Information Theoretic Criteria for Image Quality Assessment Based on Natural Scene Statistics

by

Di Zhang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2009

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

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

Abstract

Measurement of visual quality is crucial for various image and video processing applications. It is widely applied in image acquisition, media transmission, video compression, image/video restoration, etc.

The goal of image quality assessment (QA) is to develop a computable quality metric which is able to properly evaluate image quality. The primary criterion is better QA consistency with human judgment. Computational complexity and resource limitations are also concerns in a successful QA design. Many methods have been proposed up to now. At the beginning, quality measurements were directly taken from simple distance measurements, which refer to mathematically signal fidelity, such as mean squared error or Minkowsky distance. Lately, QA was extended to color space and the Fourier domain in which images are better represented. Some existing methods also consider the adaptive ability of human vision. Unfortunately, the Video Quality Experts Group indicated that none of the more sophisticated metrics showed any great advantage over other existing metrics.

This thesis proposes a general approach to the QA problem by evaluating image information entropy. An information theoretic model for the human visual system is proposed and an information theoretic solution is presented to derive the proper settings. The quality metric is validated by five subjective databases from different research labs. The key points for a successful quality metric are investigated. During the testing, our quality metric exhibits excellent consistency with the human judgments and compatibility with different databases. Other than full reference quality assessment metric, blind quality assessment metrics are also proposed. In order to predict quality without a reference image, two concepts are introduced which quantitatively describe the inter-scale dependency under a multi-resolution framework. Based on the success of the full reference quality metric, several blind quality metrics are proposed for five different types of distortions in the subjective databases. Our blind metrics outperform all existing blind metrics and also are able to deal with some distortions which have not been investigated.

Acknowledgements

I am deeply grateful to Professor Edward Jernigan from the Department of Systems Design Engineering at the University of Waterloo. It is my honor to have him as my supervisor. In the past five years, he guided my study on image quality assessment with the consistent support. He also provided the tremendous flexibility to let me try anything I want. I gained varied experiences but not only limited to my research area. He is a true educator.

I would like to thank Professors Paul Fieguth, Zhou Wang, David Clausi and Eric Dubois. I do appreciate their precious advices on my research and comments on this dissertation. I would also like to thank the quality assessment community for providing the subjective assessment data for this research.

I sincerely acknowledge my parents for their love, continuous support, suggestions and always believing in me through my long journey of science and education. I would also like to thank my wife, Yu Shen, for comforting my agitation and rage when I faced difficulties.

This research was supported in various ways by grants and/or scholarships from the National Sciences and Engineering Council, the Faculty of Engineering and the Department of Systems Design Engineering. I am truly appreciative of these aids.

Dedication

This is dedicated to the one I love.

Contents

List of Tables

List of Figures

List of Symbols

Syntax	Definition
X	scalar, random variable
$\hat{X}$	estimate of a random variable X
$\underline{X}$	column vector, random vector
$\underline{\hat{X}}$	estimate of a random vector$\underline{x}$
X_i	the ith element of column vector,random vector
$\hat{X}_i$	estimate of the ith element of a random vector $\underline{X}$
$\underline{X_i}$	the ith column vector, random vector
$\underline{X_i}_j$	the jth element in the ith column vector, random vector
$\underline{\hat{X}_i}$	estimate of the ith random vector
$\mathbf{X}$	matrix, covariance matrix
$\mathbf{X}_i$	the ith matrix
$g(\omega; \mu, \sigma)$	Gaussian distribution
$g(\underline{\omega}; \underline{\mu}, \mathbf{\Sigma})$	joint Gaussian distribution, $\mathbf{\Sigma}$ is the covariance matrix.
$p(\omega)$	the probability of ω
$f(\omega)$	the probability density of ω
$\mathbf{cov}(X, Y)$	the covariance between two random variables X and Y

Chapter 1

Introduction

1.1 Image Quality

For a long time people have been eager to understand how the human vision system responds to the visual world. Among many vision functions, comparison is a natural capability that everyone takes for granted. With this ability, people can notice the differences between two similar images. But teaching a computer to "perceive" the differences in a human manner is frustrating. There are several reasons that limited success has been achieved [47]. First, image comparison by a human entity is a subjective action. The judgment will be affected by individual physiological limits, observation behavior, personal preference, and knowledge limitation. All of these imply that simple distance metrics are not able to evaluate the perceptual differences. Second, distortion variety poses difficulties for subjective evaluation. Research [2] shows that observers could become familiar with the stimuli (such as image distortion) in order to make judgment results more repeatable. But interpretation of certain distortions with varied image contents is highly subjective. Also people usually do not quantify the difference but evaluate the difference in a fuzzy manner. Third, accumulating individual perceptible differences and combining them is a subjective process. The whole procedure of difference detection, quantification and accumulation to a single score is referred to as image quality assessment.

Why has so much effort gone into developing reliable methods for evaluating image quality? Because quality assessment metrics are widely used in every image processing area where human perception is involved.

First, most image acquisition, storage or communication systems need quality as-

sessment metrics. Images are acquired by choosing from a wide variety of possible methods according to which method gives the best visual quality. Quality assessment is also used in storage systems or digital television broadcasting. Due to limited storage capacity or channel bandwidth, raw data are usually processed by compression technologies. The basic idea of lossy compression is to reduce spatial redundancy and discard high frequency information. The question is how much information should be kept for acceptable visual quality.

Second, quality assessment metrics are used to evaluate the performance of image processing technologies, especially in image restoration areas, such as denoising. With quality assessment metrics project developers are able to choose the best solutions for their specific tasks among numerous existing methods.

Third, optimization of image processing systems usually needs quality assessment metrics. For power limitations or computation resource constraints, such as real-time application, algorithms need to be simplified to make systems run smoothly. Unfortunately this simplification usually weakens system performance. Under this circumstance, reliable quality assessment metrics will be helpful in choosing the proper modifications.

The quality assessment is critical aspect of most image processing systems. The main purpose of this thesis is to propose the reliable quality prediction methods under different circumstances, specifically, with or without reference image.

1.2 Subjective Image Quality Assessment

The straightforward way of measuring image quality is to consult human opinions because people are the ultimate viewers in most image applications. This is known as subjective image quality assessment. The basic idea of subjective quality assessment is to ask several observers to give their personal judgments for a given image test sample and then compute the mean opinion score (MOS) of the human judgments as the subjective evaluation.

Subjective testing for visual quality assessment has been defined in ITU-R Recommendation BT.500-10 (2000) [2]. The three most commonly used procedures are:

- *Double Stimulus Continuous Quality Scale* (DSCQS) [2]
- *Double Stimulus Impairment Scale* (DSIS) [2]

- *Single Stimulus Continuous Quality Evaluation* (SSCQE) [2]

The databases LIVE [99], IVC [64] and TOYAMA [46] use double stimulus continuous quality scale method for subjective testing, since DSCQS is the most reliable method shown in contextual effects study [47]. Several issues should be considered during the subjective quality assessment experiment. First, sufficient observers should be included in the experiment. At least 15 observers are recommended in [2]. Even though the observers are urged to base their score only on how different or dissimilar they perceive the images to be, the personal preference effect on evaluation results cannot be totally eliminated. More observers will help to achieve a quality score that most people agree with. Also "nonexpert" observers and "expert" observers need to be included in the observer group. A nonexpert viewer may pay more attention to the larger context, but a trained viewer may focus on the details or evaluate image quality from professional perspective. Second, to conduct appropriate subjective assessment, it is necessary to set up the proper viewing conditions which include display device requirements and viewing manner. For instance, the ratio of luminance of inactive screen to peak luminance should be less than 0.02. The viewing distance is recommended to be 1.5m, which corresponds to a viewing distance/display height ratio of 6. In addition, other procedural elements are worth noting, such as: timing of presenting test samples and selection of samples. These weaknesses result in the request for a fast, computational and reliable quality assessment method.

1.3 Outline of Thesis

The thesis is organized as follows. Chapter 1 gives a general introduction about the motivation and subjective image quality assessment. The past work on quality assessment and human vision properties is reviewed in Chapter 2. Chapter 3 talks about wavelet and image statistics, which are applied to construct an information theoretic framework for representing an image. In Chapter 4, an information theoretic framework with a human vision model will be proposed. Chapter 5 will use the framework to approach objective quality assessment with access to reference image. Chapter 6 proposes several image quality assessment methods without using reference image. Future work will be discussed in Chapter 7.

Chapter 2

The Human Visual System and Past Work

2.1 Human Visual System

The goal of image quality assessment is to simulate human assessment of image dissimilarities. Although the full functions of the human visual system, especially high level functions, are not thoroughly investigated, knowledge of the early stage of the human visual system is still helpful in understanding fundamental visual behaviors. Some human vision behaviors are modeled to simulate the human assessment of visual quality. Four of them are popular for the design of quality assessment.

1. Luminance Sensitivity
In early research, experiments results [12] showed that the human vision system responds to luminance change in a logarithmic way. Consider a light area with luminance of $I + \triangle I$ surrounded by a background with luminance of I. The just noticeable difference $\triangle I$ is proportional to the background luminance as in Equation 2.1. Several quality assessment methods [77, 121, 122, 130] incorporate this feature in their frameworks.

$$\triangle I \propto I \tag{2.1}$$

Later research [13] shows that human visual system luminance sensitivity not only depends on the overall background luminance but also the local neighborhood luminance.

2. Contrast Sensitivity
The famous Campbell-Robson [20] chart was proposed to illustrate that the human

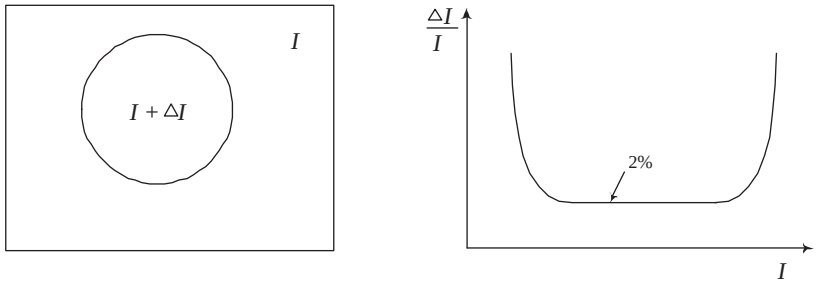


Figure 2.1: Luminance sensitivity. The just noticeable luminance difference $\triangle I$ is proportional to the background luminance I within a wide range.The experiment results show that the ratio is nearly constant at 0.02.

visual system has different responses to visual signals at different frequencies, which is referred to as contrast sensitivity. Campbell and Robson generated a chart (Figure 2.2) in which spatial frequency increases (the bars get thinner) in the horizontal direction and contrast decreases (the black/white bars merge to plain gray area) in the vertical direction. The envelop of distinguishable bars illustrates visual sensitivity to different frequencies. In theory, human vision has a cut-off frequency which equals the smallest bars that people cannot detect even with the largest contrast. Many quality assessment methods adopt this feature. Among them, Mannos and Sakrison [72] first proposed a linear spatial invariant operation $a(.)$ applied on the raw image to simulate the contrast sensitivity, or *just noticeable difference* corresponding to certain frequency. The response $A(f)$ of the operation $a(.)$ is given by Campbell and Robson in Equation 2.2. The equation describes a bandpass filter, which has maximum gain at 7.891 cycles/degree and bandwidth is 12.922 cycles/degree.

$$A(f) = 2.6(0.0192 + 0.114f)e^{-(0.114f)^{1.1}} \tag{2.2}$$

where f is in cycles/degree.

