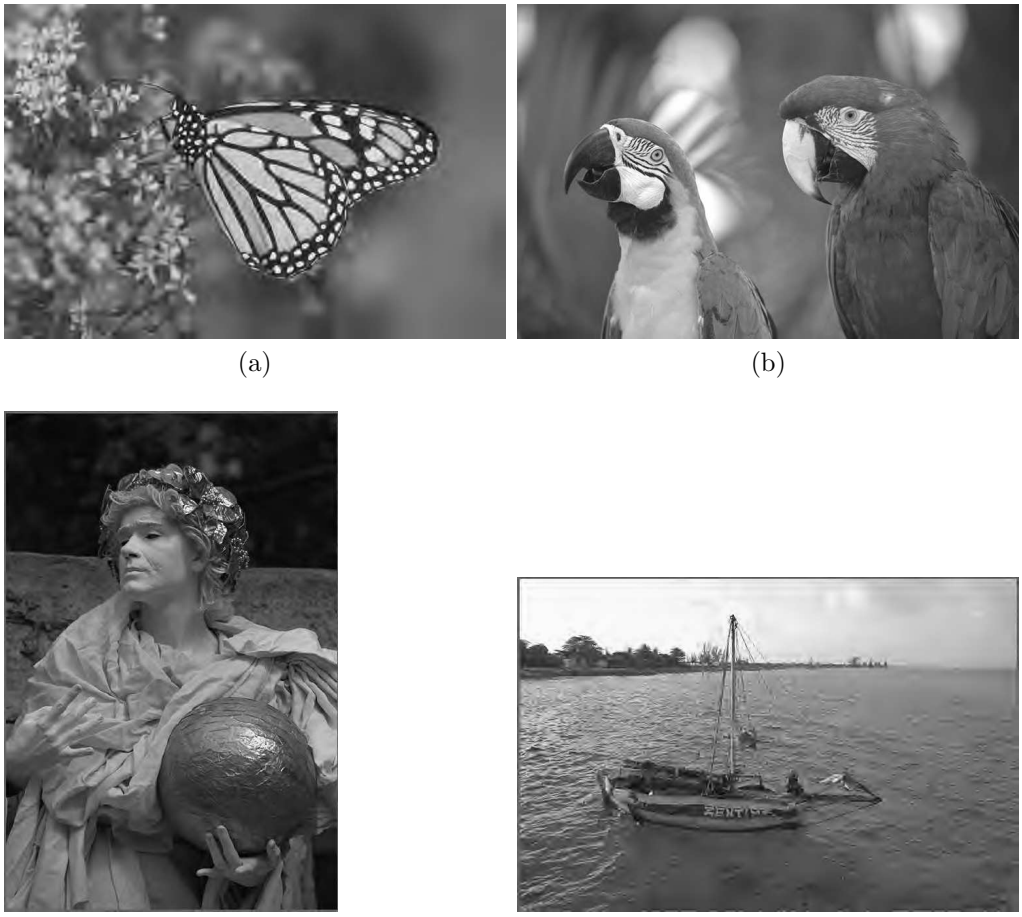


(b)



(c)

Figure 4.6: Scatter plot of DMOS versus image quality scores computed by K1-FB, K1-BF, and K2-QM. Each point, marked by asterisk or "+", represents one test image with "+" denoting outliers. The curve corresponds to the logistic function (3.9) with parameters fitted over dataset. (a) K1-FB. (b) K1-BF. (c) K2-QM.



(a) (b)

(c) (d)

Figure 4.7: Sample images corresponding to the outliers marked in Figure 4.6. (a) The "Monarch" image compressed to 0.1028 bpp. (b) The "Parrots" image compressed to 0.3819 bpp. (c) The "Statue" image compressed to 0.3777 bpp. (d) The "Sailing1" image compressed to 0.1157 bpp.

measures but is intrinsic to most NR image quality measures. With increased knowledge of human visual perception and especially the role of high-level cognitive understanding, a better emulation of the HVS would help to overcome this kind of outliers.

The compressed "Statue" image shown in Figure 4.7c exemplifies another important case of outliers. As presented in Section 4.5, the simple uniform weighting

(i.e., spatial average) is utilized to calculate image quality scores in this study. Nevertheless, the fact is that different image regions having diverse features/textures are of varying interest to the human eye and attract visual attention at different degrees. For instance, the almost blank background in the "Statue" image, which takes up around one third of the entire image, can hardly be appealing to human eyes, while the statue in the image is apparently a region of interest and undoubtedly attracts much more visual attention. However, the blank background and statue are assigned the same weights by uniform weighting. As previously mentioned in Section 4.5, a sophisticated pooling strategy based on visual importance may provide an alternative solution to this problem.

Additionally, a large number of outliers correspond to highly compressed images. A sample image is shown in Figure 4.7d, where the distortions of blur and ringing artifacts are clearly visible in the sea and the sailing boat. This kind of outliers arises mainly because the proposed quality measures predict image quality based on the distortion of blur alone and fail to take into account ringing artifacts. Perceived as annoying ripples and overshoots around sharp edges, the pseudo-structures of ringing artifacts easily mislead the proposed image quality measures to under-estimate the amount of blur and over-predict quality. It is reasonable to believe that the performance of the proposed image quality measures could be further improved with the incorporation of additional quantification of ringing artifacts.

4.7 Summary

Kurtosis-based NR image quality measures operating in the DCT domain are proposed for JPEG2000 compressed images in this chapter. Specifically, three kinds of kurtosis are discussed in this study, including frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, and 2-D kurtosis. In the frequency band-based 1-D kurtosis, the DCT coefficients with the same frequency index are grouped into a frequency band with an increased frequency index indicating higher frequency content. A 1-D frequency distribution thus can be constructed with frequency bands arranged in the sequence of increasing frequencies. Regarding the normalized 1-D frequency distribution as a univariate PDF, the frequency band-based 1-D kurtosis can be calculated. In the basis function-based 1-D kurtosis, the DCT coefficients are basically reordered following a zigzag pattern in the fashion of increasing frequency, and the pair of DCT coefficients which are symmetrical with respect to the diagonal is grouped and is represented by the maximum coefficient of the pair. A 1-D frequency distribution thus can be constructed from the resulting frequency components. Regarding the normalized 1-D frequency distribution as a univariate PDF, the basis function-based 1-D kurtosis can be calculated. In 2-D kurtosis, a 2-D frequency distribution can be constructed based on the magnitude of the DCT coefficient matrix. With the normalized 2-D frequency distribution treated as a bivariate PDF, 2-D kurtosis can be calculated. A final image quality score is computed as the average of local kurtosis, be it frequency band-based 1-D kurtosis, basis function-based 1-D kurtosis, or 2-D kurtosis, after the difference operation with the median kurtosis value. The proposed kurtosis-based NR image quality measures are simple; they do not need to extract edges/features; they are

parameter free; and their quality predictions are shown to be in good agreement with subjective quality scores.

Chapter 5

Pixel Activity-Based No-Reference Image Quality Measure: JPEG2000

A pixel activity-based NR quality measure for JPEG2000 compressed images is presented in this chapter. The proposed image quality measure works with the luminance component of the image. Its implementation comprises two stages. In the first stage, the local pixel activity is evaluated within non-overlapping image blocks, resulting in an activity map. In the second stage, a pooling approach is employed to combine the activity map into a single image quality score. This chapter is organized as follows. Section 5.1 provides the motivation of this study. Section 5.2 introduces the representation of pixel activity. Section 5.3 presents the structural content-weighted pooling approach. The effectiveness and limitations of the proposed image quality measure are discussed in Section 5.4. This chapter ends with a summary in Section 5.5.

5.1 Motivation

The proliferation of digital images in wide applications motivates the development of various image compression technologies. Lossy image compressions produce artifacts in the reconstructed images and lead to a drop in visual quality. The distortions created by JPEG2000 compression have attracted substantial research interest. Due to the coarse quantization of DWT coefficients, blur and ringing artifacts are created in JPEG2000 compressed images. The blurring distortion is characterized by the loss of image details, while ringing artifacts by ripples/oscillations around high-contrast edges. Both of them are dependent on image content and make NR quality assessment of JPEG2000 compressed images complicated. Despite its intrinsic difficulty, NR image quality assessment is desirable in practical applications where a reference image is expensive to obtain or simply not available. This chapter focuses on NR quality assessment of JPEG2000 compressed images.

A study of existing NR quality measures for JPEG2000 compressed images reveals the following issues. First, in the case of image quality measures based on the analysis of extracted edges/features [70, 71, 73, 74, 79], their effectiveness would inevitably decline when images are subject to severe distortion, given that the number of extracted edges/features decreases with increased blur. Second, the practical implementation is another concern. The threshold operation associated with the image quality measures [70, 71, 73, 74, 79] for structure/feature extraction is usually experience or image dependent. Besides, the parameters/procedures in the image quality measures [73, 79] are determined in an *ad hoc* manner. It would impose much inconvenience in practical applications. Third, many image quality measures require additional procedures and training data for parameter

determination [73–75,78,79,82,86]. This imposes an obvious restriction in practical applications especially when data are limited.

In this study, a NR quality measure is developed for JPEG2000 compressed images. Instead of extracting edges/features, the proposed quality measure predicts image quality based on the basic activity of general pixels. It therefore avoids the limitations imposed by edge/feature extraction of distorted images, i.e., decreased extracted edges/features under severe distortion and the associated inconvenient threshold operation. Besides, the proposed pooling approach does not involve any parameters and avoids additional procedures and training data for parameter determination. The proposed image quality measure exhibits satisfactory performance at reasonable computation load and easy implementation. It proves a NR quality measure of choice for JPEG2000 compressed images in practice.

5.2 Pixel Activity

With an image divided into non-overlapping blocks, pixel activity is evaluated within the individual image blocks independently.

5.2.1 Representation of Pixel Activity

Let $B^t(i,j)$ represent the intensity of a pixel (i,j) within an $M \times N$ image block, where i and j are indices of block-wide spatial locations with $1 \leq i \leq M$ and $1 \leq j \leq N$, and the superscript t denotes the block index. The horizontal difference at a pixel (i,j) is calculated by

$$B^t_{\mathrm{HD}}(i,j) = B^t(i,j) - B^t(i,j-1) \quad 1 \leq i \leq M \quad 2 \leq j \leq N \tag{5.1}$$

The sign change with respect to adjacent horizontal intensity differences can be indicated by

$$B^t_{\mathrm{HS}}(i,j) = B^t_{\mathrm{HD}}(i,j) \cdot B^t_{\mathrm{HD}}(i,j+1) \quad 1 \le i \le M \quad 2 \le j \le N-1 \tag{5.2}$$

According to the sign of $B^t_{\mathrm{HS}}(i,j)$, a pixel (i,j) within the effective block region of $1 \le i \le M$ and $2 \le j \le N-1$ can be classified into the following types:

- Horizontal *monotonic-changing* (MC) pixel if $B^t_{\mathrm{HS}}(i,j) > 0$
- Horizontal *zero-crossing* (ZC) pixel if $B^t_{\mathrm{HS}}(i,j) < 0$
- Inactive pixel if $B^t_{\mathrm{HS}}(i,j) = 0$

From the mathematical perspective, the horizontal MC pixel is related to monotonically-changing intensities in its horizontal neighborhood. For a group of consecutive horizontal MC pixels which starts at (i,j_1) and ends at (i,j_2) with $j_2 \ge j_1$, the horizontal MC spread is defined as the number of general pixels with monotonically-changing intensities:

$$\mathrm{MCS}^t_{\mathrm{H}}(k) = j_2 - j_1 + 3 \tag{5.3}$$

where k denotes the group index, and the constant 3 is derived by including the end pixels. Since each horizontal MC pixel is related to monotonically-changing intensities, the horizontal MC activity of the group is defined in equation (5.4) as the product of two components, namely, the horizontal MC spread and the number of MC pixels which is expressed by subtracting the end pixels from the horizontal MC spread:

$$\mathrm{MC}^t_{\mathrm{H}}(k) = \mathrm{MCS}^t_{\mathrm{H}}(k) \cdot (\mathrm{MCS}^t_{\mathrm{H}}(k) - 2) \tag{5.4}$$

The horizontal MC activity of an $M \times N$ block is obtained by adding up the corresponding activities from all the groups within the block:

$$\mathrm{MC}_{\mathrm{H}}^{t} = \frac{1}{MN} \sum\nolimits_{k} \mathrm{MC}_{\mathrm{H}}^{t}(k) \tag{5.5}$$

Different from the horizontal MC pixels, the horizontal ZC pixels deal with the local intensity extrema in their horizontal neighborhood. The horizontal ZC activity is defined as the number of horizontal ZC pixels within an $M \times N$ image block:

$$\mathrm{ZC}_{\mathrm{H}}^{t} = \sum_{i=1}^{M} \sum_{j=2}^{N-1} \mathrm{sign}\_(B_{\mathrm{HS}}^{t}(i,j)) \tag{5.6}$$

where

$$\mathrm{sign}\_(x) = \begin{cases} 1 & x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

Similarly, the vertical MC activity $\mathrm{MC}_{\mathrm{V}}^{t}$ and vertical ZC activity $\mathrm{ZC}_{\mathrm{V}}^{t}$ can be obtained by transposing the image block and using the same algorithm as described above.

Finally, the maximum of the horizontal and the vertical MC activity is taken as the overall MC activity of the image block, and the sum of horizontal and vertical ZC activity is taken as the overall ZC activity of the image block:

$$\begin{bmatrix} \mathrm{MC}^{t} \\ \mathrm{ZC}^{t} \end{bmatrix} = \begin{bmatrix} \max\left\{\mathrm{MC}_{\mathrm{H}}^{t}, \mathrm{MC}_{\mathrm{V}}^{t}\right\} \\ \mathrm{sum}\left\{\mathrm{ZC}_{\mathrm{H}}^{t}, \mathrm{ZC}_{\mathrm{V}}^{t}\right\} \end{bmatrix} \tag{5.8}$$

5.2.2 Meaning of Pixel Activity

Structures in an image are normally related to monotonically-changing intensities with strong structures characterized by large intensity changes and weak structures by relatively small changes. Based on the MC spread, either the horizontal or vertical MC activity is indirectly related to structures. It is intuitively evident that blur leads to increased MC spread in the horizontal direction or vertical direction or both directions. The horizontal and vertical MC activity thus turns out to provide a blur measure, with an increased value indicating increased blur. The maximum operation in equation (5.8) is intended for the sensitivity to distortion. Based on a similar principle, a feature of edge spread was proposed in [70,71] for blur measurement. Despite the different presentations in [70] and [71], edge spread is basically defined as the distance between the local intensity extrema closest to an edge pixel along its gradient orientation. The significant differences between edge spread and the proposed MC activity include the following respects. First, edge spread is measured at the extracted edges of relatively high strength, while MC activity is applied to general pixels and is related to structures of different strength levels. Second, edge spread is measured along the gradient orientation, while MC activity is measured along the horizontal and vertical directions. The general measurement along the horizontal and vertical directions provides a meaningful approximation given that the HVS pays much attention to structures along these two directions in comparison with other orientations [4]. Third, edge spread is measured across the image space, while MC activity is measured within non-overlapping image blocks with an intention to simplify operations and reduce complexity. Furthermore, we would like to point out the major advantages of the proposed MC activity over edge spread. First, edge extraction and the associated threshold selection are avoided.

Second, more stable performances can be expected since MC activity works with general pixels and is not subject to the limitation imposed by the decreased extracted edges with increased distortion.