3. Vision Channel Decomposition

Current consensus [14, 72] seems to be that the primary cortex is sensitive to different orientations, frequencies and colors carried by visual signals. This suggests the visual area functions similarly to filters with orientation and frequency selectivity. Hubel and Wiesel [14] deduced that simple cells are sensitive to specific orientations with approximate bandwidth 30^o. Along with contrast sensitivity in different frequencies, Mannos and Sakrison [72] suggested that the human visual system contains a number of independent channels with narrow bandwidths and specific center frequencies. Experiments show the frequency bandwidth of simple cells is about one octave. Usually

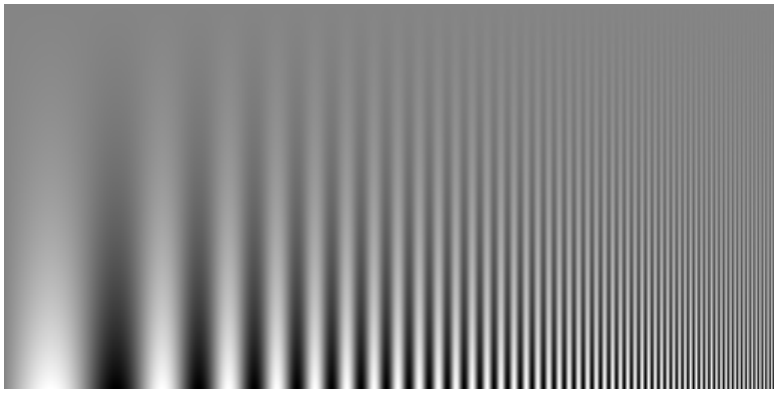


Figure 2.2: Contrast sensitivity chart (Campbell-Robson Chart). Spatial frequency increases (the bars get thinner) in the horizontal direction and contrast decreases (the black/white bars merge to plain gray area) in the vertical direction. The envelop of distinguishable bars illustrates the visual sensitivity to different frequencies

people use wavelet filters [17, 62] and direct Fourier transform [130] (only select certain frequencies, the bandwidth is 0) to decompose visual field into multi-channels.

4. Contrast Masking
Contrast Masking happens when different image components are present at the same spatial location. A simple example below gives a clear illustration. Figure 2.3(a) is a raw image. The top half of Figure 2.3(b) has extra white noise, whereas the bottom half of Figure 2.3(c) has the same amount of white noise. Generally people are more likely to notice the noise in the middle figure than the right. The presence of one component (grass) decreases the visibility of another component (noise) in Figure 2.3(c). Usually, the contrast masking effect is strongest when two components have similar frequency and orientation [65]. Based on this statement, masking is typically modeled within each channel. This vision property is also widely adopted in many quality assessment methods [17, 77, 121, 122, 130].

2.2 Quality Assessment Methods Review

Image quality assessment was originally designed for compression purposes. MSE or PSNR is computed to indicate the error created by different compression methods with different compression ratios. Although MSE or PSNR could be taken as a quality indication, these fundamental quality metrics are not consistent with human

(a) (b) (c)

Figure 2.3: Masking effect. Figure (a) is the raw image. Figure (b) has extra white noise in the top half. Figure (c) has the same amount of white noise in the bottom half.

judgments [9,37,125]. Later, people [17,72,77,130] designed quality assessment metrics by modeling human visual properties. Recently, quality assessment [15,35,69,97,121–123,126] was extended to images with varied distortions, such as noise, blur, communication distortion. Depending on reference accessibility, quality assessment methods can be classified into two broad categories.

Full-reference Quality Assessment methods [15,17,35,38,43,69,72,77,97,110,115,117,121–123,130] [66,118,132,139], in which the quality assessment method has full access to the reference image. Most existing quality assessment solutions belong to this category. Modern quality assessment metrics usually propose nonlinear models to predict human visual reactions to distortions and convert the reactions into the quality score, which are summarized in Table 2.1.

No-reference/Blind Quality Assessment methods [21,22,22,44,52,73,96,98,111,119,124,137], in which the quality assessment method has access only to the distorted image. Up to now, the solution has two different directions. One is to evaluate artificial effects due to compression, especially for JPEG. Most blind quality methods [68,111,119,124] take this direction. One approach [96] tried to exploit experimental knowledge of natural scene statistics.

There is another type of quality measure [8,23,24,27,49,59–61,74,109,127]: the reduced reference quality assessment, which aims to predict the visual quality of distorted images with only partial information about the reference images. The reduced reference quality assessment was proposed as the substitute for blind quality assessment as people [96] thought distortion is difficult to evaluate without reference information. Partial reference image will help us to understand how distortion affects the entire image if the distortion is globally applied. For local distortion, such as the fastfading distortion, it

is dangerous to estimate the distortion using only partial reference image. This thesis focuses on full-reference and blind quality assessments.

2.2.1 Existing Quality Assessment Methods

In this section, the past work on image quality assessment will be reviewed. Let us begin with some distance metrics [9], [114], [38], [69], which were considered as the quality metrics at the early stage of quality assessment.

1. **Minkowsky Distance Metric** [9]

$$D = \left[\frac{1}{M \times N}\sum_{m=1,n=1}^{M,N}|X(m,n) - X_d(m,n)|^{\gamma}\right]^{\frac{1}{\gamma}} \tag{2.3}$$

where

$X(m,n)$	pixel value at position (m,n)
X,X_d	reference image and distorted image
M,N	image width and height
γ	distance order

The Minkowski distance [9] in Equation 2.3 is a metric on Euclidean space. With different distance order γ, different distance metrics can be obtained.
If $\gamma = 1$, mean absolute error is obtained.

$$D_1 = \frac{1}{M \times N}\sum_{m=1,n=1}^{M,N}|X(m,n) - X_d(m,n)| \tag{2.4}$$

If $\gamma = 2$, root mean square error is obtained.

$$D_2 = \left[\frac{1}{M \times N}\sum_{m=1,n=1}^{M,N}|X(m,n) - X_d(m,n)|^{2}\right]^{\frac{1}{2}} \tag{2.5}$$

If $\gamma = \infty$, maximum error is obtained.

$$D_3 = \arg\max_{m,n}|X(m,n) - X_d(m,n)| \tag{2.6}$$

In the literature, mean absolute error D_1 and maximum error D_3 are seldom used to predict image quality directly. Mean squared error $(D_2)^2$ is a popular metric which can

be found in most of the literature. Minkowski distance is also used in many modern quality assessment methods [17, 130] in which image dissimilarities are computed on different frequencies Minkowski distance is applied to pool these dissimilarities to obtain a visual quality score. Common values for γ in Minkowski distance are 3.5 [130] and 4 [17].

2. **Moments of the Angles** [114]

$$D_4 = \frac{2}{\pi} \sum_{m=1,n=1}^{M,N} A(m,n) \tag{2.7}$$

$$D_5 = \sum_{m=1,n=1}^{M,N} A(m,n) \times R(m,n) \tag{2.8}$$

$$A(m,n) = cos^{-1} \frac{< \underline{X}(m,n), \underline{X_d}(m,n) >}{||\underline{X}(m,n)|| \cdot ||\underline{X_d}(m,n)||} \tag{2.9}$$

$$R(m,n) = \frac{||\underline{X}(m,n) - \underline{X_d}(m,n)||}{\max\{||\underline{X}(m,n)||, ||\underline{X_d}(m,n)||\}} + 1 \tag{2.10}$$

where

$\underline{X}(m,n)$	the vector of RGB pixel value at position (m,n) in reference image
$\underline{X_d}(m,n)$	the vector of RGB pixel value at position (m,n) in distorted image
$<,>$	inner product of two vectors
$\|\| \cdot \|\|$	vector magnitude
$\max\{a,b\}$	larger value between a and b

Trahanias [114] defined visual quality as the moments of the angles. Trahanias stated that similar colors have almost parallel orientations in RGB color space. The angular distance, which quantifies the orientation difference between two color vectors $\underline{X}(m,n)$ and $\underline{X_d}(m,n)$, is a meaningful measure of the color similarity. Trahanias also proposed D_5 which considered the magnitude difference.

The above quality metrics evaluate image quality by computing the geometric distances or dissimilarities between the reference and the test (distorted) image. The advantage of these quality assessment methods is that the result is predictable and intuitive. Alternatively, the following fundamental metrics address quality assessment by utilizing image statistics.

3. **Normalized Cross-correlation** [38]

If an image is considered to be a random field X, each pixel value $X(m,n)$ is a sam-

ple generated by this random field. D_7 indicates the strength of a linear relationship between two images (random fields).

$$D_7 = \frac{\sum_{m=1,n=1}^{M,N} X(m,n)X_d(m,n)}{\sum_{m=1,n=1}^{M,N} X(m,n)^2} \tag{2.11}$$

4. **Mutual Information** [115]

$$D_8 = I(X;X_d) \tag{2.12}$$

$$I(X;X_d) = -\sum_{x\in\mathcal{X},x_d\in\mathcal{X}_d} p(x,x_d)\log\frac{p(x,x_d)}{p(x)p(x_d)} \tag{2.13}$$

where

$\mathcal{X}$, $\mathcal{X}_d$	the set of all possible values generated by random field X and X_d
$p(x,x_d)$	joint probability of pixel value pair (x,x_d)

The concept of information entropy describes how much randomness (uncertainty) there is in an image. For instance, if an image is plain white, from an information theory point of view, the randomness within the image is zero. If an image is distorted, its information (randomness) usually changes. For instance, if a plain white image is contaminated by Gaussian noise, which is often generated by amplifier devices, the information (randomness) of the distorted image is greater than the information of the reference image. Obviously the distorted image has degraded quality, which means increased information does not guarantee better visual quality. Mutual information is proposed as a reasonable measure for quality assessment, since it describes the common information between the reference image X and distorted image X_d. By definition [28], mutual information measures the remaining information after operations, such as compression and transmission which usually cause distortions. If an independent distortion (i.e. white noise) applied to the image X is so strong that the distorted image X_d is almost independent of X, then the joint probability $p(x,x_d)$ is close to $p(x)p(x_d)$ which makes the ratio inside the log operator close to 1 so that mutual information in Equation 2.12 is close to zero. If the distorted image X_d is very similar to X, mutual information is almost equal to the information in image X. From the above discussion, we notice that the mutual information could be a qualified metric for image quality assessment.

5. **Entropy correlation coefficient** [69]

$$D_9 = \frac{2I(X;X_d)}{H(X)+H(X_d)} \tag{2.14}$$

$$H(X) = \sum_{x\in\mathcal{X}} p(x)\log p(x) \tag{2.15}$$

$$H(X_d) = \sum_{x_d\in\mathcal{X}_d} p(x_d)\log p(x_d) \tag{2.16}$$

The problem of the mutual information metric D_8 is that it is not adaptive to images with varied contents. Consider two distorted images with the same perceptual quality. The mutual information of the image with simple content (plain image) might be much less than the mutual information of the image with complex content (natural scenes). Maes *et al.* [69] introduced a normalization operation, as in Equation 2.14, that eliminated the impact of varied image contents.

6. **Peak signal to noise ratio (PSNR)**

Peak signal to noise ratio (PSNR) maps the mean squared error into a logarithmic scale. PSNR does not depend on image intensity scaling. For different dynamic ranges, i.e. [0,1] or [0 255], PSNR gives a single score for image quality. Although PSNR is not originally intended to be designed as a method for quality assessment, it is one of the most popular quality metrics. The VEQG report [47] even concluded that none of the more sophisticated metrics show any great advantage over peak signal noise ratio.

$$D_{10} = 10\log_{10}\frac{MAX_X}{MSE} \tag{2.17}$$

$$MSE = \frac{1}{M\times N}\sum_{m=1,n=1}^{M,N}|X(m,n)-X_d(m,n)|^2 \tag{2.18}$$

In the following section, some modern quality assessment methods are reviewed in historical order. Most of these methods consider essential human visual behaviors in order to achieve consistent quality evaluation performances. We will also see that the fundamental quality metrics are widely used in the modern quality assessment methods.

7. **Visual Fidelity Criterion** (1974) [72]

Mannos and Sakrison [72] proposed a method on evaluating the impact of contrast sensitivity on image quality assessment. The contrast sensitivity response was modeled as a low pass filter $\boldsymbol{A}(f_u, f_v)$. Mannos and Sakrison assumed that the response of $\boldsymbol{A}(f_u, f_v)$ has the same magnitude in different directions. $\boldsymbol{A}(f_u, f_v)$ can be simplified to $\tilde{\boldsymbol{A}}(f_r)$, where $f_r = \sqrt{f_u^2 + f_v^2}$. The dissimilarity between the reference X and the distorted image X_d is defined as the squared error between Y and Y_d, where Y and Y_d are the corresponding images processed by the contrast sensitivity model $\tilde{\boldsymbol{A}}(f_r)$. This is the first quality assessment method which modeled contrast sensitivity behavior with Equation 2.2.

$$D_{11} = \sum_{m,n} [Y(m,n) - Y_d(m,n)]^2 \tag{2.19}$$

$$\tilde{\boldsymbol{A}}(f_r) = c + (f_r/f_0)^{k_1} e^{-(f_r/f_0)^{k_2}} \tag{2.20}$$

where

$Y(m,n), Y_d(m,n)$	the results of image X and X_d processed by the contrast sensitivity model $\tilde{\boldsymbol{A}}(f_r)$
f_0, c, k_1, k_2	contrast sensitivity function parameters

There is another method proposed by Mitsa and Varkur(1993) [76] that computes the weighted signal to noise ratio of the distorted image with respect to its reference and returns the image quality in dB. The basic idea of their method is to weight the mean squared errors in several frequencies according to contrast sensitivity. The quality prediction is similar to PSNR (D_{10}) but the mean squared error (MSE) is replaced by the weighted mean squared error.

8. **DCTune** (1993) [130]

Watson [130] approached quality assessment by considering luminance sensitivity and contrast masking. The luminance sensitivity t_{ijk} at frequency $(i\pi/8, j\pi/8)$ in the kth block (8×8 microblock) is defined according to the ratio between local luminance c_{00k} and global luminance $\bar{c}_{00}$.

$$t_{ijk} = t_{ij}[\frac{c_{00k}}{\bar{c}_{00}}]^{\alpha_t} \tag{2.21}$$

where

t_{ij}	the smallest coefficient change that yields a visual perception in spatial domain at each frequency $(i\pi/8, j\pi/8)$
c_{00k}	DC component at kth block
$\bar{c}_{00}$	the mean of DC component over all blocks
α_t	the nonlinearity parameter, which is recommended to be 0.649 by Ahumada and Peterson [4]

The definition of contrast masking threshold m_{ijk} considers luminance sensitivity and contrast masking, as shown in Equation 2.22. The contrast masking effect can be made more local by manipulating the parameter w_{ij}.

$$m_{ijk} = \max\{t_{ijk}, |c_{ijk}|^{w_{ij}} \times t_{ijk}^{1-w_{ij}}\} \tag{2.22}$$

where

c_{ijk}	DCT coefficient of kth block at frequency $(i\pi/8, j\pi/8)$
w_{ij}	between 0 and 1.

Watson described the difference e_{ijk} between images in threshold units as shown in Equation 2.23. Finally, the Minkowski metric is applied to pool differences in different blocks (Equation 2.24) and frequencies (Equation 2.25) as the visual quality (Equation 2.25) for an image. Bradly chose $\beta = \beta_f = 4$.

$$d_{ijk} = \frac{e_{ijk}}{m_{ijk}} \tag{2.23}$$

$$p_{ij} = \left[\sum_k |d_{ijk}|^{\beta}\right]^{\frac{1}{\beta}} \tag{2.24}$$

$$D_{12} = \left[\sum_{ij} p_{i,j}^{\beta_f}\right]^{\frac{1}{\beta_f}} \tag{2.25}$$

9. **Picture Quality Scale** (1996) [77]

Miyahara *et al.* [77] addressed image quality assessment in a signal processing manner. The preliminary step is luminance correction.

$$X(m,n) = I(m,n)^{1/2.2} \tag{2.26}$$

where $I(m,n)$ is pixel value of the image

Miyahara *et al.* modeled human visual contrast sensitivity as an anisotropic function with two steps. $\boldsymbol{S}(\omega)$ in Equation 2.28 depicts such a sensitivity function that has the maximum magnitude at frequency of $\frac{\log(16/9)}{\sigma}$ and then monotonically decreases to zero. $\boldsymbol{O}(\omega)$ in Equation 2.29 controls the contrast sensitivity varying with the angle's change. $\boldsymbol{O}(\omega, \theta)$ has the weakest response when the angle $\theta = \tan^{-1}(v/u)$ is a multiple of $\pi/4$. This anisotropic effect performs strongest at ω_o.

$$\boldsymbol{S}_a(u,v) = \boldsymbol{S}(w)\boldsymbol{O}(w,\theta) \tag{2.27}$$

$$\boldsymbol{S}(\omega) = 1.5e^{-\sigma^2\omega^2/2} - e^{-\sigma^2\omega^2} \tag{2.28}$$

$$\boldsymbol{O}(\omega,\theta) = \frac{1 + e^{\beta(\omega-\omega_o)}\cos^4 2\theta}{1 + e^{\beta(\omega-\omega_o)}} \tag{2.29}$$

where $\sigma = 2$ $\omega = \dfrac{2\pi\sqrt{u^2+v^2}}{60}$ $\theta = \tan^{-1}(v/u)$ $\beta = 8$ $\omega_o = 1.165$

Miyahara *et al.* introduced five strategies to compute error maps. The first error factor F_1 was computed by CSF weighted squared error as shown Equation 2.30. F_1 was originally proposed by International Telecommunication Union (1982) as the quality prediction for noise degradation.

$$f_1(m,n) = [E_I(m,n) * W(m,n)]^2 \tag{2.30}$$

$$F_1 = \frac{\sum_{m,n} f_1(m,n)}{\sum_{m,n} I^2(m,n)} \tag{2.31}$$

where

$FT\{W(m,n)\} = \boldsymbol{W}(u,v) = \dfrac{1}{1 + (u^2+v^2)/f_c^2}$

$f_c = 5.56$ cycles/degree

$E_I(m,n) = I(m,n) - I_d(m,n)$

The distortion factor F_2 adopts anisotropic contrast sensitivity and contrast masking. $S_a(m,n)$ accounts for contrast sensitivity. In contrast to other methods [130], Miyahara et al. implemented the masking effect with a simple indicator function $I_T(m,n)$ with a perceptual threshold $T = 1$.

$$f_2(m,n) = I_T(m,n)\,[E_X(m,n) * S_a(m,n)]^2 \tag{2.32}$$

$$F_2 = \frac{\sum_{m,n} f_2(m,n)}{\sum_{m,n} i^2(m,n)} \tag{2.33}$$

where $FT\{S_a(m,n)\} = \boldsymbol{S}_a(u,v)$ which is defined in Equation 2.27
$E_X(m,g) = X(m,n) - X_d(m,n)$

Block disturbances factor F_3 is specifically designed for block based distortion, such as JPEG compression. The basic idea is to accumulate the discontinuities along the artificial block edges. This idea is also widely shared in many quality measurements for JPEG pictures [68, 89, 111, 119, 124]. The error map in horizontal direction $f_{3h}(m,n)$ is computed as:

$$f_{3h}(m,n) = I_h(m,n)\triangle_h^2(m,n) \tag{2.34}$$

where $\triangle_h(m,n) = E_X(m,n) - E_X(m,n+1)$
$I_h(m,n)$ is an indicator function where selects only those differences on the block boundaries by default [68, 89], since images of natural scenes rarely have periodic boundaries.