Given that monotonically-changing intensities of structures normally end with local intensity extrema, either the horizontal or vertical ZC activity is indirectly related to structures with a large value indicating the structured image content. Here the structures from both the original image content and ringing artifacts are considered as structures of a JPEG2000 compressed image. Considering that ringing artifacts mainly occur around strong structures and produce zero crossings in their neighborhood as well, it is advantageous for the ZC activity in reflecting the structural information in a JPEG2000 compressed image. The summation operation in equation (5.8) is intended to reflect the full-scale structural content. Both qualitative and quantitative evaluations are used in Section 5.4 to validate the effectiveness of ZC activity as a structural content indicator of a JPEG2000 compressed image.

5.3 Structural Content-Weighted Pooling

A spatial pooling approach is required to combine the pixel activity across an image into a single image quality score. In the proposed image quality measure, a structural content-weighted pooling approach is utilized with the general expression given by

$$Q_o = \frac{\sum_t w_t q_t}{\sum_t w_t} \tag{5.9}$$

where Q_o is the image quality score, t is the index, q_t is the local quality measurement, and w_t is a structural content-based local weight with a large value

indicating substantial structural information. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, the basic idea of the proposed pooling approach is to determine the relative contribution of the local quality measurement according to its associated structural information.

The pooling approach with the mathematical expression of weighted average in equation (5.9) has been studied in the literature. In [100], Wang *et al.* evaluated several pooling approaches among which the quality-weighted and information-weighted approaches prove the good potential to improve image quality prediction. In the quality-weighted approach, the weight is defined as the monotonic function of the local quality/distortion measure. It is intended to largely weight the image regions of poor quality. A limitation of this approach is that the weight function is still not well defined, such as the simple power-law function employed in [100]. In the information-weighted approach, the weight is defined as the function of local information content with the intention to largely weight the image regions containing much information. However, this approach is currently developed only for FR image quality measures and has not yet been extended to NR image quality measures. In [98], Moorthy *et al.* presented two pooling approaches with the same mathematical expression of weighted average, namely, visual fixation-based weighting and quality-based weighting approaches. Both approaches define quality weight based on visual importance, with the former approach predicting visual fixation points and the latter approach using percentile scoring. Whatever the approach, the parameters are selected empirically and experimentally without clear rules. In contrast, the weight in the proposed pooling approach is independent of the local quality measurement and is determined by the structural information

instead. In the comparison between the reported and proposed pooling approaches, it is observed that the difference in mathematical expression is subtle whereas the concept difference is significant.

The specific expression of the proposed pooling approach in the developed image quality measure is given in equations (5.10) and (5.11) where the image quality score Q_o is calculated as the weighted average of non-zero MC activities with valid ZC activities as weights. The ZC activity is considered valid at the spatial locations where the local MC activity is non-zero.

$$Q_o = \frac{\sum_t \mathrm{ZC}^t \cdot \mathrm{MC}^t}{\sum_t \mathrm{ZC}^t \cdot \mathrm{sign}_+(\mathrm{MC}^t)} \tag{5.10}$$

where

$$\mathrm{sign}_+(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

A salient feature of this pooling approach is that it is parameterless, which avoids additional procedures and training data for parameter determination. According to equation (5.10), a local MC activity with a large ZC activity is given much emphasis in image quality score calculation. The MC activity is thus shown the basic component and the ZC activity is used to adjust the importance of MC activity in image quality score computation. Based on the analysis in Section 5.2.2, the MC activity provides a blur measure with increased value indicating increased blur, while the ZC activity provides a structural content indicator with a large value indicating substantial structural information. As a result, an increased image quality score is generated for the declining image quality. It is a somewhat simply strategy which does not include the specialized quantification of ringing

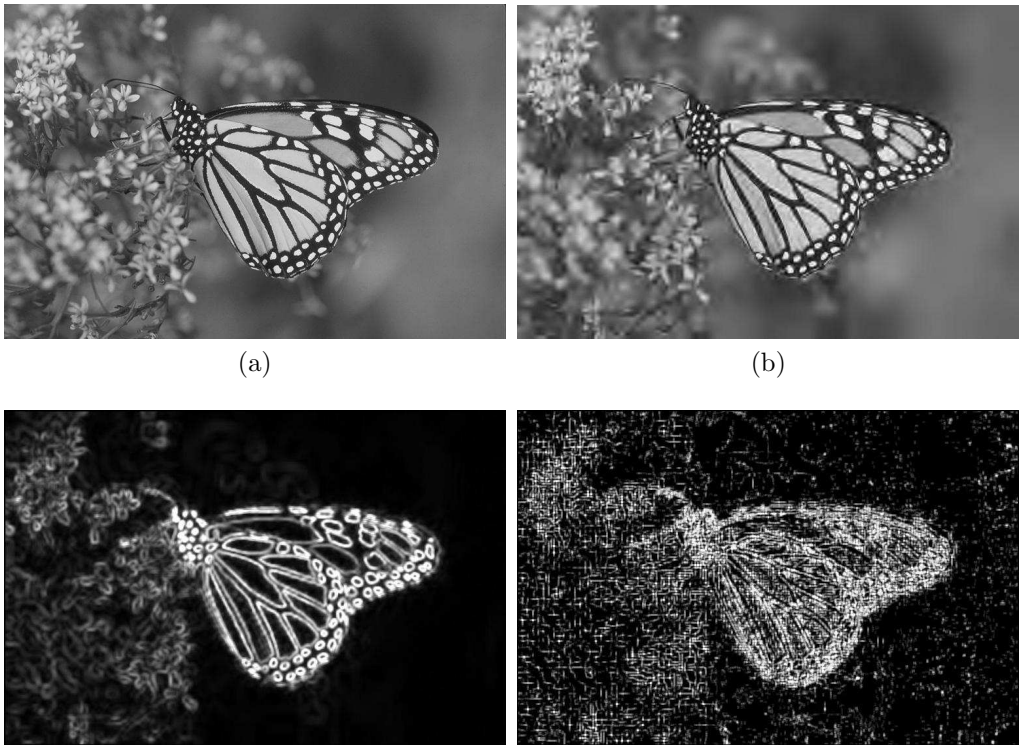
artifacts. However, the experimental results presented in Section 5.4 indicate that this simple strategy is sufficient for the proposed image quality measure to deliver satisfactory results.

5.4 Results and Discussion

The performance of the proposed image quality measure is evaluated with the JPEG2000 compressed images in the LIVE image database [88, 89], which has been widely employed for the performance evaluation of image quality measures. A total of 169 JPEG2000 compressed images was generated by the application of JPEG2000 compression at bit rates ranging from 0.028 to 3.150 bpp. The subjective quality scores are given in terms of DMOS on a scale of 0 to 100, with a larger DMOS indicating poorer visual quality. More details of the LIVE image database can be found in Section 3.4.1. In the following content, the performance of the proposed image quality measure is investigated through the visualization of ZC activity, study of the sensitivity to block size, quantitative performance evaluation, and outlier analysis.

5.4.1 Visualization of the Zero-Crossing Pixel Activity

Different expressions may be used for the structural content-based quality weight in the proposed pooling approach (introduced in Section 5.3). Unlike the proposed ZC activity, the feature of gradient captures structural information in terms of strength. To better illustrate the perceptual meaning of ZC activity, gradient is employed as a tentative expression of quality weight for comparison purposes. An intuitive comparison is provided in Figure 5.1 where the distributions of gradient



(a) (b)

(c) (d)

Figure 5.1: Visualization of the ZC activity over a sample JPEG2000 compressed image. (a) The original "Monarch" image. (b) Compressed to 0.1028 bpp. (c) The gradient map with gradient computed using the Sobel operator. (d) The ZC activity map with ZC activity computed using a 5×5 sliding window. Both the maps shown in (c) and (d) are normalized and contrast stretched for visibility, with a brighter pixel indicating a larger value.

and ZC activity are visualized over a sample JPEG2000 compressed image. The brightness of the normalized gradient and ZC activity maps in the figure represents the relative magnitude, with a brighter pixel indicates a relatively larger gradient or ZC activity. For satisfactory resolution, the gradient and ZC activity in Figure 5.1 are computed pixel-by-pixel across the image with gradient computed using the Sobel operator and ZC activity using a 5×5 sliding window. As shown in Figure 5.1b, blur and ringing artifacts are clearly visible in the compressed "Monarch" image, e.g., in the region of the wings. In the gradient map (Figure 5.1c), the

contours of flecks and stripes in the body and the wings are observed to have large gradient values. In contrast, ringing artifacts around the stripes in the wings have small gradient values because of their relatively low strength. The fact is that ringing artifacts disturb the natural behavior exhibited in an image and are usually a major attraction of visual attention in practice. A clear gap is consequently seen between the distributions of gradient and visual interest. In contrast to the clear details shown in the gradient map, the ZC activity map delivers relatively low resolution, as indicated by the quite general content in Figure 5.1d. It is mainly because the ZC activity operates solely over the sign of intensity changes. This leads to the variation among ZC activity values being not so obvious and the details not so clear. Despite the low resolution of the ZC activity map, it is clearly seen that the ZC activity effectively reflects the structural information in the image, and its distribution coincides with visual interest as a whole. In particular, the ZC activity is advantageous in capturing the weak structures of ringing artifacts, as confirmed by the notable ZC activity distributed at ringing artifacts around the stripes in the wings.

5.4.2 Quantitative Performance Evaluation

Based on the ground truth data of DMOS provided by the LIVE image database, two representative FR image quality measures of PSNR and SSIM index [32] (introduced in Section 3.1) are employed as the benchmarks to study the performance of the proposed NR image quality measure. Some state-of-the-art NR image quality measures are also included for comparison. The performance evaluation of the proposed image quality measure follows the procedures recommended by the VQEG as detailed in Section 3.4.2.

5.4.2.1 Sensitivity to Image Block Size

Table 5.1: Performances of the proposed image quality measure implemented using non-overlapping square image blocks of different sizes

Block size	CC	RMSE	OR	ROCC
5	0.923	9.849	0.568	0.910
6	0.930	9.375	0.562	0.922
7	0.930	9.384	0.556	0.922
8	0.937	8.901	0.533	0.931
9	0.932	9.270	0.538	0.925
10	0.932	9.275	0.556	0.925
11	0.931	9.296	0.544	0.926
12	0.932	9.264	0.562	0.928
13	0.931	9.313	0.556	0.926
14	0.930	9.407	0.568	0.925
15	0.930	9.362	0.562	0.926

The sensitivity of the proposed image quality measure to different image block sizes is studied in this section. A minimum block size of 5 is selected so as to effectively compute the MC and the ZC activity. With the proposed image quality measure implemented using non-overlapping square blocks of size ranging from 5 to 15 pixels, the results indicated by the evaluation criteria are summarized in Table 5.1. It can be seen that the proposed image quality measure achieves consistently close correlation with subjective quality scores and remarkably stable performance when the processing block size is subject to a wide range.

Figure 5.2 shows the change in processing time with the proposed image quality measure implemented using different image block sizes. The data in Figure 5.2 are the average processing time over an image of 768×512 pixels with normal computer configuration. The plot in Figure 5.2 indicates that the processing time significantly decreases from around 3 seconds to less than 0.5 seconds when block

size increases from 5 to 15 pixels.

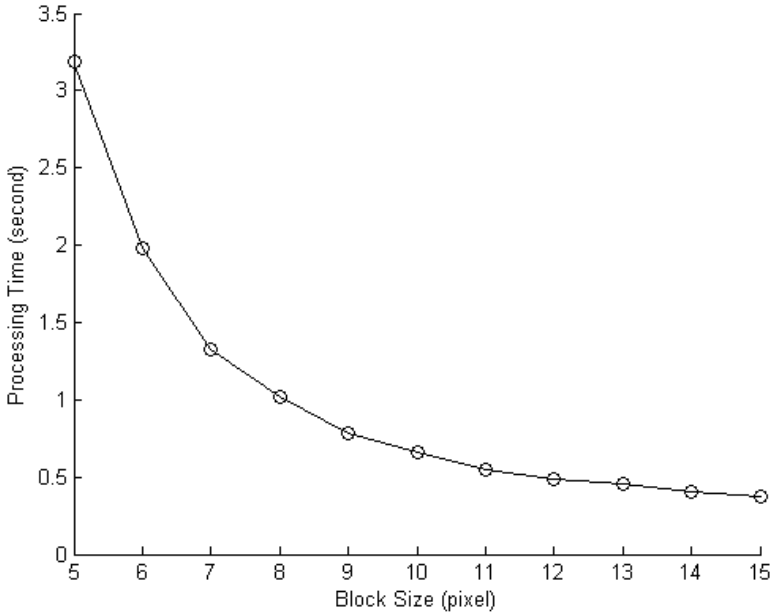


Figure 5.2: Change in processing time with the proposed image quality measure implemented using non-overlapping square image blocks of different sizes. The time is the average processing time over an image of 768×512 pixels in size.

From the results listed in Table 5.1 and the plot shown in Figure 5.2, it is observed that the choice of block size, assuming within a reasonable range, largely influences processing time but does not produce obvious differences in results. A common image block size of 8×8 is employed to implement the proposed image quality measure in the following tests.

5.4.2.2 Performance Comparisons of Image Quality Measures

The results of evaluation criteria pertaining to PSNR, SSIM index, and the proposed image quality measure are provided in Table 5.2. The negative value of ROCC in the table indicates an inversely proportional relationship between DMOS and image quality scores computed by quality measures. It does not happen to

CC because of the compensation by nonlinear mapping (refer to Section 3.4.2). It means the magnitude of ROCC when ROCC is referred to in the following content. As shown in Table 5.2, the performance of the proposed NR image quality measure is better than the FR quality measure of PSNR and is comparable to the FR quality measure of SSIM index. The performance of the proposed image quality measure is actually good for such a NR image quality measure.

Table 5.2: Performance evaluation of the proposed NR image quality measure with the FR quality measures of PSNR and SSIM index as benchmarks. The proposed image quality measure is implemented using 8×8 image blocks.

Quality measure	CC	RMSE	OR	ROCC
PSNR	0.900	11.151	0.615	-0.895
SSIM index	0.967	6.551	0.473	-0.961
Proposed measure	0.937	8.901	0.533	0.931

Table 5.3: Performance comparison between the proposed NR image quality measure and the state-of-the-art NR image quality measures. The proposed image quality measure is implemented using 8×8 image blocks.

Quality measure	CC	RMSE	OR	ROCC
[70]	0.86	N/A	N/A	0.84
[73]	0.857	N/A	N/A	N/A
[74]	0.896	N/A	N/A	0.872
[75]	0.93	N/A	N/A	N/A
[78]	0.93	9.20	0.0396	0.99
[82]	0.8086	14.8427	N/A	0.7995
[86]	N/A	N/A	N/A	0.9219
Proposed measure	0.937	8.901	0.533	0.931

The proposed image quality measure is further compared with those reported NR image quality measures that are applicable to JPEG2000 compressed images. The performance tests for the image quality measures in [70, 73–75, 78, 82, 86] used

JPEG2000 compressed images from the LIVE image database. All the results reported in [70,73–75,78,82,86] together with those obtained by the proposed image quality measure are listed in Table 5.3. Due to the different versions of the subjective quality data, different test procedures, and also limited information and results released in [70, 73–75, 78, 82, 86], the results in Table 5.3 provide an intuitive but instructive comparison. It can be seen from the table that the proposed image quality measure is competitive with the state-of-the-art NR image quality measures. Besides, the proposed image quality measure takes around 1 second to process an image using 8×8 blocks, and the processing time can be further reduced with a larger block size. As an intuitive comparison, the model in [75] was reported around 20 seconds and even 5 seconds with a simplified model to process an image from the same image database, which had not yet taken into consideration the time of the training procedure to determine its parameters. Apparently, the proposed image quality measure is computationally efficient. Moreover, the proposed image quality measure exhibits clear advantages in the following respects: it avoids structure/feature extraction of distorted images, and it avoids additional procedures and training data for parameter determination.