The distortion factor F_3 is defined as the combined block effect in horizontal and vertical direction.

$$F_3 = \left[\sum_{m,n} f_{3h}(m,n) + \sum_{m,n} f_{3v}(m,n)\right]^{1/2} \tag{2.35}$$

The fourth distortion factor F_4 considers the distortion which disrupts strong local spatial correlation. The distortion map $f_4(m,n)$ is computed as the average correlations of the block centering at (m,n) with its neighbors.

$$F_4 = \sum_{m,n} f_4(m,n) \tag{2.36}$$

Miyahara et al. named the fifth distortion factor F_5 as errors in the vicinity of high contrast image transitions. The basic idea of F_5 is still based on the statement that the visibility of disturbance is reduced in active areas (texture-rich area). s_h and s_v are defined such that the pixel value changes within very small (2 pixels) distance exponentially affect the perception of differences.

$$f_5(m,n) = |e_w(m,n)|\,[s_h(m,n) + s_v(m,n)] \tag{2.37}$$

$$F_5 = \sum_{m,n} f_5(m,n) \tag{2.38}$$

where $s_h(m,n) = e^{-0.02|i(m,n-1)-i(m,n+1)|}$
$s_v(m,n) = e^{-0.02|i(m-1,n)-i(m+1,n)|}$

Picture Quality Scale (PQS) is quite different from other methods. Those methods usually follow the track of visual signal in human vision system, and simulate the subsequent response. Instead, PQS computes several distortion factors in a parallel way. Some distortion factors (F_2 and F_5) even correspond to the same vision feature. To pool all these distortion factors, Miyahara et al. used principle component analysis. The key idea is to divide the whole image into several slices (for instance N_0 slices), and compute the distortion vector $[F_1, ..., F_5]^T$ for each slice. Then principle component analysis is carried out to fuse the distortion factors by taking the first several eigenvalues of the correlation matrix.

10. **Visible Difference Predictor** (1999) [17]

Bradley [17] proposed a wavelet visible difference predictor in 1999. He might be the first that adopted multi-resolution analysis in quality assessment. His method has four stages. First, an image is decomposed into wavelet domain (Daubechies Biorthogonal 9/7 wavelet). The contrast sensitivity in each subband (specific orientation θ and frequency f) is calculated as the minimum noticeable threshold $y(\theta, f)$ in the spatial domain [131].

$$\log y(\theta, f) = \log 0.495 + 0.466[\log f - \log(0.401 g_\theta)]^2 \tag{2.39}$$

where

- θ LL, LH, HL or HH subband
- g_θ is 1.501,1 and 0.534 for the LL, LH/HL and HH subbands

Bradley used a "worst case" formula that converts contrast sensitivity to the wavelet domain.

$$n_w(\theta, f) = \frac{y(\theta, f)}{i_\theta \times p_l^{2(l-1)}} \tag{2.40}$$

where

- l wavelet decomposition level
- i_θ 0.6217, 0.6723 or 0.7271 for θ = LL, LH/HL or HH

The contrast masking, modeling signal content sensitivity, is the third stage. The threshold for masking at each specific position is given in Equation 2.41. Similar to the DCTune method [77], the masking effect can be made more local by manipulating parameter $b(f)$. Bradley chose $b(f)$ to be 1.

$$T_e(\theta, f, i, j) = \max\{n_w(\theta, f), b(f) \times |w(i, j)|\} \tag{2.41}$$

where $w(i,j)$ is wavelet coefficient.

Bradley extended contrast masking further to mutual masking. The basic idea of mutual masking is to consider the flat area due to over compression especially for highly textured area in the original image. Bradley suggested $T_{em}(\theta,f,i,j) = \min\{T_e(\theta,f,i,j), \dot{T}_e(\theta,f,i,j)\}$ as the masking threshold, where $\dot{T}_e(\theta,f,i,j)$ is the threshold for distorted (compressed) image.

Instead of pooling differences by Minkowski metric [17,130], Bradley applied a detection probability for the difference at position (i,j), first proposed by Daly [30].

$$P_b(\theta,f,i,j) = 1 - \exp(-|\frac{\triangle w(\theta,f,i,j)}{T_{em}(\theta,f,i,j) \times \alpha}|^{\beta}) \tag{2.42}$$

where

$\triangle w(\theta,f,i,j)$	wavelet coefficient difference between the reference image and distorted image
α, β	decision threshold, $\alpha = 4.0, \beta = 2.0$ in [30]

Detection probability is accumulated at position (i,j) across different orientation $\theta =$ HL/LH and HH. The impairment score D_{14} is the Minkowski sum of the detection probabilities $P_d(f,i,j)$.

$$P_d(f,i,j) = 1 - \prod_{\theta} [1 - P_b(\theta,f,i,j)] \tag{2.43}$$

$$D_{14} = \left[\sum_{i,j} |P_d(i,j)|^{\beta}\right]^{1/\beta} \tag{2.44}$$

Later, Lai *et al.* (2000) [62] proposed a perceptual difference map based on the Haar wavelet. Their method implements a similar contrast sensitivity function as perceptual image fidelity [110]. Furthermore, Lai et al. also considered the interactions among channels. Lai stated that the visibility of a stimulus at certain frequency could be affected by the stimulus in adjacent frequency channels. One simple example is the blocking artifact propagating across frequencies. Hence Lai et al. modeled contrast sensitivity as a function of several frequencies. The errors between the reference and distorted image are computed point-to-point in the wavelet domain and pooled by the Minkowski metric.

11. **Universal Image Quality Index** (2002) [121] and **Structural Similarity** (2002) [122]

Instead of simulating human visual reaction to distortions, Wang [121,122,126] proposed a QA metric by modeling any image distortion as a combination of three components: *loss of correlation*, *contrast degradation*, and *luminance distortion*. Universal objective image quality is a simple mathematically defined quality index without any human vision system model explicitly employed. An image is treated as a random field, and each pixel value is a sample generated by this field. Fundamental statistics properties, such as mean and standard deviation, are defined as background luminance or texture activities.

$$\text{loss of correlation:} \qquad s(x, x_d) = \frac{\sigma_{xx_d}}{\sigma_x \sigma_{x_d}} \tag{2.45}$$

$$\text{luminance distortion:} \qquad l(x, x_d) = \frac{2\mu_x \mu_{x_d}}{\mu_x^2 + \mu_{x_d}^2} \tag{2.46}$$

$$\text{contrast distortion:} \qquad c(x, x_d) = \frac{2\sigma_x \sigma_{x_d}}{\sigma_x^2 + \sigma_{x_d}^2} \tag{2.47}$$

where

μ_x, μ_{x_d}	mean luminance of the reference X and distorted image x_d
σ_x, σ_{x_d}	standard deviation of pixel value
σ_{xx_d}	luminance correlation between the reference x and distorted image X_d

Wang defined the image visual quality as the combination of $s(x, x_d)$, $l(x, x_d)$ and $c(x, x_d)$ in Equation 2.48

$$D_{15} \quad = \quad s(x, x_d) l(x, x_d) c(x, x_d) \tag{2.48}$$

Later, Wang proposed a improved quality metric, structural similarity [122] as shown in Equation 2.52. The improved metric adds the constants (c_1, c_2 and c_3) to avoid instability. Wang also introduced the parameters (α, β and γ) to adjust the relative importance of the three components: contrast comparison, structure comparison and luminance change.

$$\text{structure comparison:} \qquad s(x, x_d) = \frac{\sigma_{xx_d} + c_1}{\sigma_x \sigma_{x_d} + c_1} \tag{2.49}$$

$$\text{luminance change:} \qquad l(x, x_d) = \frac{2\mu_x \mu_{x_d} + c_2}{\mu_x^2 + \mu_{x_d}^2 + c_2} \tag{2.50}$$

$$\text{contrast comparison:} \qquad c(x, x_d) = \frac{2\sigma_x \sigma_{x_d} + c_3}{\sigma_x^2 + \sigma_{x_d}^2 + c_3} \tag{2.51}$$

$$D_{16} \quad = \quad [s(x, x_d)]^\alpha [l(x, x_d)]^\beta [c(x, x_d)]^\gamma \tag{2.52}$$

The structure similarity is defined on the entire image. In practice, an image is divided into several local windows, and the overall structure similarity (MSSIM) [129] is accumulated over local windows.

12. **Visual Information Fidelity** (2006) [97]

Visual information fidelity is a quality assessment approach based on information theory. Sheikh et al. [97] defined the human visual system as a typical additive noise channel. An image X is treated as a random signal and sent in at one end. The other end, the brain, receives the visual information Y, which is defined in Equation 2.53.

$$Y = X + V \tag{2.53}$$

where V is vision noise and obeys normal distribution with zero mean and σ_V^2

The mutual information between X and its perceptual image Y can be computed as:

$$I(X; Y) = log(1 + \frac{\sigma_X^2}{\sigma_V^2}) \tag{2.54}$$

In contrast to the existing methods, Sheikh et al. built a relationship with the reference x and distorted image x_d with a distortion model [58].

$$X_d = kX + Z \tag{2.55}$$

where k is scalar and Z is Gaussian noise with zero mean and σ_Z^2

Based on the distortion model, the common information between X and the perceptual image Y_d of the distorted image X_d can be computed as:

$$I(X; Y_d) = log(1 + \frac{k^2 \sigma_X^2}{\sigma_Z^2 + \sigma_V^2}) \tag{2.56}$$

From Equation 2.56, it can be seen that as the scalar k decreases (blur effect) or noise Z increases (noise, compression effect or quantization noise), $I(X;Y_d)$ is going to decrease. The visual information fidelity in each frequency band is defined as the ratio between the two mutual informations $I(X;Y)$ and $I(X;Y_d)$. The overall visual information fidelity can be computed as the ratio between two mutual information in all channels.

In the following section, several blind quality assessment methods are reviewed. All of them are specially designed for image with JPEG or JPEG2000 format.

13. **No-reference JPEG image quality index** (2004) [124]

The traditional no-reference quality assessment for JPEG compression is based on block effect detection. Since JPEG standard uses 8x8 microblock, the discontinuity across the block boundary results in a grid structure. The grid structure (similar to 2-D impulse train) has a significant pattern in the frequency domain. The block effect can be evaluated by this pattern [16, 68]. Differently, Wang [124] proposed a quality prediction for JPEG compression by considering three different activities. First, a neighborhood difference image $d_h(m,n)$ is computed in horizontal direction. The average difference at each block border (for JPEG block is 8×8) is computed as the discontinuity B_h at block border in horizontal direction.

$$d_h(m,n) = X_d(m,n+1) - X_d(m,n) \tag{2.57}$$

$$B_h = \frac{1}{M\lfloor N/8 \rfloor} \sum_{i=1}^{\lfloor N/8 \rfloor - 1} |d_h(i,8j)| \tag{2.58}$$

The pixel activity is defined as:

$$A_h = \frac{1}{7}[\frac{8}{M(N-1)} \sum_{m}^{M} \sum_{n}^{N-1} |d_h(m,n)| - B_h] \tag{2.59}$$

The binary zero-cross rate at (m,n) is defined as:

$$z_h(m,n) = \begin{cases} 1, & \text{for } d_h(m,n-1) \times d_h(m,n+1) < 0 \quad (2.60) \\ 0, & \text{else} \quad (2.61) \end{cases}$$

The overall zero-crossing rate is computed as:

$$Z_h = \frac{1}{M(N-2)} \sum_{m}^{M} \sum_{n}^{N-2} z_h(m,n) \tag{2.62}$$

B_v, A_v and Z_v in the vertical direction can be computed similarly. The quality of an JPEG image can be predicted as a nonlinear combination of three activities in Equation 2.63. α, β and γ are model parameters that are estimated with subjective data.

$$D_{18} = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \tag{2.63}$$

$$\text{where} \qquad B = (B_h + B_v)/2 \tag{2.64}$$

$$A = (A_h + A_v)/2 \tag{2.65}$$

$$Z = (Z_h + Z_v)/2 \tag{2.66}$$

Several no-reference quality assessment methods [67,68,138] have been proposed for JPEG compression. Wu and Yuen (1997) [138] computed the difference along the block edges, and weighted the difference by local variance. The basic idea of Xin (2002) [67] is to evaluate edge sharpness. While Liu and Bovick (2002) [16,68] chose to count on the strong frequency pattern in frequency domain. Sazzad *et al.* (2008) [53,92] were trying to extend Wang's method to JPEG2000 compression.

14. **No-reference quality assessment for JPEG2000 using natural image statistics** (2005) [96]

Sheikh *et al.* [96] approached blind quality assessment for JPEG2000 by using the consistency of wavelet coefficients across different resolutions. Consider the hidden Markov tree model (HMT) proposed by Crouse *et al.* [29]. The HMT model assigns each wavelet coefficient with two hidden state (large and small). For a child coefficient and its parent, there is a state pair assigned, which has four possible combinations (*ss*,*ll*,*ls*,*sl*). Sheikh et al. stated that the probability of *ss* ($p(ss)$) gives the best indication of the quality loss. $p(ss)$ can be estimated by using the wavelets secondary property [18]. But $p(ss)$ can not be directly used as the objective score. A quality feature in ith wavelet scale is defined as:

$$q_i = K_i \left[1 - \exp\left(-\frac{p_i(ss) - u_i}{T_i} \right) \right] \tag{2.67}$$

where

q_i	the transformed feature (predicted image quality) for the ith scale
K_i,T_i, u_i	curve fitting parameters that are learned from the training data [96]

An image might represent different statistics on different scales, hence a weighted average of the transformed feature is used for quality prediction as shown below.

$$D_{19} = \sum_i w_i \times q_i \tag{2.68}$$

Sheikh *et al.* are the first to approach the blind quality assessment by using image statistics and achieve fairly consistent results with human judgements.

15. **No-reference JPEG2000 image quality index** (2008) [90,91]

Sazzad *et al.* [90,91] proposed a no-reference JPEG2000 quality assessment. In order to achieve good consistency with human judgments, the method extracts seven spatial features. The first feature S is the complexity of the texture. Instead of using the local standard deviation s of luminance, Sazzad et al. computed the local standard deviation S of s. The second feature is the absolute difference measure A of a central pixel from the second closet neighborhood pixels. The edge information measure Z is defined in the above no-reference JPGE quality method (Equation 2.66). The next two features are defined as histogram measure in the horizontal direction. In order to evaluate the possible edge loss due to compression, Sazzad et al. used a Sobel filter to emphasize the edges. Let h_{f_0},h_{f_1} , h_{f_2} and h_0, h_1, h_2 be the value of histogram at $0, 1, 2$ for the distorted (compressed) image and distorted image after Sobel filter. Sazzad et al. stated that the compression differs h_{f_0} ,h_{f_1},h_{f_2} from h_0,h_1,h_2. Then two histogram features H_f and H are defined:

$$H_f = \frac{h_{f_0} + h_{f_1} + h_{f_2}}{3} \tag{2.69}$$

$$H = \frac{h_0 + h_1 + h_2}{3} \tag{2.70}$$

Similarly, Sazzad et al. defined the other two histogram features V_f and V in the vertical direction. Finally, a highly nonlinear combination method is proposed as shown in Equation 2.71.

$$\begin{aligned} C = & [\gamma_1 log(S+1) + \gamma_2 log(A+1) + \gamma_3 log(Z+\gamma_4)] \\ & \times [\gamma_5 log(H_f+1) + \gamma_6 log(V_f+1) \\ & + \gamma_7 log(H+1) + \gamma_8 log(V+1) + \gamma_9] \end{aligned} \tag{2.71}$$

$$D = \frac{4}{1 + e^{-1.0217(C-3)}} + 1 \tag{2.72}$$

There are also few other blind quality assessment methods proposed [21,22,52,111,119]. Tong *et al.* [111] identified the edge distortion using principal component analysis. Holliman *et al.*(2002) [52] and Campisi *et al.*(2003) [21,22] evaluated image quality by estimating the damage to the embedded watermarks. Venkatesh *et al.* (2007) [119] addressed blind quality assessment problem using a GAP-RBF neural network. Up to now, very little literature on JPEG2000 compression has been successful. Within the few existing no-reference quality assessment methods, Sazzad's method has the best consistency with the current subjective databases.

2.2.2 Quality Method Evaluation

Before we start to test the performance of the existing quality metrics, it is necessary to spend some time on the evaluation standards. For a database, there are usually several hundred test samples. To evaluate a quality metric, it is important to define the "similarity" between two sets of quality measures: the subjective quality scores and the objective(evaluated) quality scores. In this thesis, we use two metrics to measure the "similarity": Spearman correlation.

The Spearman correlation is defined between two sets $\{x_i\}$ and $\{y_i\}$. r_{x_i} and r_{y_i} are rankings for each x_i and y_i in their own sets. Then Spearman correlation is defined in Equation 2.73, where $d_i = r_{x_i} - r_{y_i}$ and n is the number of values in each data set. Spearman correlation varies from -1 to 1. As for quality assessment, the larger absolute Spearman correlation value between the subjective quality and the objective quality is preferable.

$$\rho = \frac{1 - 6\sum d_i^2}{n(n^2 - 1)} \tag{2.73}$$

Furthermore Tourancheau [113] suggested that it is important to test the quality metric with different subjective databases in case the quality metric is over trained for certain database. Hence in this thesis, five different subjective databases are used during the validation of the quality metrics.

2.2.3 Summary of Existing Methods

From the review section, we notice that image quality assessment until now has three stages. At the beginning, people simply used the error/distance/dissimilarity metric for quality evaluation. These methods have been proved to be inconsistent with human judgments, even though some famous error metrics (MSE, PSNR) are still widely applied in image restoration and compression areas. Later, psychophysical evidence revealed some fundamental visual behaviors. People started to model human visual reactions and tried to build the connections between these reactions and image quality assessment. One representative method is the visual fidelity criterion by Mannos and Sakrison [72]. Mannos and Sakrison attempted to weight different aspects of the error signal according to their visibility, as determined by the contrast sensitivity. This approach was extended by many other researchers [17, 77, 130] by considering more human visual features. Recently, people no longer insisted on simulating visual behaviors. One reason might be the difficulties in modeling highly nonlinear and adaptive vision functions. Some simple linear/nonlinear models were proposed based on the observations of some common vision behaviors. For instance, the contrast masking effect on quality evaluation was modeled as a function of local intensity variance [122]. People also tended to use some classic signal analysis methods, such as principle component analysis [77, 79, 111].

In the first section of this chapter, we reviewed four human visual properties, which have great popularity among quality assessment methods. Almost all literature on quality assessment adopts one or more of them. Table 2.1 lists the popularity of these four features. The table also lists how these vision features are modeled. Function $f(x)$ computes the human visual response if the stimulus is x. x can be pixel luminance value, transform coefficient or operation result. As shown in the table, most literature models the human visual features as nonlinear functions. i.e. the exponential, logarithmic or power function. For full-reference quality assessment, contrast sensitivity is adopted by over 60% of the quality metrics. This vision feature is incorporated in our method in Chapter 5.

In this section, five subjective databases are used to test the existing quality assessment methods. They are the LIVE [99], IVC [64], TOYAMA [55], Cornell A57 [25] and TID2008 [81] databases. Each database emphasizes different distortions. For instance, LIVE database contains five distortions: JPEG, JPEG2000, white noise, Gaussian blur and communication error. IVC, TOYAMA and Cornel A57 databases focus on compression distortion: JPEG and JPEG2000. TIC2008 database contains several noise distortions, such as additive noise in color components [81], impulse noise [81] and

Vision Feature	Quality Assessment Method	Model
Luminance Sensitivity	Method 8 [130]	$f(x) \sim x \times r^{\alpha}$
	Method 9 [77]	$f(x) \sim x^{\alpha}$
	Method 11 [121, 122]	$f(x) \sim \frac{x}{t}$
Contrast Sensitivity	Method 7 [72]	$f(x) \sim log(x)$
	Method 8 [130]	$f(x) \sim \max\{a, b\}$
	Method 9 [77]	$f(x) \sim e^{-x}$
		$f(x) \sim log(x)$
	Method 10 [17]	$f(x) \sim e^{-x}$
		$f(x) \sim \max\{a, b\}$
	Method 11 [121, 122]	$f(x) \sim \frac{1}{\sigma_x^2}$
	Method 12 [97]	$f(x) \sim log(x)$
Vision Channel Decomposition	Method 7 [72]	Fourier Domain
	Method 10 [17]	Fourier Domain
	Method 12 [97]	Wavelet Domain
Masking	Method 8 [130]	$f(x) \sim \frac{x}{t}$
	Method 9 [77]	$f(x) \sim \frac{x}{t}$
	Method 10 [17]	$f(x) \sim \exp(-x^{\alpha})$
	Method 11 [121, 122]	$f(x) \sim \frac{x}{t}$

Table 2.1: Human visual properties and modern quality assessment methods. Function $f(x)$ computes the human visual response if the stimulus is x. As shown in the table, most literature models the human vision features as nonlinear functions. i.e. the exponential, logarithmic or power functions.

JPEG	448	JPEG2000	442	Gaussian blur	300
White noise	274	Fastfading	174	Additive noise in color components	174
Spatially correlated noise	100	Masked noise	100	High frequency noise	100
Impulse noise	100	Quantization noise	100	Image denoising	100
JPEG transmission errors	100	JPEG2000 transmission errors	100	Non eccentricity pattern noise	100
Local block-wise distortion	100	Mean shift (intensity shift)	100	Contrast change	100

Table 2.2: The details of the test samples

high frequency noise [81]. These five databases contain over three thousand test images, which cover eighteen types of distortions. Table 2.2 displays the number of test samples with different distortions. Around 1/3 test samples are JPEG and JPEG2000 pictures, which reflects the urge for quality assessment on compressed images.