5.4.2.3 Quantitative Validation of the Zero-Crossing Pixel Activity

Besides the illustration in Section 5.4.1, the validity of ZC activity merits further quantitative evaluation. The proposed image quality measure is compared with a tentative implementation based on the feature of gradient instead of ZC activity. Specifically, gradient is computed at each pixel using the Sobel operator, and the gradient values within each processing block are added up to replace the role of ZC activity for the image quality score calculation in equation (5.10). The results

of evaluation criteria with respect to the proposed image quality measure and the tentative implementation are listed in Table 5.4. It is observed that the tentative implementation itself delivers good results that even outperform those of PSNR as listed in Table 5.2. The good performance of the tentative implementation indirectly validates the proposed pooling approach since the only difference between the proposed image quality measure and the tentative implementation is the different expressions of the structural content-based quality weight. Furthermore, the proposed image quality measure exhibits clear superiority over the tentative implementation according to each evaluation criterion. Based on the qualitative illustration in Section 5.4.1 and the quantitative evaluation here, we may conclude that the effectiveness of ZC activity as a structural content-based quality weight for JPEG2000 compressed images is fully validated.

Table 5.4: Performance comparison between the proposed image quality measure and the tentative implementation. Both of them are implemented using 8×8 image blocks.

Quality measure	CC	RMSE	OR	ROCC
Tentative implementation	0.911	10.515	0.609	0.908
Proposed Measure	0.937	8.901	0.533	0.931

5.4.3 Outlier Analysis

The scatter plot of DMOS versus image quality scores computed by the proposed image quality measure is presented in Figure 5.3. Each point in the figure, marked by asterisk or "+", represents one test image. With most points close to the fitted logistic curve (corresponding to equation (3.9)) in the scatter plots, the proposed image quality measure is observed to provide satisfactory prediction of DMOS for

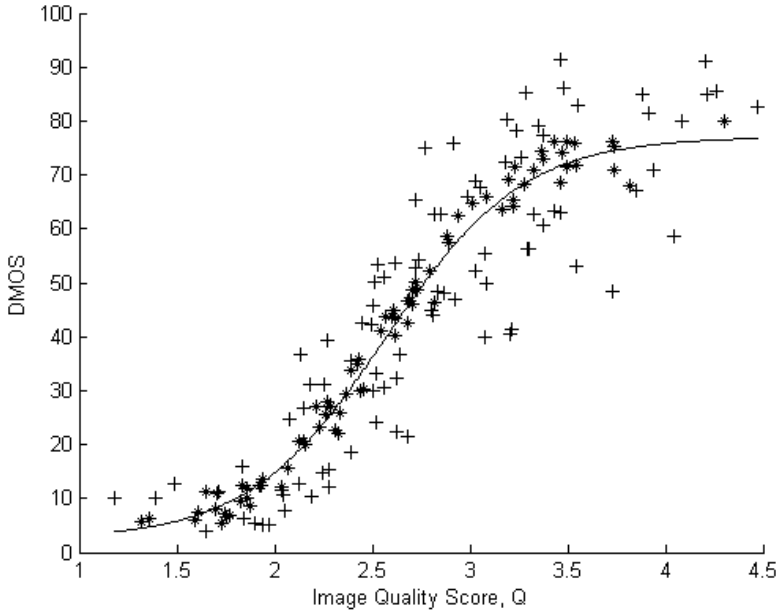
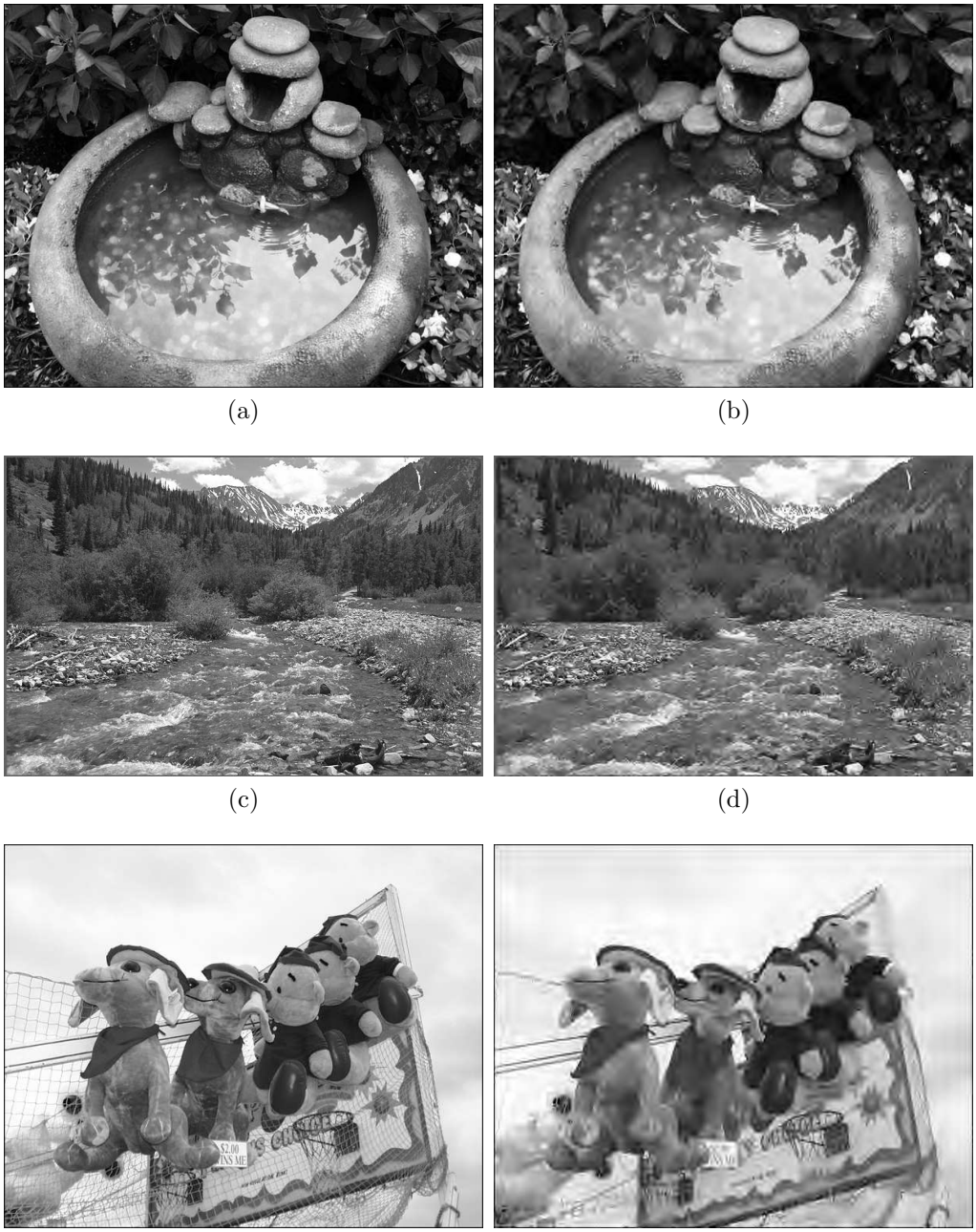


Figure 5.3: Scatter plot of DMOS versus image quality scores computed by the proposed image quality measure. Each point, marked by asterisk or "+", represents one test image with "+" denoting outliers. The curve corresponds to the logistic function (3.9) with parameters fitted over dataset.

most test images. As described in Section 3.4.2, the outliers (denoted by "+" in Figure 5.3) are those data points whose prediction error exceeds twice the standard deviation of individual subjective quality scores. We take a close look at some representative outliers in this section.

The most significant outliers correspond to the compressed "Monarch" images, a sample image of which is shown in Figure 5.1b. Note that the blur of flowers in the background is largely due to a close-up shot with very short depth-of-field rather than distortion. Given that the HVS does not consider the out-of-focus background as quality degradation whereas the proposed image quality measure can not recognize this, the proposed quality measure under-predicts image quality in this case. The problem of short depth-of-field was also reported in [70] and is present in most NR image quality measures.



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

Figure 5.4: Sample images corresponding to the outlier marked in Figure 5.3. (a) The original "Coins in fountain" image. (b) Compressed to 0.3285 bpp. (c) The original "Stream" image. (d) Compressed to 0.1920 bpp. (e) The original "Carnival dolls" image. (f) Compressed to 0.1235 bpp.

Another important case of outliers corresponds to the compressed "Coins in fountain" images, a sample image of which is shown in Figure 5.4b. It can be seen that the blur of reflection and coins in the water is essentially the real image content rather than distortion. However, the proposed image quality measure is not able to distinguish blur between distortion and real image content and under-predicts image quality in this scenario. In one form or another, this problem is intrinsic to most NR image quality measures.

The compressed "Stream" image shown in Figure 5.4d exemplifies another typical case of outliers. It is noted that the stones at the banks are a kind of quasi-texture where the visible distortion is not uniform with the distortion in other regions. The proposed image quality measure does not take into account the masking property of the HVS (Section 2.2) and over-predicts image quality in this case. The problem of texture is reported in [75] as well.

Another representative case of outliers corresponds to the compressed "Carnival dolls" images, a sample image of which is shown in Figure 5.4f. Given that the ZC activity only operates over the sign of intensity changes, the detailed structures of web are given the same significance with the major structures of dolls in Figure 5.4f. Since the structures of web are nearly totally missing, it should come as little surprise that the proposed quality measure under-predicts image quality in this case. This case of outlier, in large part, results from the simple structural content indicator of ZC activity in the proposed image quality measure. A sophisticated structural content indicator is advantageous in overcoming this problem. However, the major objective of this study is to design simple and effective image quality measures, so we choose the ZC activity.

5.5 Summary

A pixel activity-based NR image quality measure is developed for JPEG2000 compressed images in this chapter. Instead of extracting edges/features, the proposed quality measure predicts image quality based on the basic activity of general pixels and avoids the issues imposed by edge/feature extraction of distorted images, i.e., decreased extracted edges/features with increased distortion and inconvenience incurred by the associated threshold operation. In this study, the pixel activity is expressed in terms of the MC activity and the ZC activity within non-overlapping image blocks. The MC activity proves related to structures in an image with increased values indicating increased blurring; and the ZC activity turns out to be also related to structures with a large value indicating structured image content. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, a structural content-weighted pooling approach is proposed with the basic idea to determine the contribution of the local quality measurement according to its associated structural information. Specifically, the image quality score in the proposed quality measure is calculated as the weighted average of non-zero MC activities with valid ZC activities as quality weight. The pooling approach is parameterless and avoids additional procedures and training data for parameter determination. With satisfactory performance and ease of implementation, the proposed NR image quality measure proves an effective alternative to those complicated quality measures for JPEG2000 compressed images.

Chapter 6

Structural Activity-Based Framework for No-Reference Image Quality Assessment

Under the assumption that human visual perception is highly sensitive to the structural information in a scene, a concept of *structural activity* (SA) together with a model of an SA measure is proposed as a new framework for NR image quality assessment in this chapter. In this framework, image quality is estimated by quantifying SA information of different visual significance in an image. This chapter is organized as follows. Section 6.1 provides the motivation of this study. Section 6.2 presents the concept of SA and generalizes the idea of SA from a variety of NR image quality measures in the literature. Section 6.3 elaborates on the model of SA measure and proposes some examples of its implementation. The effectiveness and limitations of the SA measure are discussed in Section 6.4. This chapter ends with a summary in Section 6.5.

6.1 Motivation

The goal of objective image quality assessment is to develop computational models to quantitatively predict perceived image quality. The issue of objective image quality assessment has been approached from different perspectives. The classical fidelity measures of MSE and its variant PSNR have found widespread use due to their simplicity and mathematical convenience. However, they have been widely criticized for the limited accuracy when predicting perceived visual quality, e.g., [6–10]. Perceptual image quality measures have been developed to model the relevant functional components in the HVS - see a review in [11, 12]. Based on the assumption that the HVS is highly adapted to extracting structural information from visual scenes, the structurally-oriented viewpoint claims to attribute image quality degradation to the loss of structural information [32]. Despite being founded on different principles, these image quality measures are fundamentally connected to each other with the common theme of comparing a distorted image with a reference image. In other words, they are all FR image quality measures and require full access to a distortion-free image. However, a reference image is expensive to obtain or simply not available in many practical applications. In these applications, NR image quality measures, which appraise quality solely from a distorted image without any reference, are highly desirable. Due to its intrinsic difficulty, the field of NR image quality assessment is still in its infancy and remains largely unexplored to date. So far, the development of NR image quality measures lags the advances in FR image quality assessment. As an open research field with substantial practical potential, NR image quality assessment has attracted increasing research attention.

We approach the issue of NR image quality assessment from the structural and HVS-based perspectives. Under the assumption that human visual perception is highly sensitive to the structural information in a scene, a concept of SA together with a model of an SA measure is proposed as a new framework for NR image quality assessment. In this framework, image quality is estimated based on the quantification of comprehensive SA information in an image. Despite no explicit HVS modeling, SA information is differentiated according to visual significance and is assigned the corresponding importance in image quality prediction.

Due to the enormous difficulty and very early stages of NR image quality assessment, it is necessary to limit the scope of the SA measure in this study. Most existing NR image quality measures are developed with the foreknowledge of the distortion. General-purpose NR image quality assessment applicable to all kinds of distortions is an extremely difficult task. Some general NR image quality measures were reported [82, 86], with the framework of [82] based on natural scene statistics, and the blind image quality index of [86] based on the statistics of the DCT coefficients. Despite the progress in general-purpose NR image quality assessment, the scope of these image quality measures [82, 86], i.e., the image distortions that these image quality measures apply to, heavily relies on the image distortions under training. We assume the prior knowledge of the distortion and work with those images corrupted by a single distortion in this study. Fortunately, the distortion process is known in many applications, and the task of developing distortion-specific NR image quality measures is of practical importance. Different from those existing NR image quality measures that are developed for a particular distortion, the SA measure provides a more general quality assessment model

with its scope extended to a number of distortion types. Specifically, the performance of the SA measure is investigated in the context of white noise, Gaussian blur, and JPEG and JPEG2000 compression in this study. We believe the effective implementation of the SA measure or a hybrid quality measure combing the SA measure and other suitable image quality measures could extend the scope to a broader class of distortion types as well as multiple distortions.

6.2 Structural Activity

The concept of SA is defined as those properties, either in the spatial domain or other domains, whose variation can indicate the change of structural information in an image. SA can be a local concept concerned with the local structural information or a global concept concerned with structural information across most or all the image. In fact, SA is the underlying idea of many reported NR image quality measures. We would like to take a close look at those measures to reveal the implicit idea of SA.

The blur measures presented in [70, 71] are based on the measurement of edge spread, which is counted as the number of pixels with monotonically-changing intensities along the gradient orientation at an edge pixel. Apparently, the feature of edge spread is the spatial extent of an edge pixel with increased value indicating increased blur. Since edges deliver the most significant structural information in an image, edge spread provides an SA feature in nature, namely, increased edge spread indicates weakened SA and declining visual quality. A notable property of edge spread is that it does not distinguish among blur sources. The property allows the application of edge spread-based image quality measures to different types of

blur.

Blocking artifacts in JPEG compressed images are a kind of pseudo structure since they arise from the distortion and are not the genuine structures of the original image content. Most existing image quality measures, e.g., [55, 57, 60, 61], evaluate the effect of blocking artifacts based on their visibility or strength. Structure strength turns out to be another eligible SA feature; increased visibility of pseudo structures indicates increased SA and declining image quality.