The subjective databases were evaluated by different quality assessment methodologies. The LIVE [99], IVC [64] and TOYAMA [55] databases adopt DMOS method defined in ITU-R BT.500-10 [2], in which the quality assessment is achieved by pairwise comparison between a test image and its reference. The Cornell A57 [25] database uses a customized method: viewers were instructed to position the distorted images in such an order that the distance between each distorted image and the reference was linearly proportional to their subjective assessment of distortion. Then the subjective quality judgments are computed as a function of the distance. More details about assessment procedure of Cornell A57 database can be found in [25]. The TID2008 [81] database uses another subjective quality assessment method proposed by Ponomarenko *et al.* [81]. The assessment procedure has two stages. First, the pairwise comparisons between test images (not a test image and its reference) are conducted so that the test samples are ordered according to their visual quality. Second, the quantitative evaluation of image quality is determined according to the order obtained in the first stage. More details can be found in [81].

To test the existing quality assessment metrics, Spearman correlations between the subjective judgments and the objective quality scores by these methods are computed and listed in Table 2.2.3. From the test results, we notice that the modern

	JPEG2000		JPEG	
	Sheikh's method [96]	Sazzad's method [90]	Wang's method [124]	Sazzad's method [90]
LIVE [99]	0.8621	0.8980	0.9010	0.8962
Toyama [55]	0.7799	0.8169	0.8942	0.7648
IVC [64]	0.8567	0.7922	0.9333	0.9472
TID2008 [81]	0.4933	0.8750	0.9159	0.5318

Table 2.3: The performance of the blind quality assessment metrics. The table lists the absolute value of Spearman correlation between the subjective judgments and objective scores by each method. If the value is close to 1, the method has consistent performance with human judgments.

quality metrics have more stable performance on five subjective databases. Among all the quality metrics, the visible difference predictor [17], structural similarity [122], multi-scale structural similarity [128] and visual information fidelity [97] have the best performances. We also notice that even the best methods in Table 2.2.3 are not compatible to all five databases. For instance, structural similarity performs badly on the Cornell A57 database [25] and visual information fidelity does not properly handle the Cornell A57 [25] and TID2008 [81] databases. We can summarize the problems with current quality assessment as follows.

- Some distortions, such as impulse noise in the Cornell A57 database [25], are just recently taken into consideration. The "old" methods are not able to evaluate these new distortions.
- Different human visual features were modeled to evaluate image visual quality by many methods, as listed in Table 2.1. Some methods [17, 97, 122] adopt more than one human visual feature. But which human visual feature dominates the assessment performance on certain distortion (i.e. JPEG or JPEG2000 distortion) is unknown. In the existing literature, people [80, 102] are more interesting in which distortion can be well evaluated.

The existing blind quality methods are also tested in this section. Table 2.3 listed the performance of blind quality metrics over four subjective databases: the LIVE [99], TOYAMA [55], IVC [64] and TID2008 [81] databases. The problems with the existing blind quality assessment can be summarized as follows.

- For JPEG2000 distortion, all tested methods do not have stable performance on four subjective databases. Sheikh's method [98] has acceptable performance

on the LIVE [99] database but cannot deal with the TID2008 [81] database. Sazzad's method [54] performs relatively better than Sheikh's method, but does not properly evaluate the TOYAMA [55] and IVC [64] databases. Hence there is no method that can deal with the JPEG2000 distortion and maintain a stable performance. For the JPEG distortion, Sazzad's method has good performance on the LIVE [99] , TOYAMA [55] and IVC [64] databases , but is not compatible with the TID2008 [81] database. Wang's method [124] has the excellent performance on the four databases, but there is still some room for the improvement.

- The strengthes on full reference quality assessment were not utilized in the development of blind quality assessment metrics. Full reference quality assessment methods [17,77,97,122,130] emphasize modeling the human visual reactions to image differences. But blind quality assessment metrics [90,96,124] usually choose a highly adhoc strategy to interpolate the artificial effects caused by distortions.

- A blind quality metric usually extracts several features and combines the features by using complicated nonlinear formulas. The proper explanation of mapping is missed. For instance, Sazzad [90] defined the visual quality score D (Equation. 2.75) as a exponential function of C, where C is computed highly nonlinearly as a combination of seven image features S, S, A, Z, H_f, V_f, H and V. But Sazzad did not explain how Equation 2.74 was derived. Also a blind quality metric usually adopts several parameters (i.e. Sazzad method has over 10 parameters). In order to tune these parameters to achieve the best performance, a training procedure is involved, introducing the risk of over training to a specific group of images. The parameters are very dependent on the subjective database.

$$\begin{aligned} C &= [\gamma_1 log(S+1) + \gamma_2 log(A+1) + \gamma_3 log(Z+\gamma_4)] \\ &\quad \times [\gamma_5 log(H_f+1) + \gamma_6 log(V_f+1) \\ &\quad + \gamma_7 log(H+1) + \gamma_8 log(V+1) + \gamma_9] \end{aligned} \tag{2.74}$$

$$D = \frac{4}{1+e^{-1.0217(C-3)}} + 1 \tag{2.75}$$

- All blind quality assessment metrics are specially designed for images with JPEG format or JPEG2000 format. There is no general solution to evaluate the visual quality of an image with any type of distortion.

In this thesis, we propose an information concept which describes the amount of image information that can be perceived by a human observer. This concept yields an efficient solution to quality assessment. Our solution exhibits the following characteristics.

- Full reference quality assessment (Chapter 5)
  - Our solution has better compatibility to five subjective databases compared with the other existing methods.
  - The key points for the success of our solution are investigated.
  - Our solution provides a foundation to develop the blind quality assessment metrics.
- Blind quality assessment (Chapter 6)
  - Based on the success of our full reference quality method, several blind quality metrics are proposed for five common distortions, which are the JPEG, JPEG2000, white noise, Gaussian blur and fastfading distortions (communication error). The expression of blind quality metrics are directly developed from our full reference quality metric.
  - Our blind quality metrics only contain 2 or 3 parameters. These parameters are constant for different subjective databases. No training procedure is required.
  - Our blind quality metrics have excellent compatibility to different subjective databases and outperform the existing methods.

2.3 Summary

This chapter starts with a review of some aspects of human vision: luminance sensitivity, contrast sensitivity, vision channel decomposition and masking. In the second part of the chapter, 15 quality methods are reviewed in detail. We specify the human visual features applied in each method and summarize the quality methods in a family tree. According to the literature survey, contrast sensitivity is the most popular human visual feature and is usually modeled as exponential/logarithmic function. We summarize the main problems with the current quality assessment and point out the research direction for the remainder of the thesis.

	LIVE	TOYAMA	IVC	Cornell A57	TID2008
Absolute Error	0.8862	0.6751	0.6040	0.3824	0.2640
Mean Squared Error	0.8999	0.6825	0.6386	0.4844	0.5245
Maximum Error	0.9149	0.5654	0.6821	0.3824	0.3843
Moments of the Angles (D_4)	0.7518	0.2937	0.5357	-*	0.0056
Moments of the Angles (D_5)	0.8172	0.3832	0.5519	-*	0.0526
Normalized Cross-correlation	0.8935	0.6128	0.5944	0.4835	0.4787
Mutual Information	0.8644	0.5122	0.5818	0.2434	0.4462
Entropy Correlation Coefficient	0.8867	0.5672	0.5780	0.1473	0.4059
Peak Signal Noise Ratio	0.8999	0.6825	0.6386	0.3113	0.5245
Visual Fidelity Criterion	0.9106	0.5938	0.5972	0.4888	0.5402
DCTune	0.8584	0.6117	0.8119	0.2616	0.4374
Picture Quality Scale	0.9338	0.6650	0.8552	0.0671	0.4252
Visible Difference Predictor	0.9178	0.6916	0.6801	0.2110	0.6784
Visual Signal Noise Ratio	0.6568	0.7635	0.8418	0.9339	0.7118
Structural Similarity	0.9440	0.7885	0.9162	0.8077	0.7642
Multi-scale structural Similarity	0.9444	0.7705	0.9043	0.6500	0.8527
Visual Information Fidelity	0.9695	0.7924	0.9274	0.6224	0.7490

Table 2.4: The performance of existing QA methods. The table lists the absolute value of the Spearman correlation between the subjective judgments and the objective evaluations by the listed methods. The method with the larger Spearman correlation is preferable. The modern quality metrics are listed in the shaded area. *Test samples in Cornell A57 database are all gray level images. Moments of the angles metrics (D_4, D_5) can only work on colorful images.

Chapter 3

Image Statistics

This chapter begins with a brief review on the wavelet transform and image statistical modeling in the wavelet domain. The primary purpose of this chapter is to introduce the notation and to provide a suitable background for the establishment of our information theoretic framework for image quality assessment.

3.1 Wavelet Transforms

Wavelet analysis was proposed based on the work of Joseph Fourier in the nineteenth century. Fourier presented the foundations of the theory of frequency analysis, in which signals are represented in the space constructed by sinusoidal bases. Fourier analysis has been proven to be tremendously powerful [107]. However, researchers became more and more interested in the representation of signals in a new domain. This new domain is constructed by non-sinusoidal bases which may have compact support. The first non-sinusoidal base with compact support is the Haar wavelet base proposed by Alfred Haar [50] in 1909. In the late nineteen-eighties, Daubechies [32], Mallat [70] and Meyer [75] further explored and applied the ideas of wavelet transforms to address signal analysis. There are two types of the wavelet transforms popular in signal analysis. One is the orthogonal wavelet [70]. The bases of such wavelets are orthogonal. But orthogonal wavelets are not popular in image precessing, because most orthogonal wavelet bases are not symmetrical except for the Haar wavelet [32]. It is well known in the subband filtering community that symmetry and exact reconstruction are incompatible (except for the Haar wavelet) if the same FIR filters are used for reconstruction and decomposition. The other is the biorthogonal wavelet [70], which uses two different wavelet bases for analysis and synthesis. One important difference be-

Orthogonal wavelet	Biorthogonal wavelet
db1	bior1.1
db2	bior2.4
db3	bior3.7

Table 3.1: The list of the tested wavelets. 38 wavelets are applied to our solution in Chapter 5 to determine the impact of different wavelets on quality assessment performance. The table lists some examples.

tween the orthogonal and biorthogonal wavelet is that the biorthogonal wavelet bases can be designed to be symmetrical. However real cartesian-separable wavelets might be not good for image analysis since they have poor orientation selectivity (vertical, horizontal and diagonal directions) [101]. Simoncelli and Freeman proposed the steerable pyramid [104] under the wavelet framework in 1990 which has better orientation selectivity (6 orientations).

In Chapter 5, 38 wavelets are applied to our solution in order to determine the impact of various wavelet bases on the quality assessment performance. The tested wavelets include orthogonal, biorthogonal wavelets and the steerable pyramid. Table 3.1 lists a few of them. In the first column of the table, three Daubechies compactly supported wavelets [70] (db1, db2 and db3) are listed in the form of "db$N$". The number N (1, 2 or 3) after "db" indicates the vanishing moments [70] of the wavelet. Theoretically, for a wavelet with large vanishing moments, the wavelet coefficients will tend to be zero in the smooth regions of the signal where it is well represented by its Taylor series. In fact, the experimental research [19] suggests that the number of vanishing moments required depends heavily on the application. In the second column of the table, three biorthogonal wavelets(bior1.1, bior2.4 and bior3.7) are listed in the form of "bior$M.N$". M and N indicates the vanishing moments of wavelet bases for analysis and synthesis. The benefit of the steerable pyramid is that it is translation-invariance (i.e. the subbands are aliasing-free) and rotation-invariant (i.e. the subbands are steerable) other than orientation selectivity in 6 directions.

3.2 Image Statistics in Wavelet Domain

The goal of our research is to evaluate the visual quality based on the amount of image information that can be perceived by human beings. To measure the information, the first step is to statistically describe images. This section reviews three statistical

models that have been widely used in wavelet analysis.

1. ***Simple Gaussian Model***

The simple Gaussian model assumes that wavelet coefficients in one subband obey the normal distortion and is independent of the other subbands.

$$f(\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_{\omega_i}} e^{-\frac{(\omega_i - \mu_{\omega_i})^2}{2\sigma_{\omega_i}^2}} \tag{3.1}$$

$$f(\omega_i, \omega_j) = f(\omega_i) f(\omega_j) \tag{3.2}$$

where

ω_i	the coefficient in the ith scale
μ_{ω_i}, $\sigma^2_{\omega_i}$	the mean and variance of the ith scale

Instead of sticking to a Gaussian assumption, researchers developed marginal or joint statistical models to describe the non-Gaussianity statistics. It is known that the marginal/joint distribution of coefficients under multi-resolution framework is non-Gaussian, as shown in Figure 3.1. Based on this observation, many statistical models [29] [26, 39–41, 82–85] are proposed, which are well summarized by Azimifar in [10]. Among the existing models, the following two models are representative and depict the image statistics in two different strategies. Crouse *et al.* [29] proposed a model to describe the intra-scale marginal statistics and also put the emphasis on the inter-scale dependencies, while Portilla *et al.* [82–85] focused on the intra-scale joint statistics. In this thesis, the inter-scale statistics are considered within the same direction. If not specified, the inter-scale or intra-scale statistics are only considered in the horizontal direction.

2. ***Hidden Markov Tree*** [29]

Wavelet coefficients can be represented by a hidden Markov tree which has the following two properties.

a. *Intra-scale statistics*: Each wavelet coefficient w_i at Scale i is modeled as being in one of two states θ_i: "large($\mathcal{L}$)", corresponding to a wavelet component containing significant contributions of signal energy, or "small($\mathcal{S}$)", representing coefficients with little energy. The long-tailed distribution in the ith scale can be described by a

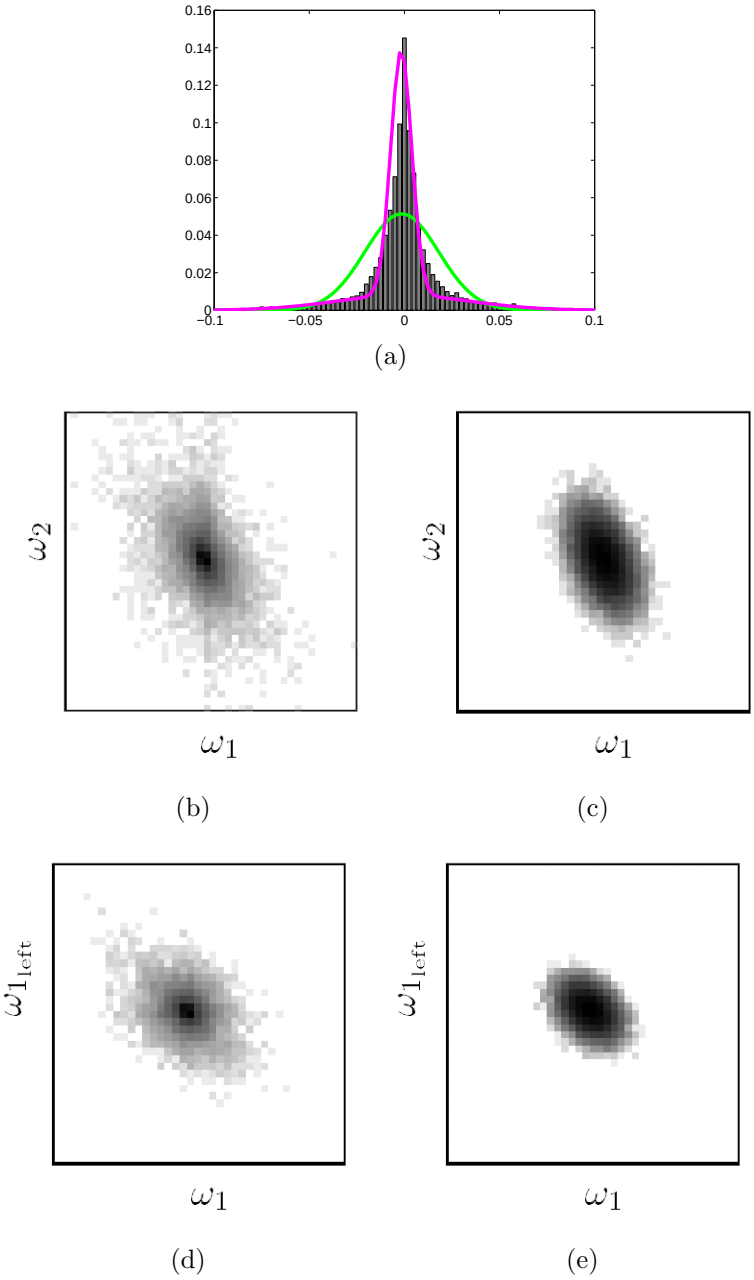


Figure 3.1: Non-Gaussianity image statistics under the db1 wavelet framework. Figure (a) plots the marginal distribution of the wavelet coefficients. The pink line indicates the distribution modeled by a Gaussian mixture model. The green line indicates the distribution modeled by a simple Gaussian model. Figure (b) displays the actual 2-D distribution of a child coefficient and its parent. Figure (c) displays the joint Gaussian distribution with the same covariance matrix as Figure (b). Figure (d) displays the joint statistics of a coefficient and its neighbor coefficient (the one on its left). Figure (e) displays the joint Gaussian distribution with the same covariance matrix as Figure (d).

two-state, zero-mean mixture model.

$$f(w_i) = p(\theta_i = \mathcal{s})g(w_i; 0, \sigma^2_{\mathcal{S};i}) + p(\theta_i = \mathcal{L})g(w_i; 0, \sigma^2_{\mathcal{L};w_i}) \tag{3.3}$$

$$\text{with } p(\theta_i = \mathcal{s}) + p(\theta_i = \mathcal{L}) = 1,\ g(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

b. *Inter-scale statistics*: A state transition matrix is defined as the state-to-state link between a child in the ith scale and its parent in the $i+1$th scale.

$$p(\theta_i|\theta_{i+1}) = \begin{bmatrix} p(\mathcal{s}|\mathcal{s}) & p(\mathcal{L}|\mathcal{s}) \\ p(\mathcal{s}|\mathcal{L}) & p(\mathcal{L}|\mathcal{L}) \end{bmatrix} \tag{3.4}$$

In the HMT model, the conditional independence property is assumed in Equation 3.5. In words, ω_i in the ith scale is conditionally independent of all other ω_j $(i \neq j)$ given its state θ_i. It means that $f(\underline{\omega}|\underline{\theta})$ is joint Gaussian. The probability density of $\underline{\omega}$ is given in Equation 3.6 [29].

$$f(\underline{\omega}|\underline{\theta}) = \prod_i f(\omega_i|\theta_i) \tag{3.5}$$

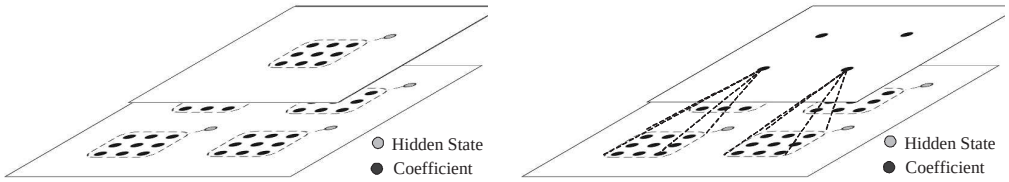
$$f(\underline{\omega}) = \sum_{\underline{\theta}} f(\underline{\omega}, \underline{\theta}) = \sum_{\underline{\theta}} f(\underline{\omega}|\underline{\theta}) f(\underline{\theta}) \tag{3.6}$$

There are several improved versions of the HMT model [39–41]. The improved hidden Markov tree model [40] captures the additional correlations between two adjacent scales: inter-scale statistics are not only limited to a coefficient and its parent, but also the inter-scale coefficient around its parent. The hidden Markov model-three subbands model [41] simulates the joint statistics across different directions (horizontal, vertical and diagonal direction).

3. ***Gaussian Scale Mixtures*** [84]

The Gaussian scale mixture (GSM) model describes the joint distribution of wavelet coefficients in a defined neighborhood as:

$$f(\underline{w_i}) = \sum_j p(z_j) g(\underline{w_i}; \mathbf{0}, z_j \mathbf{\Sigma}_{\underline{w_i}}) \tag{3.7}$$



(a) Intra-scale neighborhood 3×3 (b) Inter-scale neighborhood $3 \times 3 + 1$

Figure 3.2: The neighborhood defined in Gaussian scale mixture model

where

$\underline{w_i}$ the vector formed by the coefficients in the defined neighborhood, such as 3×3 window.

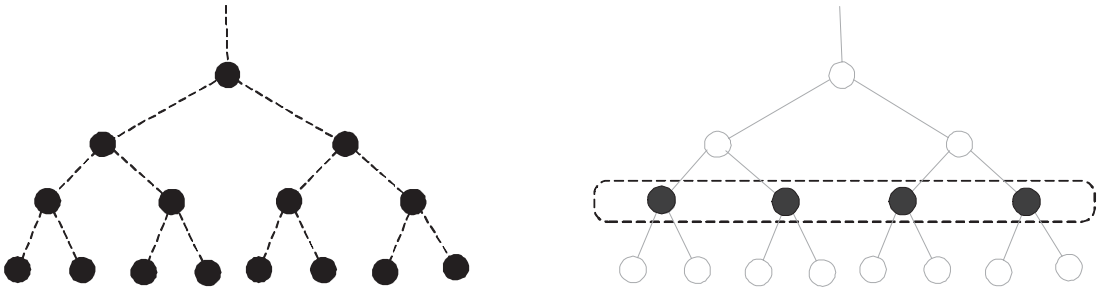
z_j the hidden state, which is logarithmically sampled across a wide range.