Among the NR quality measures that are able to work for JPEG2000 compressed images, the model of [75] captures the nonlinear dependencies among wavelet coefficients, and appraises quality via quantifying the deviation of these dependencies from the expected statistics. Some researchers claimed that these nonlinear dependencies could be explained solely by the presence of spatial structures in natural images [101]. Different from those SA features in the spatial domain, the feature of nonlinear dependencies among wavelet coefficients reflects SA in the wavelet domain. Another example is the ZC rate in [78], which is defined as the number of local intensity extrema termed zero crossings in the literature. From the mathematical perspective, most structures are characterized by the monotonically-changing intensities at different strength levels. Considering that monotonically-changing intensities normally end with local intensity extrema, the feature of ZC rate is indirectly related to structures with a large value indicating an image region rich in structures. Apart from the genuine structures, the pseudo structures of ringing artifacts created by compression generate zero crossings in their neighborhood as well. It is shown that both the distorted genuine structures and pseudo structures may influence the distribution of zero crossings. In other words, the ZC rate reflects structural distortion and provides another SA

feature. Additionally, the principal component analysis employed in [73] and the anisotropic diffusion process in [58] are just different ways to extract SA, and their derived features are lent to the idea of SA.

Although the researchers claimed good results, those reported image quality measures have some limitations. The image quality measures [70,71,73] are heavily dependent on reliable edge information. Given the reduction in extracted edges with increased blur, the effectiveness of these measures would inevitably decline when images are subjected to severe distortion. A sophisticated edge detection method may help to alleviate the problem, such as using the Canny edge detector in [71] instead of the Sobel operator in [70] for edge extraction, but this is not an innovative solution. In addition, the operation of edge extraction easily leads to the loss of useful structural information, since only the structures of relatively high strength are kept whereas detailed structures of low strength are discarded during edge extraction. Another concern is practical implementation. Since the image quality measures [70,71,73] all rely on structure or feature extraction, the selection of a suitable threshold, which is largely experience/image-dependent, would be inconvenient in practical applications. Similarly, the choice of projection axes in [73] is also made in an *ad hoc* manner. Moreover, the measures developed in [57,73,75,78] require additional procedures and training data for model parameter determination. This imposes a restriction especially when data are limited.

As can be seen, the implicit idea of SA is generalized from a variety of NR image quality measures. With both strengths and weaknesses of these SA-motivated quality measures in mind, a new model of an SA measure is developed in the following section.

6.3 Structural Activity Measure

A new model of an SA measure, with its block diagram illustrated in Figure 6.1, is applied to the luminance component of an image. The flexible structure of the SA measure allows the easy incorporation of additional components and facilitates the extension of the model to various applications. The implementation of the SA measure comprises two stages. In the first stage, both the model components of SA weight and local SA are evaluated pixel-by-pixel over the entire image, resulting in an SA weight map and an SA map, respectively. In the second stage, the model component of global SA combines the two maps into a single numerical value as an image quality score. The three model components, together with some example implementations, are elaborated on in the following content. The implementations described here mainly serve to demonstrate the effectiveness of the SA measure.

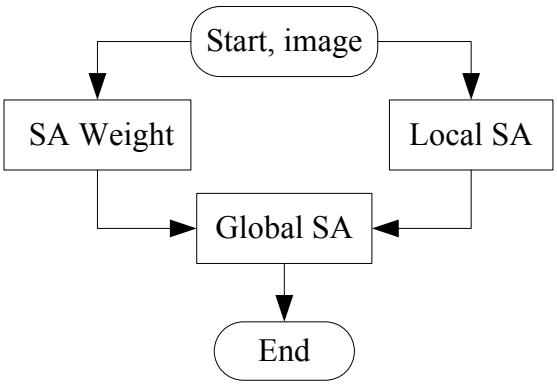


Figure 6.1: Block diagram of the SA measure.

6.3.1 Structural Activity Weight

The motivation for SA weight is to automatically adjust the relative contribution among the local SA information according to visual significance. Due to

the complexity and limited knowledge of the HVS, there is so far no widely accepted way to interpret visual attention. A number of studies regarding visual importance-motivated image quality assessment have been reported in the literature, e.g., [98,102,103]. Visual importance is expressed using some well-recognized influencing factors in [102], color contrast, texture contrast, motion, and cognitive features in [103], and fixation points together with local quality levels in [98]. Assuming that human visual perception is highly sensitive to the structural information, visual significance is described from the structural perspective in this study: a larger SA weight is assigned to the spatial location where the SA information is more likely to attract visual attention. As specific examples, SA weight is expressed in terms of structure strength (denoted by SAW-SS) and zero-crossing activity (SAW-ZC) in the following content.

6.3.1.1 Structure Strength-Based Structural Activity Weight

As discussed in Section 6.2, strong structures are more visible to the HVS and are inclined to attract more visual attention as compared to weak structures. Structure strength is therefore employed as an alternative expression of SA weight. A number of methods are available for structure strength calculation, such as the widely used Sobel operator. A *multistage median filter* (MMF)-based approach is proposed to compute structure strength with the intention to indirectly reinforce the detailed structures of relatively low strength. First, the definition of MMF is introduced below.

Let W be a $(2w+1) \times (2w+1)$ square filter window centered at a pixel (i, j) with integers i and j as spatial indices. The pixels in W can be divided into four

subsets:

$$\begin{aligned} W_1(i,j) &= \{I(i,j-m); -w \le m \le w\} \\ W_2(i,j) &= \{I(i-m,j); -w \le m \le w\} \\ W_3(i,j) &= \{I(i+m,j-m); -w \le m \le w\} \\ W_4(i,j) &= \{I(i-m,j-m); -w \le m \le w\} \end{aligned} \tag{6.1}$$

Let $z_k(i,j)$ denote the median pixel of the subset $W_k(i,j)$, where the integer k with $1 \le k \le 4$ is the index:

$$z_k(i,j) = \text{med}(W_k(i,j)) \quad 1 \le k \le 4 \tag{6.2}$$

With

$$\begin{aligned} y_{\min}(i,j) &= \min_{1\le k\le 4}(z_k(i,j)) \\ y_{\max}(i,j) &= \max_{1\le k\le 4}(z_k(i,j)) \end{aligned} \tag{6.3}$$

the output of MMF at pixel (i,j) is defined by

$$I_{\text{MMF}}(i,j) = \text{med}(y_{\min}(i,j), y_{\max}(i,j), I(i,j)) \tag{6.4}$$

where $\min(\cdot)$, $\max(\cdot)$, and $\text{med}(\cdot)$ denote the operations for the minimum, maximum, and median, respectively. In our implementation, $w = 2$. A detailed description of MMF can be found in [104].

Structures are characterized by abrupt intensity changes and correspond to the high-frequency components in an image. With the low-frequency component estimated as the mean of local median pixels in equation (6.5), the high-frequency component is computed as the difference between the filtered image and the low-frequency component. The structure strength and also the local SA weight at pixel (i,j) is thus expressed in equation (6.6) as the magnitude of the high-frequency

component:

$$I_{\text{mean}}(i,j) = \frac{\sum_{k=1}^{4} z_k(i,j)}{4} \tag{6.5}$$

$$\text{SAW}_{\text{SS}}(i,j) = |I_{\text{MMF}}(i,j) - I_{\text{mean}}(i,j)| \tag{6.6}$$

For a better understanding, the proposed approach is compared with Wang's MMF-based edge detector [105]. The major difference between these two methods include: (a) MMF serves to keep image details while smoothing noise in Wang's detector, but preserves details from distorted images with the distortion not limited to noise in the proposed approach; and (b) the low-frequency component is estimated as the mean of the general pixels in Wang's detector, but the mean of the median pixels in the proposed approach. Given that MMF basically works with the median pixels rather than general pixels, the mean of median pixels has more dependencies with MMF output in comparison with the mean of general pixels. As a result, structure strength is quantified at different degrees by these two methods. Specifically, in comparison with Wang's detector, structure strength obtained by the proposed approach tends to have a moderate distribution, which is intended to indirectly enhance the detailed structures of relatively low strength. Both qualitative and quantitative studies are used in Section 6.4 to reveal and validate the perceptual implication of the proposed approach.

6.3.1.2 Zero Crossing-Based Structural Activity Weight

Let W be a $N_1 \times N_2$ filter window centered at a pixel (i,j). The horizontal difference at a pixel (m,n) is calculated by

$$d_{\text{H}}(m,n) = I(m,n) - I(m,n-1) \quad 1 \le m \le N_1, \quad 2 \le n \le N_2 \tag{6.7}$$

where m and n are window-wide spatial indices. In our implementation, $N_1 = N_2 = 5$. The horizontal ZC activity at pixel (i, j) is defined by

$$\mathrm{ZC_H}(i,j) = \sum_{m=1}^{N_1} \sum_{n=2}^{N_2-1} \mathrm{sign\_}(\mathrm{sign}(d_\mathrm{H}(m,n)) \cdot \mathrm{sign}(d_\mathrm{H}(m,n+1))) \tag{6.8}$$

where

$$\mathrm{sign}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

and

$$\mathrm{sign\_}(x) = \begin{cases} 1 & x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.10}$$

Similarly, the vertical ZC activity $\mathrm{ZC_V}(i, j)$ can be computed by transposing window W and using the same algorithm as that for the horizontal ZC activity. Finally, ZC activity and also the local SA weight at pixel (i, j) is defined by

$$\mathrm{SAW_{ZC}}(i,j) = \mathrm{ZC_H}(i,j) + \mathrm{ZC_V}(i,j) \tag{6.11}$$

As analyzed in Section 6.2, the feature of ZC activity is indirectly related to structures with a large value indicating an image region rich in structures. Based on the assumptions that the HVS is highly sensitive to the structural information and the image regions rich in structures tend to attract visual attention, ZC activity is employed as another alternative expression of SA weight.

6.3.2 Local Structural Activity

The feature of direction spread is proposed as an example of the local SA expression in this section. The basic principle of direction spread measurement is to track the pixels with monotonically-changing intensities along the horizontal and vertical directions at a general pixel. Specifically, the number of pixels with strictly increasing or decreasing intensities is first counted along the horizontal and vertical directions respectively, and then direction spread is given as the maximum of these two counts.

As illustrated in Figure 6.2, horizontal (H) and vertical (V) directions are defined at a pixel (i, j). Their positive and negative sub-directions are denoted by +H & -H and +V & -V, respectively. Directional traces, starting from a pixel (i, j)

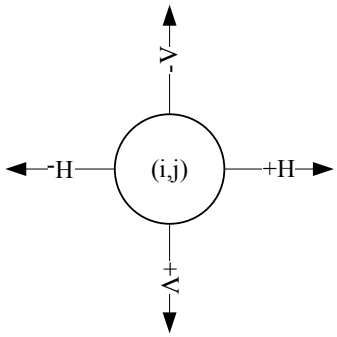


Figure 6.2: Directions at a pixel (i, j).

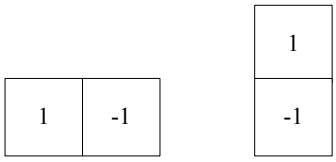


Figure 6.3: Masks applied to an image with the weight "1" aligned with pixels from the directional traces. The mask on the left is applied to the pixels from $T_{+\mathrm{H}}(i,j)$ & $T_{+\mathrm{H}}(i,j)$, and the mask on the right to the pixels from $T_{+\mathrm{V}}(i,j)$ & $T_{+\mathrm{V}}(i,j)$.

and extending along the different sub-directions, are represented by

$$\begin{aligned} T_{+\mathrm{H}}(i,j) &= \left\{ \; I(i, j+m_{+\mathrm{H}}); \quad m_{+\mathrm{H}} = 0,1,2,... \; \right\} \\ T_{-\mathrm{H}}(i,j) &= \left\{ \; I(i, j-m_{-\mathrm{H}}); \quad m_{-\mathrm{H}} = 1,2,... \; \right\} \\ T_{+\mathrm{V}}(i,j) &= \left\{ \; I(i+m_{+\mathrm{V}}, j); \quad m_{+\mathrm{V}} = 0,1,2,... \; \right\} \\ T_{-\mathrm{V}}(i,j) &= \left\{ \; I(i-m_{-\mathrm{V}}, j); \quad m_{-\mathrm{V}} = 1,2,... \; \right\} \end{aligned} \tag{6.12}$$

where integers m_{+D} and m_{-D} with $D = \{\mathrm{H}, \mathrm{V}\}$ (denote H and V directions) orderly index pixels along the positive and negative sub-directions and limit the pixel location within the effective image region. Different masks, as shown in Figure 6.3, are applied to the image with the weight "1" aligned with the pixels from the respective directional traces. The measurement of direction spread at a pixel (i,j) follows several steps below.

Step 1. Initial judgment.

Start from the first pixels $T^0_{+D}(i,j)$ and $T^1_{-D}(i,j)$ of the respective directional traces. $T^0_{+D}(i,j)$ is $T^{m_{+D}}_{+D}(i,j)$ with $m_{+D} = 0$, and $T^1_{-D}(i,j)$ is $T^{m_{-D}}_{-D}(i,j)$ with $m_{-D} = 1$, where $T^{m_{+D}}_{+D}(i,j)$ and $T^{m_{-D}}_{-D}(i,j)$ specify the pixels within the directional traces in equation (6.12) along the positive and negative sub-directions, respectively. Apply the corresponding masks and compute the correlation results

$MT^0_{+D}(i,j)$ and $MT^1_{-D}(i,j)$. If $\text{sign}_+(\text{sign}(MT^0_{+D}(i,j)) \cdot \text{sign}(MT^1_{-D}(i,j))) = 0$, the direction spread in the D direction is set as zero; otherwise, proceed to step 2. Here

$$\text{sign}_+(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.13}$$

Step 2. Tracking.

Continuously move the masks along the directional traces pixel-by-pixel as long as $\text{sign}_+(\text{sign}(MT^{m_{+D}}_{+D}(i,j)) \cdot \text{sign}(MT^0_{+D}(i,j))) = 1 \quad m_{+D} = 1, 2, ...$ and $\text{sign}_+(\text{sign}(MT^{m_{-D}}_{-D}(i,j)) \cdot \text{sign}(MT^1_{-D}(i,j))) = 1 \quad m_{-D} = 2, 3, ...$. The tracking operation is implemented in the respective directional traces independently.

Step 3. Compute direction spreads in the H *and* V *directions.*

Compute the direction spread in the D direction by

$$\text{DS}_D(i,j) = m_{+D} + m_{-D} + 1 \tag{6.14}$$

Step 4. Compute direction spread at pixel (i,j).

The direction spread at pixel (i,j) is defined in equation (6.15) as the maximum of the direction spreads in H and V directions. The maximum operation is intended to make direction spread sensitive to distortion:

$$\text{DS}(i,j) = \max_{D=\{\text{H,V}\}} (\text{DS}_D(i,j)) \tag{6.15}$$

We would like to compare direction spread with the feature of edge spread reported in [70,71]. Their significant difference includes: (a) edge spread works with the extracted edges of relatively high strength, while direction spread is applied to general pixels and serves to cover structures of different strength levels; and (b)

edge spread is measured along the gradient orientation, while direction spread is measured along the horizontal and vertical directions. The general measurement along the horizontal and vertical directions provides a meaningful approximation given that the HVS pays much attention to structures along these two directions in comparison with other orientations [4]. Furthermore, we would like to point out the major advantages of direction spread over edge spread: (a) the operation of extracting edges is avoided; (b) more stable performance can be expected since direction spread overcomes the limitations imposed by structure/feature extraction of severely distorted images; and (c) more complete SA information is preserved.

A special case comes from JPEG compressed images. The quantization operation in JPEG compression is implemented in each coding block independently and leads to non-uniformly distributed distortions. We propose to restrict the measurement of direction spread within 8×8 coding blocks (the default DCT block size) of a JPEG compressed image rather than across the image space. However, it leads to no SA defined at block boundaries since direction spread should be measured within the effective image region. Assuming an image of $M \times N$ in size and t as the index of the inter-block boundaries, the SA at horizontal block boundaries is defined in equation(6.16) and at vertical block boundaries in equation(6.17) to complete the local SA definition in the case of JPEG compression. According to Weber-Fechner law as introduced in Section 2.2, the HVS is sensitive to the relative variation of luminance. As can be seen, equations (6.16) and (6.17) are essentially consistent with Weber-Fechner law. The SA at block boundaries is thus shown to be quantified in a perceptually meaningful manner.

$$B_{\mathrm{H}}(t) = \frac{\sum_{k=-7}^{0} |I(8i, 8j+k) - I(8i+1, 8j+k)|}{\sum_{k=-7}^{0} I(8i, 8j+k) + I(8i+1, 8j+k)} \tag{6.16}$$

$$B_{\mathrm{V}}(t) = \frac{\sum_{k=-7}^{0} |I(8i+k, 8j) - I(8i+k, 8j+1)|}{\sum_{k=-7}^{0} I(8i+k, 8j) + I(8i+k, 8j+1)} \tag{6.17}$$

where $i = 1, 2, ..., \lfloor M/8 \rfloor - 1$ and $j = 1, 2, ..., \lfloor N/8 \rfloor - 1$, $t = (\lfloor N/8 \rfloor - 1) \times (i-1) + j$ and $1 \leq t \leq (\lfloor M/8 \rfloor - 1) \times (\lfloor N/8 \rfloor - 1)$, and $\lfloor \cdot \rfloor$ denotes the operation of rounding the operand to the nearest integer towards minus infinity.

6.3.3 Global Structural Activity

A pooling approach is required to combine all the SA weights and local SA into a single scalar of global SA as image quality score. In this study, a visual significance-weighted pooling approach is proposed with the general expression

$$Q_o = \frac{\sum_i w_i q_i}{\sum_i w_i} \tag{6.18}$$

where Q_o is the image quality score, i is the index, q_i is the local quality/distortion estimate, and w_i is visual significance-based local weight with a large value indicating substantial visual interest. The basic idea is to determine the relative contribution of local quality/distortion estimate according to visual significance. It is motivated by the fact that the HVS, as the only "correct" quality assessment model, should determine the importance of the local quality/distortion estimate. For a good understanding of the proposed pooling strategy, we compare it with Wang's two pooling approaches [100]. The two approaches presented in [100], namely, local quality/distortion-weighted pooling and information content-weighted pooling, prove the good potential to improve image quality prediction. Hereafter, these two approaches are referred to as quality- and information-weighted approaches. In the quality-weighted approach, the weight is defined as the monotonic function of the

local quality/distortion measure with the intention to emphasize the image regions of poor quality. In the information-weighted approach, the weight is defined as the function of local information content with the intention to emphasize the image regions with substantial information. In fact, both approaches may converge to the common idea of visual significance, with visual attention assumed to be attracted by the image regions of poor quality in the quality-weighted approach and the image regions of substantial information in the information-weighted approach. With respect to the proposed model of the SA measure, some examples of SA weight expression are already proposed in Section 6.3.1 to interpret visual significance from the structural perspective.

The specific expressions of the proposed pooling approach are presented according to different distortion types in the following subsections. A common feature of these expressions is that they are parameterless, which avoids additional procedures or training data for parameter determination.

6.3.3.1 Gaussian Blur and White Noise

When an image is subject to Gaussian blur or white noise, the proposed pooling approach is expressed by

$$\mathrm{SA}_m = \frac{\sum_i \sum_j \mathrm{SAW}_m(i,j) \cdot \mathrm{DS}(i,j)}{\sum_i \sum_j \mathrm{SAW}_m(i,j) \cdot \mathrm{sign}_+(\mathrm{DS}(i,j))} \quad m = \{\mathrm{SS,ZC}\} \tag{6.19}$$

With valid SA weights assigning spatially varying importance, the global SA is calculated in equation (6.19) as the weighted average of non-zero local SA (i.e., direction spread). SA weights are deemed valid at the spatial locations where the local direction spread is non-zero. Increased global SA indicates declining

image quality in the case of Gaussian blur, while decreased global SA reflects degraded image quality in the case of white noise. Compared with the observation that increased direction spread indicates weakened SA due to blur, the relation between direction spread and white noise is not so explicit. Although the individual direction spread is interrupted by noise in a random way, it is found that the overall direction spread statistically decreases and reflects the increased SA of pseudo structures produced by noise. This observation regarding white noise will be further validated in Section 6.4.2.1.

6.3.3.2 JPEG Compression

As presented in Section 6.3.2, the local SA is expressed differently at the inner pixels and boundaries of coding blocks in JPEG compressed images. Accordingly, the global SA of a JPEG compressed image is computed in two steps. First, the part of SA at inner pixels of coding blocks is calculated using equation(6.20) and the part of SA at block boundaries using equation(6.21). Here, SA weight comes into play only at the inner pixels of coding blocks in equation(6.20) and has no effect at block boundaries in equation(6.21). Second, the global SA is computed in equation(6.22) as the combination of the two component SA:

$$\mathrm{SA}_{m(\mathrm{inner})} = \frac{\sum_{(i,j)\in\{(i,j)\}_{\mathrm{inner}}} \mathrm{SAW}_m(i,j)\cdot \mathrm{DS}(i,j)}{\sum_{(i,j)\in\{(i,j)\}_{\mathrm{inner}}} \mathrm{SAW}_m(i,j)\cdot \mathrm{sign}_+(\mathrm{DS}(i,j))} \tag{6.20}$$

$$\mathrm{SA}_{\mathrm{boundary}} = \frac{\sum_{t=1}^{(\lfloor M/8\rfloor-1)\times(\lfloor N/8\rfloor-1)} B_{\mathrm{H}}(t)+B_{\mathrm{V}}(t)}{\sum_{t=1}^{(\lfloor M/8\rfloor-1)\times(\lfloor N/8\rfloor-1)} \mathrm{sign}_+(B_{\mathrm{H}}(t))+\mathrm{sign}_+(B_{\mathrm{V}}(t))} + 1 \tag{6.21}$$

$$\mathrm{SA}_m = (\mathrm{SA}_{m(\mathrm{inner})})^{\alpha}\cdot(\mathrm{SA}_{\mathrm{boundary}})^{\beta} \tag{6.22}$$

where $m = \{\mathrm{SS}, \mathrm{ZC}\}$, $\{(i,j)\}_{\text{inner}}$ represents the set of spatial indices regarding all the inner pixels of coding blocks, the constant 1 in equation(6.21) represents a baseline minimal value, $\alpha > 0$ and $\beta > 0$ in equation(6.22) are parameters used to adjust the relative emphasis between the two parts SA. $\alpha = \beta = 1$ is chosen in this study. It is easy to see $\mathrm{SA}_{m(\text{inner})}$ increases with increased blur, $\mathrm{SA}_{\text{boundary}}$ increases when blockiness becomes more visible, and the resulting global image quality score SA_m increases with declining image quality.

6.3.3.3 JPEG2000 Compression

According to the presentation in Section 6.3.2, direction spread has meaningful values at structures with gradual intensity changes, while zero value at structures characterized by abrupt intensity changes with a representation of 2-D step function. Although the structures with abrupt intensity changes are an ideal case and seldom appear in natural images, it is inadequate for direction spread to cover the SA of ringing artifacts in JPEG2000 compressed images. The computation of global SA is complemented in equation(6.23) by incorporating SA weights at spatial locations where direction spread has no effect. When blur and ringing artifacts become more visible, the global SA increases and indicates degraded image quality.

$$\mathrm{SA}_m = \frac{\sum_i \sum_j \mathrm{SAW}_m(i,j) \cdot \mathrm{DS}(i,j) + \mathrm{SAW}_m(i,j) \cdot \mathrm{sign}_0(\mathrm{DS}(i,j))}{\sum_i \sum_j \mathrm{SAW}_m(i,j)} \quad m = \{\mathrm{SS}, \mathrm{ZC}\} \tag{6.23}$$

where

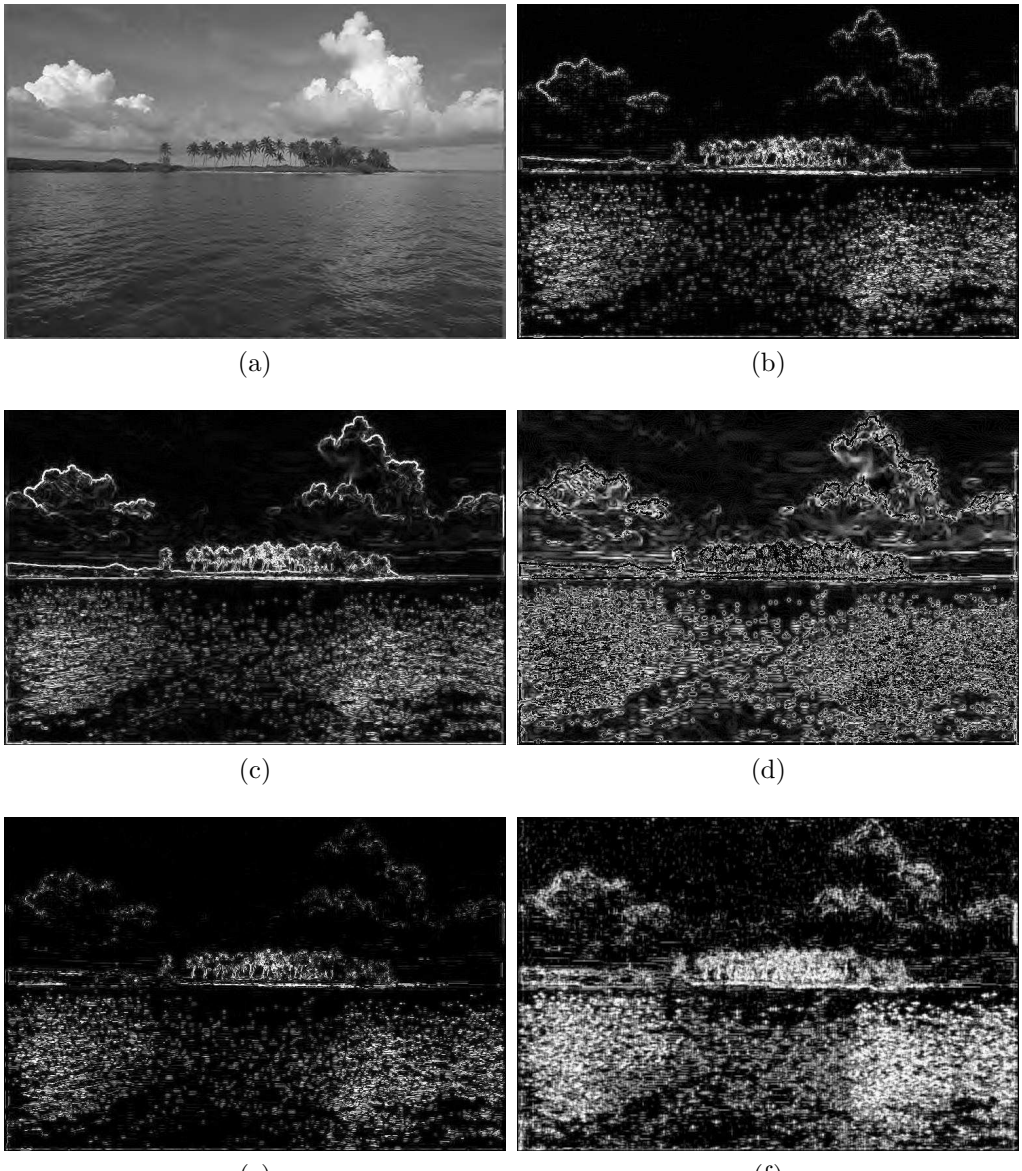
$$\mathrm{sign}_0(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.24}$$

6.4 Results and Discussion

The performance of the proposed model of the SA measure is evaluated using the LIVE image database [88,89], which has been widely employed for the performance evaluation of image quality measures. Specifically, the tests involve the datasets of white noise contaminated images, Gaussian blurred images, JPEG and JPEG2000 compressed images. Table 6.1 provides the information regarding the number of images in each dataset, parameters of distortion, and subjective quality scores. The distortion parameter varies according to distortion type. Specifically, the parameter is the standard deviation of white Gaussian noise (in pixels) in the case of white noise, the standard deviation of Gaussian filter (pixels) in the case of Gaussian blur, and bit rate (bpp) in the case of JPEG and JPEG2000 compression. The subjective quality scores are given in terms of the realigned DMOS on a scale of 0 to 100, with a larger DMOS indicating poorer visual quality. More details of the LIVE image database can be found in Section 3.4.1. In the following content, the performance of the SA measure is investigated through the visualization of SA weight, quantitative performance evaluation, and outlier analysis.

Table 6.1: Information of the LIVE image database: number of images in each dataset, parameters of the distortion, and subjective quality scores.

Dataset	Number of images	Parameters	Subjective quality scores
White noise	145	0.012-2.00	Realigned DMOS (exclude the reference images)
Gaussian blur	145	0.42-15.00	
JPEG	175	0.15-3.34	
JPEG2000	169	0.028-3.15	



(a) (b)

(c) (d)

(e) (f)

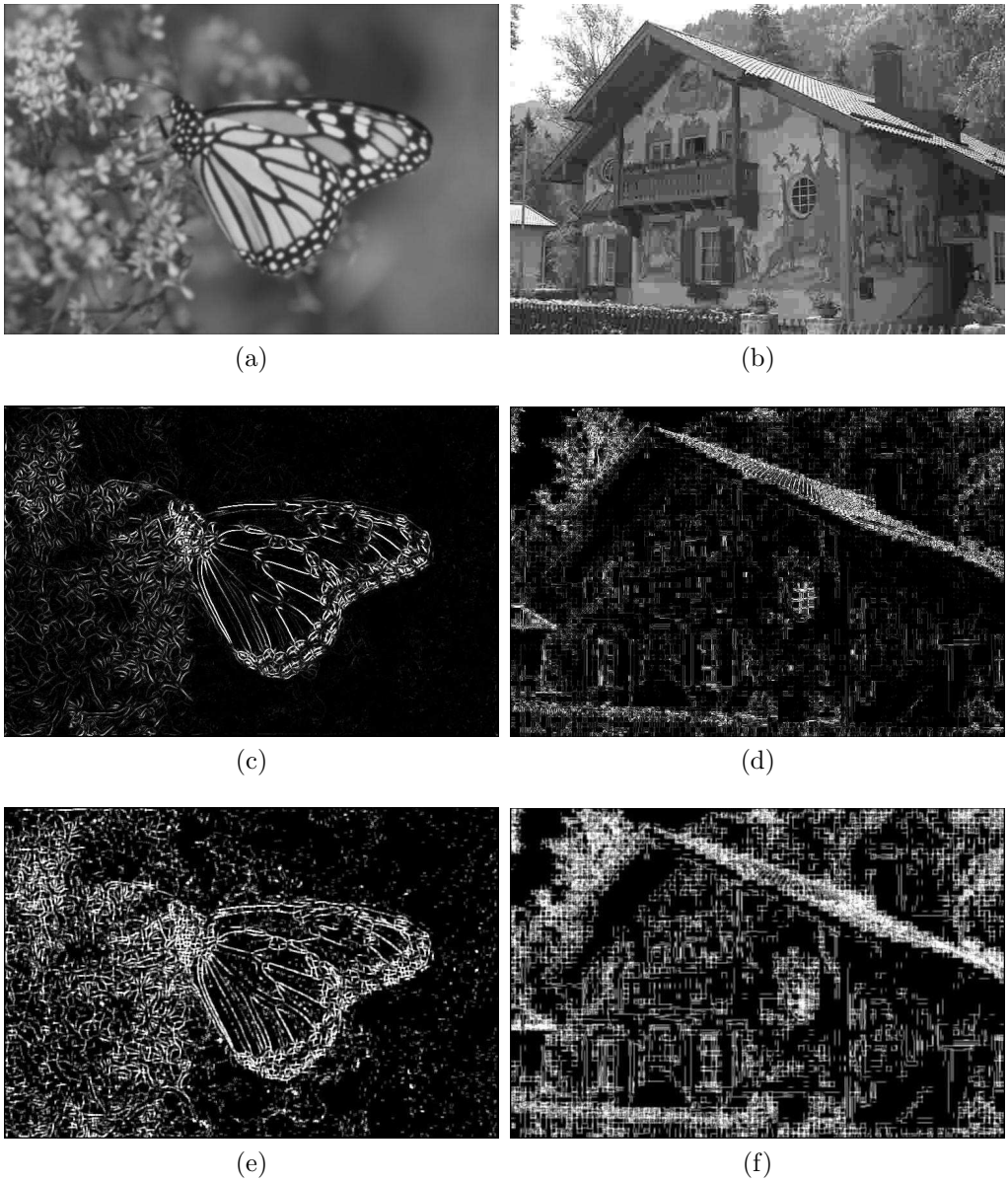
Figure 6.4: Visualization of structure strength and SA weight over a sample JPEG2000 image. (a) The "Ocean" JPEG2000 image compressed to 0.1914 bpp. The structure strength map shown in (b) is obtained by Wang's detector, (c) by the Sobel operator, and (d) by implementing an additional thresholding over (c). The thresholding in (d) is performed in a way that all the structure strength values larger than 10% of the largest value are set as zero. The SA weight map shown in (e) is obtained by SAW-SS and (f) by SAW-ZC. All the maps are normalized and contrast stretched for visibility, with a brighter pixel indicating a larger value.

6.4.1 Visualization of Structural Activity Weight

We first illustrate the perceptual implication of the proposed MMF-based approach with a set of intuitive examples. Based on a sample "Ocean" JPEG2000 compressed image shown in Figure 6.4a, a set of structure strength maps are obtained in different ways and are visualized in Figures 6.4b-6.4e. The brightness of these normalized structure strength maps represents the relative magnitude, with a brighter pixel indicating a relatively larger strength value. We discuss the relative rather than the absolute magnitude of structure strength in the following content. Specifically, Figure 6.4b is obtained by Wang's detector [105], Figure 6.4c by the Sobel operator, Figure 6.4d by applying an additional thresholding step to Figure 6.4c, and Figure 6.4e by the proposed MMF-based approach (denoted by SAW-SS in Figure 6.4). Thresholding in Figure 6.4d is utilized only for demonstration purposes and is implemented such that all the structure strength values larger than 10% of the largest value are set as zero. Note the visible blur and ringing artifacts in the compressed "Ocean" image (Figure 6.4a), especially in the region of the ocean. With reference to the structure strength maps obtained by Wang's detector (Figure 6.4b) and the Sobel operator (Figure 6.4c), it is observed that the contours of the clouds, island, trees, and the waves of the ocean have high strength. In contrast, the detailed structures like ringing artifacts at the waves have much lower strength. The fact is that the pseudo structures of ringing artifacts disturb the natural behavior exhibited in an image, and are usually a major attraction of visual attention. A clear gap is consequently seen between the distributions of structure strength and visual interest. To better illustrate this issue, an additional thresholding step is applied to the image of Figure 6.4c to ignore the largest structure strength values, and a new map is derived (Figure 6.4d). Comparing Figures 6.4c with 6.4d, it can be

observed that the strength of ringing artifacts is clearly enhanced in Figure 6.4d, which is obtained with the thresholding step. Obviously, a practical solution to the revealed issue, i.e., the gap between the distributions of structure strength and visual interest, can not be set up in this way. As theoretically analyzed in Section 6.3.1.1, the proposed MMF-based approach is intended to automatically moderate the distribution of structure strength so as to indirectly enhance the detailed structures of relatively low strength. As can be seen, the structure strength computed by the MMF-based approach (denoted by SAW-SS in Figure 6.4e) indeed appears more moderate in distribution in comparison with Wang's detector (Figure 6.4b) and the Sobel operator (Figure 6.4c), as exemplified by the distinctly decreased strength at the contours of the clouds, island, and trees in Figure 6.4e. As a result, the strength of the ringing artifacts in the structure strength distribution is indirectly enhanced, and the MMF-based approach is shown to accomplish its goal.

More SA weight maps are shown in Figures 6.4 and 6.5, where SA weight is expressed in terms of SAW-SS and SAW-ZC, and the image distortions include Gaussian blur, and JPEG and JPEG2000 compression. The brightness of these normalized SA weight maps represents the relative magnitude, with a brighter pixel indicating a relatively larger SA weight. It is noted that SAW-SS provides better resolution than SAW-ZC, as confirmed by the more distinct details shown in Figures 6.4e, 6.5c, and 6.5d (obtained by SAW-SS) in comparison with the coarse content shown in Figures 6.4f, 6.5e, and 6.5f (SAW-ZC). It is mainly because SAW-SS is derived based on the magnitude of intensity changes, while SAW-ZC is based on the sign of intensity changes. Despite their different expressions, SAW-SS and SAW-ZC are demonstrated perceptually meaningful with their distributions consistent with visual interest as a whole.



(a) (b)

(c) (d)

(e) (f)

Figure 6.5: Visualization of SA weight over sample Gaussian blurred and JPEG compressed images. (a) The "Monarch" Gaussian blurred image with the parameter of 1.8515. (b) The "Painted house" JPEG image compressed to 0.2994 bpp. The SA weight maps shown in (c) and (d) are obtained by SAW-SS, and (e) and (f) by SAW-ZC. All the SA weight maps are normalized and contrast stretched for visibility, with a brighter pixel indicating a larger value.

6.4.2 Quantitative Performance Evaluation

Hereafter, SA-SS and SA-ZC are used to denote the implementations of the SA measure based on SAW-SS and SAW-ZC, respectively. Based on the ground truth data of DMOS provided by the LIVE image database, two representative FR image quality measures of PSNR and SSIM index [32] (introduced in Section 3.1) are employed as the benchmarks to study the performance of the SA measure. The performance evaluation of the SA measure follows the procedures recommended by the VQEG as detailed in Section 3.4.2. Table 6.2 lists the results of evaluation criteria regarding all the image quality measures under study. The negative sign of ROCC in Table 6.2 indicates an inversely proportional relationship between subjective and objective quality scores. This does not occur with CC because of the compensation by nonlinear mapping (refer to Section 3.4.2). It means the magnitude of ROCC when ROCC is referred to in the following content.

6.4.2.1 White Noise

From Table 6.2, it is shown that all the image quality measures under test exhibit good performances. PSNR delivers the best results and proves an excellent FR quality measure in the case of white noise. SA-SS outperforms SA-ZC according to each evaluation criterion, and the performance of SA-SS is basically as good as that of SSIM index. The validity of the SA measure, be it SA-SS or SA-ZC, is justified through its good prediction of subjective quality scores. We further validate the logic underlying the SA measure presented in Section 6.3.3.1: the overall direction spread statistically decreases and reflects increased noise. With image quality scores calculated as the average of non-zero direction spreads across an image, the following results can be obtained: CC of 0.959, RMSE of 8.016,

Table 6.2: Performances of image quality measures over the LIVE image database.

Quality measure	Type	White noise			
		CC	RMSE	OR	ROCC
PSNR	FR	0.986	4.769	0.262	-0.985
SSIM index	FR	0.970	6.883	0.531	-0.969
SA-SS	NR	0.971	6.839	0.531	-0.959
SA-ZC	NR	0.952	8.717	0.669	-0.931
Quality measure	Type	Gaussian blur			
		CC	RMSE	OR	ROCC
PSNR	FR	0.783	11.641	0.738	-0.782
SSIM index	FR	0.945	6.119	0.552	-0.952
SA-SS	NR	0.907	7.876	0.669	0.888
SA-ZC	NR	0.906	7.920	0.641	0.892
Quality measure	Type	JPEG compression			
		CC	RMSE	OR	ROCC
PSNR	FR	0.888	14.826	0.634	-0.881
SSIM index	FR	0.979	6.576	0.394	-0.976
SA-SS	NR	0.937	11.234	0.554	0.932
SA-ZC	NR	0.923	12.417	0.531	0.918
Quality measure	Type	JPEG2000 compression			
		CC	RMSE	OR	ROCC
PSNR	FR	0.900	11.151	0.615	-0.895
SSIM index	FR	0.967	6.551	0.473	-0.961
SA-SS	NR	0.911	10.515	0.627	0.908
SA-ZC	NR	0.929	9.426	0.574	0.925

OR of 0.621, and ROCC of -0.941. Clearly, the feature of direction spread itself provides good image quality prediction. The results validate the logic and allow us to make a reasonable conclusion: direction spread plays a leading role in SA-SS and SA-ZC for the image quality prediction of white noise contaminated images.

6.4.2.2 Gaussian Blur

The performance of SSIM index is shown superior to and PSNR inferior to that of other image quality measures under test. SA-SS performs slightly better than

SS-ZC in terms of lower RMSE and better prediction accuracy (higher CC), while SS-ZC is better than SA-SS in terms of better prediction consistency (lower OR) and better prediction monotonicity (higher ROCC).

6.4.2.3 JPEG Compression

SSIM index delivers the best results, and SA-SS together with SA-ZC exhibit satisfactory performances which are demonstrably better than PSNR according to each evaluation criterion. SA-SS is superior to SA-ZC in terms of lower RMSE, better prediction accuracy (higher CC), and better prediction monotonicity (higher ROCC), while SA-ZC is better than SA-SS in terms of better prediction consistency (lower OR).

6.4.2.4 JPEG2000 Compression

PSNR is shown as an effective measure for JPEG2000 compressed images. SSIM index achieves further improvement over PSNR and supplies the best results among the image quality measures under test. SA-ZC outperforms PSNR according to each evaluation criterion. SA-SS is better than PSNR in terms of lower RMSE, better prediction accuracy (higher CC), and better prediction monotonicity (higher ROCC), while PSNR is better than SA-SS in terms of better prediction consistency (lower OR).

6.4.2.5 Performance Summary

SA-SS and SA-ZC perform the best in the case of white noise, followed by JPEG and JPEG2000 compression, and then Gaussian blur. SA-SS and SA-ZC are shown

to be good NR image quality measures in the case of white noise, and the performance of SA-SS is as good as that of the SSIM index. It is demonstrated that the feature of direction spread plays a leading role in SA-SS and SA-ZC for the quality estimation of white noise contaminated images. Moreover, the performances of SA-SS and SA-ZC are generally better than that of PSNR over a majority of the datasets including Gaussian blur and JPEG and JPEG2000 compression. With the FR quality measures of PSNR and SSIM index as benchmarks, the results in Table 6.2 suggest that the performances of the NR image quality measures of SA-SS and SA-ZC are promising. Besides, comparing the two implementations of the SA measure, SA-SS is generally superior to SA-ZC in the case of white noise and JPEG compression, while SA-ZC outperforms SA-SS in the case of JPEG2000 compression, and SA-SS and SA-ZC have similar performances in the case of Gaussian blur. Furthermore, we would like to point out some additional advantages of SA-SS and SA-ZC that have not been revealed in Table 6.2: (a) they do not require structure or feature extraction; (b) they are not subject to the limitations imposed by structure or feature extraction of severely distorted images; and (c) they avoid additional procedures and training data for parameter determination. Based on their favorable performances in various respects, we may conclude that the overall performances of SA-SS and SA-ZC are indeed satisfactory as far as such an NR image quality assessment model is concerned.

6.4.2.6 Quantitative Validation of the Multistage Median Filter-Based Approach

Besides the illustration in Section 6.4.1, the proposed MMF-based approach for SAW-SS computation merits further validation. We employ the Sobel operator

Table 6.3: Performance comparison between SA-SS and the tentative implementations.

Quality measure	Gaussian blur			
	CC	RMSE	OR	ROCC
SA-Sobel	0.867	9.331	0.717	0.841
SA-Wang	0.882	8.824	0.703	0.855
SA-SS	0.907	7.876	0.669	0.888
Quality measure	JPEG compression			
	CC	RMSE	OR	ROCC
SA-Sobel	0.908	13.498	0.611	0.903
SA-Wang	0.948	10.243	0.469	0.940
SA-SS	0.937	11.234	0.554	0.932
Quality measure	JPEG2000 compression			
	CC	RMSE	OR	ROCC
SA-Sobel	0.826	14.393	0.740	0.822
SA-Wang	0.881	12.075	0.710	0.877
SA-SS	0.911	10.515	0.627	0.908

and Wang's detector [105] to replace the proposed MMF-based approach to compute SAW-SS in the implementation of the SA measure. We refer to the resultant two tentative implementations as SA-Sobel and SA-Wang. The results of evaluation criteria are provided in Table 6.3 where white noise is not included since the leading role of direction spread is already justified in Section 6.4.2.1. Our observations of Table 6.3 include two aspects. On one hand, comparing Table 6.3 with Table 6.2, it can be seen that SA-Wang outperforms PSNR in the case of Gaussian blur, outperforms PSNR and SA-SS in the case of JPEG compression, and is comparable to PSNR in the case of JPEG2000 compression. The good performance of SA-Wang further validates the model of the SA measure since the only difference between SA-Wang and SA-SS is the different computation methods of SAW-SS. On the other hand, SA-SS is generally superior to SA-Sobel and SA-Wang especially in the case of Gaussian blur and JPEG2000 compression. Note that all these

three implementations achieve clear performance increases in the case of JPEG compression. We believe that it is primarily due to the separate SA quantification at block boundaries rather than the direct effect of SAW-SS. Based on the qualitative illustration in Section 6.4.1 and quantitative evaluation provided here, we conclude that the proposed MMF-based approach for SAW-SS is fully validated.

6.4.3 Outlier Analysis

The scatter plots of DMOS versus image quality scores computed by SA-SS and SA-ZC are shown in Figure 6.6, where SA-SS is selected for the distortions of white noise and JPEG compression and SA-ZC for Gaussian blur and JPEG2000 compression. Each point in the figure, marked by asterisk or "+", represents one test image. With most points close to the fitted logistic curve (corresponding to equation (3.9)) in the scatter plots, SA-SS and SA-ZC are observed to provide good prediction of DMOS for most test images. The outliers (denoted by "+" in Figure 6.6) are those data points whose prediction deviation exceeds twice the standard deviation of individual subjective quality scores. We take a close look at some representative outliers in this section.

With Gaussian blur, it is found that the most significant outliers arise from the blurred "Monarch" images, a sample image of which is shown in Figure 6.5a. It is easy for the HVS to recognize that the blur of the out-of-focus flowers in the background is due to the macro shot rather than distortion, but it is difficult for the SA measure to realize this. The SA measure fails to imitate the behaviors of the HVS and under-predicts image quality in this case.

The sample image shown in Figure 6.7a exemplifies a typical case of the outliers pertaining to JPEG compressed images. According to equations (6.16) and (6.17),

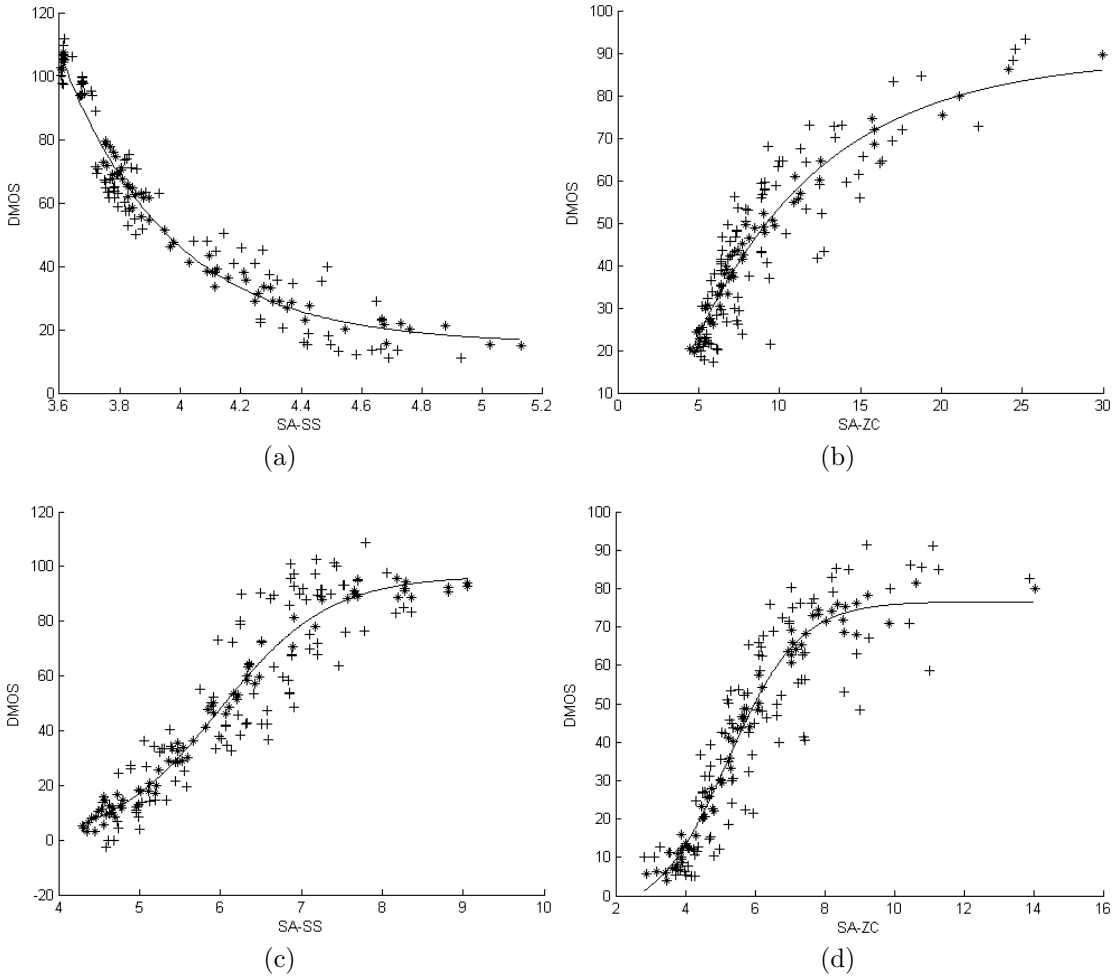
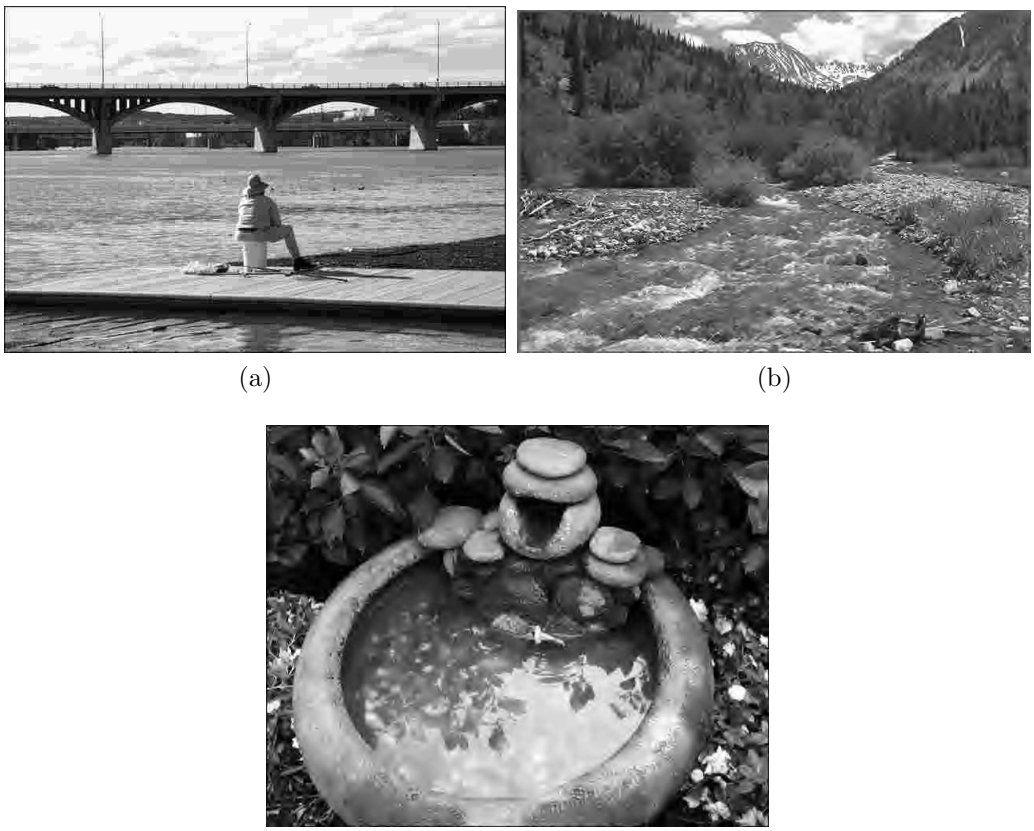


Figure 6.6: Scatter plots of DMOS versus image quality scores computed by SA-SS and SA-ZC. Each point, marked by asterisk or "+", represents one test image with "+" denoting outliers. The curve corresponds to the logistic function (3.9) with parameters fitted over dataset. (a) SA-SS for white noise. (b) SA-ZC for Gaussian blur. (c) SA-SS for JPEG compression. (d) SA-ZC for JPEG2000 compression.

SA-SS and SA-ZC are not able to identify the origin of the SA at the boundaries of coding blocks which can be either compression distortion or the real image content. Since so many horizontal fine structures of the waves in Figure 6.7a are very likely to appear at the boundaries of coding blocks in part, it should come as little surprise that SA-SS under-predicts image quality in this situation.



(a) (b)

(c)

Figure 6.7: Sample images corresponding to the outlier marked in Figure 6.6. (a) The "Man fishing" JPEG image compressed to 0.4229 bpp. (b) The "Stream" JPEG2000 image compressed to 0.1920 bpp. (c) The "Coins in fountain" JPEG2000 image compressed to 0.1874 bpp.

With respect to JPEG2000 compression, an important part of outliers are due to the compressed "Stream" image, for which a sample image is shown in Figure 6.7b. Note that the stones at the banks are a kind of quasi-texture where the perceived distortion is not as disturbing as elsewhere, e.g., the bush and trees, due to the masking properties of the HVS. The SA measure fails to recognize textures and over-predicts image quality in this case. Another significant part of outliers result from the compressed "Coins in fountain" images, a sample image of which is shown in Figure 6.7c. The HVS can easily recognize that the blur of reflection

and coins in the water is image content rather than distortion, whereas the SA measure can not recognize this and under-predicts image quality in this scenario.

To this end, the investigation into outliers provides us more insights into the SA measure. Despite the promising performance of the SA measure, it is observed that an NR image quality assessment model based on SA alone is not adequate for universally accurate image quality prediction. It is desirable to take into account the properties of the HVS, especially the role of high-level cognitive understanding in visual perception. In fact, the issues exposed here are not limited to the SA measure alone but intrinsic to most NR image quality measures. With increased knowledge of the HVS, a better emulation of the HVS would help to further reduce the gap between the SA measure and human visual perception for more accurate image quality prediction.

6.5 Summary

Under the assumption that human visual perception is highly sensitive to the structural information in a scene, an SA-based framework is proposed for NR image quality assessment in this chapter. The proposed framework predicts image quality based on the quantification of SA information of different visual significance. Despite being a new concept, SA is shown to be an implicit idea underlying a variety of existing NR image quality measures. With both strengths and weaknesses of those SA-motivated image quality measures in mind, a new model of an SA measure is developed. The components of the SA measure include SA weight, local SA, and global SA. SA weight and local SA are evaluated pixel-by-pixel over the entire image, and global SA combines the resulting two maps into a single numerical

value as image quality score. The motivation of the SA weight is to automatically adjust the relative contributions among the local SA information according to visual significance. In this study, SA weight describes visual significance from the structural perspective; a larger SA weight is assigned to the spatial location where the associated SA information is more likely to attract visual attention. As two specific examples, SA weight is expressed in terms of structure strength (SAW-SS) and zero-crossing activity (SAW-ZC). For SAW-SS, an MMF-based approach is proposed with the intention to indirectly reinforce the detailed structures of relatively low strength. As to the model component of local SA, a feature of direction spread is proposed. The basic principle of direction spread measurement is to track the pixels with monotonically-changing intensities along the horizontal and vertical directions at a general pixel. The feature of direction spread does not require edge extraction, overcomes the limitations imposed by structure or feature extraction of severely distorted images, and allows more complete SA information to be preserved. Additionally, the local SA at the coding block boundaries is also defined in the case of JPEG compression. For the model component of global SA, a visual significance-weighted pooling approach is proposed. The specific expressions of the pooling approach are developed for different distortion types. A common feature of these expressions is that they are parameterless, which avoids additional procedures or training data for parameter determination. It is demonstrated that the SA measure predicts image quality in close consistency with subjective quality scores over a variety of distortions, including white noise, Gaussian blur, and JPEG and JPEG2000 compression. We consider the model of the SA measure together with the proposed examples of its implementation in this study as an encouraging starting point. The effort in this study may motivate the further exploration of

the concept of SA and drive the success of other SA-motivated NR image quality measures in the future.

Chapter 7

Conclusion and Future Work

7.1 Summary

This thesis focuses on the development of NR image quality measures. The goal is to develop computational models to quantitatively predict perceived quality solely from a distorted image without any reference to a distortion-free image. Because of its intrinsic difficulty, the field of NR image quality assessment is still in its infancy and is full of challenges.

Due to the widespread use of image compression technologies in numerous applications nowadays, compression has become a significant source of image distortions. The distortions created by the widely used JPEG2000 compression are heavily dependent on image content without regularity to follow and make the task of NR image quality assessment complicated. A study of existing NR quality measures for JPEG2000 compressed images reveals the following limitations. In the case of structures/features-based NR image quality measures, their effectiveness would inevitably decline when images are subject to severe distortion. It is mainly due to the decreased number of detected structures/features with increased distortion.

The thresholding operation associated with the structure/feature extraction is also experience/image-dependent and would be inconvenient in practical applications. Another concern comes from those NR image quality measures that involve parameters to be determined. In some image quality measures, parameters are calibrated in an *ad hoc* manner without clear rules. And in some other image quality measures, parameters have to be learned from subjective quality data. In this case, the resulting additional procedures for parameter learning usually increase implementation complexity, while the need for training data is an obvious constraint when data are limited. With these limitations in mind, kurtosis-based NR image quality measures were presented in Chapter 4 and a pixel activity-based NR image quality measure was presented in Chapter 5, which all target the distortions resulting from JPEG2000 compression. One important objective in the design of these NR image quality measures is to achieve low complexity and easy implementation without sacrificing effectiveness. There is little doubt that a sophisticated quality measure is advantageous in achieving good results. However, the implementation expense is usually increased. With both satisfactory performances and easy implementation, the developed NR quality measures prove effective alternatives to those complicated quality measures for JPEG2000 compressed images.

General-purpose NR image quality assessment applicable to all kinds of distortions is an extremely difficult task. We seek to approach the goal of general purpose by developing an SA-based framework in Chapter 6. With an intensive investigation into the current techniques, the implicit idea of SA is generalized from a variety of existing NR image quality measures, and a large part of the reported NR image quality measures are shown to be motivated by the concept of SA. Taking into account both merits and demerits revealed from those SA-motivated NR

image quality measures, a general model of an SA measure is developed to blindly predict image quality. Different from the NR image quality measures presented in Chapters 4 and 5, which are developed for the particular distortions resulting from JPEG2000 compression, the SA measure provides a more general model with its scope extended to a number of distortion types. Specifically, the effectiveness of the SA measure is validated in the context of white noise, Gaussian blur, and JPEG and JPEG2000 compression in this thesis. We believe the effective implementation of the SA measure or a hybrid image quality measure combing the SA measure and other suitable measures could extend the scope to a broader class of distortion types as well as to multiple distortions.

All the NR image quality measures developed in this thesis are able to work with the distortion created by JPEG2000 compression. The pixel activity-based image quality measure presented in Chapter 5 gives the best performances, followed by the kurtosis-based image quality measures in Chapter 4 and the implementations of the SA measure in Chapter 6. The kurtosis-based image quality measures and the implementations of the SA measure generally perform similarly.

7.2 Contributions

Kurtosis-based NR image quality measures were presented in Chapter 4. They are developed based on either 1-D or 2-D kurtosis in the DCT domain of general image blocks. The major contributions of this study include the following aspects:

- To the best of our knowledge, 1-D kurtosis-based image quality measures do not exist in the current literature. Although the validity of kurtosis-based image quality measures has been established in the previous studies, the

former work fails to clearly demonstrate how kurtosis works in image quality prediction. In this study, we construct frequency distributions based on either frequency bands or DCT basis functions for the calculation of 1-D kurtosis in the DCT domain. Based on the proposed 1-D kurtosis, we explicitly demonstrate the working principle of kurtosis in image quality prediction.

- The proposed image quality measures are simple and easy to implement; they work with general image blocks and overcome the limitations imposed by structure/feature extraction of distorted images; they do not involve parameters and avoid additional procedures and training data for parameter determination; and they are proven effective in predicting perceived visual quality of JPEG2000 compressed images.

- The investigation into outliers provides more insights into the proposed image quality measures and sheds light on future directions.

A pixel activity-based NR image quality measure was presented in Chapter 5. Instead of extracting structures/features from an image, the proposed NR quality measure predicts image quality based on the pixel activity in terms of the MC and the ZC activity. The major contributions of this study include the following aspects:

- The proposed NR image quality measure is developed with reasonable computation expense and easy implementation.

- With image quality predicted based on the basic activity of general pixels, the proposed NR image quality measure overcomes the limitations imposed by structure/feature extraction of distorted images.

- The proposed structural content-weighted pooling approach does not involve parameters from the HVS, training data, or stabilizing constants. This avoids any additional procedures and training data for parameter determination.

- The proposed expressions of pixel activity for image quality prediction is corroborated through both qualitative and quantitative evaluations.

- The proposed image quality measure achieves consistently close correlation with subjective quality scores when the processing block size is subject to a wide range. The performance of the proposed NR image quality measure is shown better than the classical FR quality measure of PSNR, comparable to the representative FR quality measure of SSIM index, and competitive among the state-of-the-art NR image quality measures.

- The investigation into outliers provides a good understanding of the underlying deficiencies and points out the direction for the future work.

Assuming that human visual perception is highly sensitive to the structural information in a scene, we propose to quantify SA as a good approximation of perceived visual quality. An SA-based framework for NR image quality assessment together with a model named the SA measure was presented in Chapter 6. The major contributions of this study include the following aspects:

- The SA measure blindly predicts image quality based on the comprehensive SA information of different visual significance in an image. The model does not require structure/feature extraction and overcomes the limitations imposed by structure/feature extraction of severely distorted images.

- The model of the SA measure allows flexible implementations. As specific examples, the feature of direction spread was proposed to implement the model component of local SA and the features of structure strength and zero-crossing activity to implement the model component of SA weight.

- The SA measure does not involve any parameters from the HVS, training data, or stabilizing constants. It is independent of any additional procedures or training data for parameter determination.

- The SA measure is validated with a variety of distortions including white noise, Gaussian blur, and JPEG and JPEG2000 compression. The effectiveness of the SA measure is demonstrated in the comparison with subjective quality scores as well as the classical FR quality measure of PSNR and the representative FR quality measure of SSIM index.

- The investigation into outliers in image quality prediction highlights the importance of incorporating HVS features into the SA measure for the further improvement in the accuracy of image quality prediction.

7.3 Future Work

A number of issues deserve further investigation.

First, the incorporation of suitable HVS features or an appropriate HVS model at acceptable complexity may further improve the performances of the NR image quality measures presented in this thesis. As indicated in the outlier analysis in Chapters 4, 5, and 6, the incorporation of relevant properties of the HVS or an appropriate HVS model could help to overcome outliers and allow more accurate

image quality prediction. However, this may increase computational complexity. The parameters introduced during the modeling of the HVS, which usually require additional procedures or training data for parameter determination, may also increase implementation complexity. It thus could be worthwhile investigating how best to incorporate suitable HVS features or an appropriate HVS model at acceptable complexity without compromising the overall performance.

Second, with respect to the kurtosis-based NR image quality measures presented in Chapter 4, current image quality prediction is based on the quantification of the blurring distortion alone. A direct extension is to include the quantification of ringing artifacts. It is also desirable to study how to incorporate appropriate spatial features to compensate for the over-sensitivity of kurtosis so as to further narrow the gap between the quantitative quality measures and human visual perception. Additionally, other approaches of constructing frequency distribution for the computation of kurtosis could also be a direction that is worth further investigation.

Third, with respect to the pixel activity-based NR image quality measure presented in Chapter 5, the expression of pixel activity in image quality prediction is not restricted to the expressions proposed in this study. Further exploration into other suitable expressions of pixel activity may contribute to the further improvement in performance.

Fourth, with respect to the SA-based NR image quality assessment framework presented in Chapter 6, it would be beneficial to incorporate an additional model component for the identification of distortion type. Despite a step towards general-purpose NR image quality assessment, the SA measure still assumes the prior knowledge of distortion type when predicting image quality. The incorporation of

an additional model component to identify distortion type could be a promising direction to allow a more general quality assessment model.

Finally, it would also be interesting to develop hybrid quality measures that combine the image quality measures presented in this thesis with other suitable quality measures to extend the scope to other single-distortion types as well as to multiple distortions. Furthermore, the NR image quality measures proposed in this thesis have the potential to be extended to color images and video sequences.

Bibliography

[1] S. Winkler, *Digital Video Quality - Vision Models and Metrics.* John Wiley & Sons, Ltd, 2005.

[2] H. R. Wu and K. R. Rao, *Digital Image Video Quality and Perceptual Coding.* CRC, 2005.

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

[4] B. A. Wandell, *Foudations of Vision.* Sinauer Associates, 1995.

[5] L. K. Cormack, "Computational models of early human vision," in *Handbook of Image and Video Processing*, 2nd ed., A. C. Bovik, Ed. Elsevier Academic Press, 2005.

[6] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*, A. B. Watson, Ed. MA: MIT Press, 1993, pp. 207–220.

[7] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Communications*, vol. 43, no. 12, pp. 2959–2965, 1995.

[8] S. Winkler, "A perceptual distortion metric for digital color video," in *Proc. SPIE*, vol. 3644, 1999, pp. 175–184.

[9] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, Orlando, FL, 2002, pp. 3313–3316.

[10] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it?" *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[11] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, A. C. Bovik, Ed. New York: Academic Press, 2000.

[12] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Processing*, vol. 78, pp. 231–252, 1999.

[13] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: The MIT Press, 1993, pp. 163–178.

[14] J.Lubin, "A visual discrimination mode for image system design and evaluation," in *Visual Models for Target Detection and Recognition*, E. Peli, Ed. Singapore: World Scientific Publishers, 1995, pp. 207–220.

[15] JNDmetrix Technology Sarnoff Corp., "Sarnoff JND Metrix, Evaluation Version," 2003 [Online], available: http://www.sarnoff.com/products-services/video-vision/jndmetrix/downloads.asp.

[16] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[17] E. Peli, "Contrast in complex images," *Journal of Optical Society of America*, vol. 7, no. 10, pp. 2032–2040, 1990.

[18] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.

[19] P. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. SPIE*, vol. 2179, 1994, pp. 127–141.

[20] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, 1994, pp. 982–986.

[21] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multi-scale transforms," *IEEE Trans. Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.

[22] A. B. Watson, "DCTune: A technique for visual optimization of DCT quantization matrices for individual images," in *Society for Information Display Digest of Technical Papers*, vol. XXIV, 1993, pp. 946–949.

[23] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 713–724, 1995.

[24] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Trans. Communications*, vol. 46, no. 9, pp. 1215–1226, 1998.

[25] S. Winkler, "A perceptual distortion metric for digital color images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Chicago, 1998, pp. 399–403.

[26] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 4, no. 4, pp. 636–650, 2000.

[27] Z. Wang, A. C. Bovik, and L. Lu, "Foveated wavelet image quality index," in *Proc. SPIE Int. Society for Optical Engineering*, vol. 4472, 2001, pp. 42–52.

[28] W. Lin, L. Dong, and P. Xue, "Visual distortion gauge based on discrimination of noticeable contrast changes," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 7, pp. 900–909, 2005.

[29] X. H. Zhang, W. S. Lin, and P. Xue, "Improved estimation for just-noticeable visual distortion," *Signal Processing*, vol. 85, pp. 795–808, 2005.

[30] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[31] "Visual Signal-to-Noise Ratio," [Online], available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[33] Z. Wang, "The SSIM Index for Image Quality Assessment," [Online], available: http://www.ece.uwaterloo.ca/ z70wang/research/ssim/ssim_index.m.

[34] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[35] Z. Wang, "A Universal Image Quality Index," [Online], available: http://www.cns.nyu.edu/ zwang/.

[36] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, vol. 2, Asilomar, CA, 2003, pp. 1398–1402.

[37] Z. Wang, "Multi-Scale Structural Similarity Index," [Online], available: http://live.ece.utexas.edu/research/quality/ms_ssim.zip.

[38] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Philadephia, 2005, pp. 573–576.

[39] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.

[40] H. R. Sheikh, "An Information Fidelity Criterion for Image Quality Assessment," [Online], available: http://live.ece.utexas.edu/research/quality/ifcvec_release.zip.

[41] H. R. Sheikh and A. C. Bovik, "Image informatoin and visual quality," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.

[42] H. R. Sheikh, "Visual Information Fidelity Measure for Image Quality Assessment," [Online], available: http://live.ece.utexas.edu/research/quality/vifvec_release.zip.

[43] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in wavelet domain," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[44] D. V. Weken, M. Nachtegael, and E. E. Kerre, "Using similarity measures and homogeneity for the comparison of images," *Image and Vision Computing*, vol. 22, no. 9, pp. 695–702, 2004.

[45] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based grayscale image qualty measure for local and global assessment," *IEEE Trans. Image Processing*, vol. 15, no. 2, pp. 422–429, 2006.

[46] H. S. Han, D. O. Kim, and R. H. Park, "Structural information-based image quality assessment using LU factorization," *IEEE Trans. Consumer Electronics*, vol. 55, no. 1, pp. 165–171, 2009.

[47] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human Vision*

and Electronic Imaging X, Proc. SPIE, vol. 5666, San Jose, CA, 2005, pp. 149–159.

[48] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[49] S. G. Mallat, "Multifrequency channel decomposition of images and wavelet models," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2091–2110, 1989.

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

[51] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1409–1423, 2009.

[52] "VQEG: The Video Quality Experts Group," [Online], available: http://www.vqeg.org.

[53] G. K. Wallace, "The JPEG still picture compression standard," *IEEE Trans. Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[54] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Processing*, vol. 70, no. 3, pp. 247–278, 1998.

[55] H. R. Wu and M. Yuen, "A generalized block-edge impairment metric for video coding," *IEEE Signal Processing Letters*, vol. 4, no. 11, pp. 317–320, 1997.

[56] L. Meesters and J. B. Martens, "A single-ended blockiness measure for JPEG coded images," *Signal Processing*, vol. 82, no. 3, pp. 369–387, 2002.

[57] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2002, pp. 477–480.

[58] X. Li, "Blind image quality assessment," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Rochester, 2002, pp. 449–452.

[59] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, and X. Yang, "A locally adaptive algorithm for measuring blocking artifacts in images and videos," *Signal Processing: Image Communication*, vol. 19, no. 6, pp. 499–506, 2004.

[60] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 2000, pp. 981–984.

[61] S. Liu and A. C. Bovik, "Efficient DCT-domain blind measurement and reduction of blocking artifacts," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 12, no. 12, pp. 1139–1149, 2002.

[62] T. Brandao and M. P. Queluz, "No-reference image quality assessment based on DCT-domain statistics," *Signal Processing*, vol. 88, no. 4, pp. 822–833, 2008.

[63] M. Jung, D. Leger, and M. Gazalet, "Univariant assessment of the quality of images," *Journal of Electronic Imaging*, vol. 11, no. 3, pp. 354–364, 2002.

[64] P. Gastaldo and R. Zunino, "Neural networks for the no-reference assessment of perceived quality," *Journal of Electronic Imaging*, vol. 14, no. 3, pp. 033 004–1–11, 2005.

[65] R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Processing*, vol. 87, no. 6, pp. 1493–1503, 2007.

[66] J. B. Martens, "The Hermite transform - theory," *IEEE Trans. Acoustics Speech Signal Processing*, vol. 38, no. 9, pp. 1595–1618, 1990.

[67] A. Skodras, C. Christopoulos, and T. Ebrahimi, "The JPEG2000 still image compression standard," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 36–58, 2001.

[68] S. H. Oguz, Y. H. Hu, and T. Q. Nguyen, "Image coding ringing artifact reduction using morphological post-filtering," in *Proc. IEEE 2nd Workshop on Multimedia Signal Processing*, 1998, pp. 628–633.

[69] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[70] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 163–172, 2004.

[71] E. P. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang, and F. Moschetti, "A no-reference quality metric for measuring image blur," in *Proc. IEEE 7th Int. Symp. Signal Processing and Its Applications*, vol. 1, 2003, pp. 469–472.

[72] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[73] H. Tong, M. Li, H. J. Zhang, and C. Zhang, "No-reference quality assessment for JPEG2000 compressed images," in *Proc. IEEE Int. Conf. Image Processing*, vol. 5, 2004, pp. 3539–3542.

[74] R. Barland and A. Saadane, "Reference free quality metric for JPEG2000 compressed images," in *Proc. IEEE 8th Int. Symp. Signal Processing and Its Applications*, vol. 1, 2005, pp. 351–354.

[75] H. R. Sheikh and A. C. Bovik, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1918–1927, 2005.

[76] H. R. Sheikh, "No Reference Image Quality Assessment for JPEG2000," [Online], available: http://live.ece.utexas.edu/research/quality/jp2knr_release.zip.

[77] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Trans. Image Processing*, vol. 8, no. 12, pp. 1688–1701, 1999.

[78] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, "No reference image quality assessment for JPEG2000 based on spatial features," *Signal Processing: Image Communication*, vol. 23, no. 4, pp. 257–268, 2008.

[79] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits and Syetems for Video Technology*, vol. 20, no. 4, pp. 529–539, 2010.

[80] H.Liu, N. Klomp, and I. Heynderickx, "A perceptually relevant approach to ringing region detection," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1414–1426, 2010.

[81] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Computer Vision*, 1998, pp. 836–846.

[82] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.

[83] "Blind Image Quality Index," [Online], available: http://live.ece.utexas.edu/research/quality/BIQI_release.zip.

[84] A. Srivastava, A. Lee, E. Simoncelli, and S. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 17–33, 2003.

[85] V. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Germany: Springer Verlag, 2000.

[86] M. A. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 6, pp. 583–586, 2010.

[87] "BLIINDS: BLind Image Integrity Notator using DCT Statistics," [Online], available: http://live.ece.utexas.edu/research/quality/BLIINDS.zip.

[88] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2," 2005 [Online], available: http://live.ece.utexas.edu/research/quality/.

[89] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3441–3452, 2006.

[90] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase I," 2000 [Online], available: http://www.vqeg.org/.

[91] VQEG , "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II," 2003 [Online], available: http://www.vqeg.org/.

[92] VQEG, "Final Report from the Video Quality Experts Group on the Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, phase I," 2009 [Online], available: http://www.vqeg.org/.

[93] G. A. F. Seber and C. J. Wild, *Nonlinear regression*. Wiley, 2003.

[94] N. F. Zhang, M. T. Postek, R. D. Larrabee, A. E. Vladar, W. J. Keery, and S. N. Jones, "Image sharpness measurement in the scanning electron microscope-part III," *Scanning*, vol. 21, no. 4, pp. 246–252, 1999.

[95] J. Caviedes and F. Oberti, "A new sharpness metric based on local kurtosis, edge and energy information," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 147–161, 2004.

[96] L. T. DeCarlo, "On the meaning and use of kurtosis," *Psychological Methods*, vol. 2, no. 3, pp. 292–307, 1997.

[97] J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Processing Letters*, vol. 10, no. 10, pp. 289–292, 2003.

[98] A. K. Moorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193–201, 2009.

[99] H. Liu and I. Heynderickx, "Studying the added value of visual attention in objective image quality metrics based on eye movement data," in *Proc. IEEE Int. Conf. Image Processing*, 2009, pp. 3097–3100.

[100] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE Int. Conf. Image Processing*, 2006, pp. 2945–2948.

[101] D. L. Donoho and A. G. Flesia, "Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics?" *Network: Computation in Neural Systems*, vol. 12, no. 3, pp. 371–393, 2001.

[102] W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, 1998, pp. 414–418.

[103] Z. K. Lu, W. S. Lin, X. K. Yang, E. P. Ong, and S. Yao, "Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1928–1942, 2005.

[104] G. R. Arce and R. E. Foster, "Detail-preserving ranked-order based filters for image processing," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 37, no. 1, pp. 83–98, 1989.

[105] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 5, pp. 886–890, 2007.

Appendix A

Publications

J. Zhang, T. M. Le, "A new no-reference quality metric for JPEG2000 images," *IEEE Trans. Consumer Electronics*, vol. 56, no. 2, pp. 743-750, 2010.

J. Zhang, S. H. Ong, and T. M. Le, "Kurtosis-based no-reference quality assessment of JPEG2000 images," *Signal Processing: Image Communication*, vol. 26, no. 1, pp. 13-23, 2011.

J. Zhang, T. M. Le, S. H. Ong, and T. Q. Nguyen, "No-reference image quality assessment using structural activity," *Signal Processing*, vol. 91, no. 11, pp. 2575-2588, 2011.