The basic version of the GSM model only considers the intra-scale statistics by limiting the defined neighborhood within a scale. Also the GSM model is able to describe the inter-scale statistics by defining a $3 \times 3 + 1$ neighborhood. The extra coefficient is the parent coefficient in the coarse scale, as shown in Figure 3.2. For different hidden states, the corresponding covariance matrix $z_j \mathbf{\Sigma}_{\underline{w_i}}$ only varies in magnitude. It means that the GSM model describes the non-Gaussianity statistics in one specific direction. There are other improved GSM models that consider orientation adaption [51], space variance [48].

Figure 3.3 summarizes the characteristics of the HMT and GSM models. Figure 3.3(a) illustrates that the HMT model describes image statistics across the whole tree (black nodes on the tree). The dashed lines indicate the hidden state dependency (transfer matrix) between a parent coefficient and its children. Figure 3.3(b) shows that the GSM model emphasizes the intra-scale statistics. The dashed circle means the joint statistics within a certain scale. In the last section of Chapter 4, these three statistical models are used to capture image statistics and compute the visual information that can be perceived by human beings.

3.3 Image Energy Model

There is another type of image model which describes the image energy spectra in the Fourier domain or the wavelet domain. Field [42] found that the energy spectra of natural images are quite characteristic and are quite different from that of white



(a) Neighborhood of HMT model (b) Neighborhood of GSM model

Figure 3.3: The difference between the standard HMT and GSM model. Figure (a) illustrates that the HMT model describes image statistics across the whole tree (black nodes on the tree). The dashed lines indicate the hidden state transfer matrix. Figure (b) shows that the GSM model emphasizes the intra-scale statistics. The dashed circle means the joint statistics within a certain scale.

noise. They show the greatest amplitude at low frequencies and decreasing amplitude as the frequency increases. The amplitude falls off quickly by a factor of roughly $1/f$. This $1/f$ behavior fall off energy spectrum was also found in the wavelet domain by Wornell [134–136].

In this section, we propose two simple models to describe the relationship among the energy in different scales.

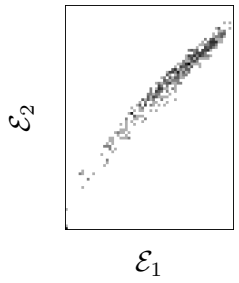
Definition 3.3.1 *The energy $\mathcal{E}_i$ in the ith scale is defined as the wavelet coefficient variance.*

600 natural images with excellent quality were decomposed with the db2 wavelet and the energy in the first two scales were computed. Figure 3.4 displays the energy of the first two scales in log-log coordinates. The test results illustrate the strong dependency between the energy in the first two scales. If the energy $\mathcal{E}_i$ in the ith scale is treated as a random variable, the linear correlation between two random variables $\log \mathcal{E}_i$ and $\log \mathcal{E}_{i+1}$ can be computed according to Equation 3.8, which is called the linear correlation of the energy or the energy correlation. For the first two adjacent scales, the energy correlations are 0.9694, 0.9718 and 0.9686 for the horizontal, vertical and diagonal directions.

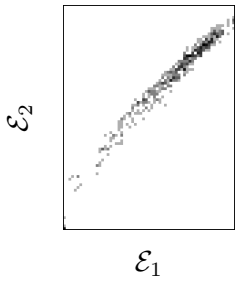
$$\rho_{\log \mathcal{E}_i, \log \mathcal{E}_{i+1}} = \frac{\operatorname{cov}(\log \mathcal{E}_i, \log \mathcal{E}_{i+1})}{\sigma_{\log \mathcal{E}_i}\, \sigma_{\log \mathcal{E}_{i+1}}} \tag{3.8}$$

	1H	2H	3H	4H	1D	2D	3D	4D	1V	2V	3V	4V
1H	1	**0.9694**	0.8963	0.7604	0.1437	0.0601	-0.0684	-0.1525	0.2983	0.2114	0.0686	-0.0360
2H		1	**0.9491**	0.8201	0.1206	0.0903	-0.0228	-0.1026	0.2287	0.2029	0.0782	-0.0192
3H			1	**0.8791**	0.0679	0.0636	0.0361	-0.0387	0.1390	0.1353	0.1082	0.0184
4H				1	0.0075	0.0163	0.0056	0.1206	0.0458	0.0533	0.0421	0.1747
1D					1	**0.9718**	0.9099	0.8133	0.2320	0.2119	0.1586	0.0975
2D						1	**0.9611**	0.8743	0.1610	0.1958	0.1690	0.1194
3D							1	**0.9238**	0.0714	0.1256	.01963	0.1508
4D								1	0.0036	0.0565	0.1265	0.2665
1V									1	**0.9686**	0.8946	0.7715
2V										1	**0.9498**	0.8345
3V											1	**0.8963**
4V												1

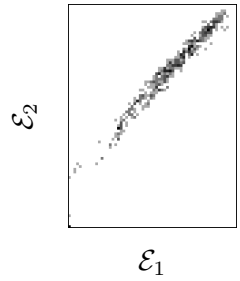
Table 3.2: The linear correlation of the energy in different subbands. The test results indicate that the energy correlation is strongest between two adjacent scales in the same direction. Also there are notable correlations between non-adjacent scales (such as 1H and 3H) in the same direction. The correlation decreases greatly when two subbands belong to different directions (such as 1H and 1D).



(a) Horizontal direction



(b) Vertical direction



(c) Diagonal direction

Figure 3.4: 2-D histogram of coefficient energy in two adjacent scales. 2-D histogram is generated by over 600 natural pictures. The multi-resolution framework is the db2 wavelet. The linear correlations of energy in two adjacent scales are 0.9694, 0.9718 and 0.9686 in the horizontal, vertical and diagonal directions.

Now the experiments are extended to all subbands under the db2 wavelet framework. The linear correlations between the energy of different subbands are computed and listed in Table 3.3. The test results indicate that the linear correlation of the energy is strongest between the two adjacent scales in the same direction. Also there are notable correlations between non-adjacent scales (such as 1H and 3H) in the same direction. The correlation decreases greatly when two subbands belong to different directions (such as 1H and 1D).

Definition 3.3.2 *Based on the strong correlation among the energy of different scales in the same direction, the energy in the ith scale can be estimated with the energy in the higher scale, as shown Equation 3.9, where* β_i *is a constant and* $\underline{\alpha_i}$ *is a constant*

vector for the ith scale. This method is called the ***energy consistency model****.*

$$\log \hat{\mathcal{E}}_i = \begin{bmatrix} \log \mathcal{E}_{i+1} & \log \mathcal{E}_{i+2} & \dots \end{bmatrix} \begin{bmatrix} \underline{\alpha_i}_1 \\ \underline{\alpha_i}_2 \\ \vdots \end{bmatrix} + \beta_i \tag{3.9}$$

The parameters ($\underline{\alpha_i}$ and β_i) estimation is the classic linear least squares problem [63], as shown in Equation 3.10. 600 natural images are used to estimated $\underline{\alpha_i}$ and β_i. Let us denote the energy in the ith scale for the jth image as $\mathcal{E}_{i,j}$, $\underline{\alpha_i}$ and β_i can estimated as follows [63].

$$\underline{\hat{\alpha}}_i, \hat{\beta}_i = \arg\min_{\underline{\alpha_i}, \beta_i} || \mathbf{Y}_i - \mathbf{E}_i \begin{bmatrix} \underline{\alpha_i} \\ \beta_i \end{bmatrix} ||^2 \tag{3.10}$$

where $\mathbf{E}_i = \begin{pmatrix} \log \mathcal{E}_{i+1,1} & \log \mathcal{E}_{i+2,1} & \dots & \log \mathcal{E}_{M,1} & 1 \\ \log \mathcal{E}_{i+1,2} & \log \mathcal{E}_{i+2,2} & \dots & \log \mathcal{E}_{M,2} & 1 \\ \vdots & \vdots & \ddots & & 1 \\ \log \mathcal{E}_{i+1,600} & \log \mathcal{E}_{i+2,600} & \dots & \log \mathcal{E}_{M,600} & 1 \end{pmatrix}$, $\mathbf{Y}_i = \begin{bmatrix} \log \mathcal{E}_{i,1} \\ \log \mathcal{E}_{i,2} \\ \vdots \\ \log \mathcal{E}_{i,600} \end{bmatrix}$ and $||.||$ is the magnitude of the vector.

Basically the energy consistency model implies that the energy at the high frequency can be estimated the energy at the low frequency. One application of this energy consistency model is to estimate the distortion energy in each scale if only the distorted image is accessible. If the distortion barely changes images at the low frequency but disturbs the image at the high frequency greatly, the energy of the original image at the high frequency can be estimated with the energy consistency model and the distortion in the high frequency may be derived. This idea provides a solution for blind quality assessment since the distortion energy the at high frequency has strong connection with the visual quality of an image, which will be revealed in Chapter 6.

In the following session, we will discuss the estimation of the distortion energy. First let us take a look at a common distortion: white noise distortion. Suppose white noise D with variance σ_D^2 is applied to a natural image as Equation 3.11. Let $\mathcal{E}_i$ and $\mathcal{E}_{di}$ donate the energy of the ith scale in the original image X and the distorted image X_d respectively. Under the orthogonal wavelet framework, the noise in each scale is still white noise $g(0, \sigma_D^2)$ and independent of the image [70]. Hence the relationship between

$\mathcal{E}_i$ and $\mathcal{E}_{di}$ can be written in Equation 3.12.

$$X_d = X + D \tag{3.11}$$
$$\mathcal{E}_{di} = \mathcal{E}_i + \sigma_D^2 \tag{3.12}$$

Figure 3.5 displays the energy in each scale and the corresponding images in the spatial domain. Although the white noise variance in every scale is the same, it affects images less and less as the scale level increases. For instance, the energy of the noisy images in the high scales (the 4th and 5th scales) are almost the same as the original image as shown in Figure 3.5(d). Hence the energy in the 4th and 5th scales of the noisy image can be taken as the energy in the original image[1]. According to the energy consistency model, $\log \mathcal{E}_3$ in the 3th scale can be estimated by Equation 3.13, where $\log \hat{\mathcal{E}}_5 \simeq \log \mathcal{E}_{d5}$ and $\log \mathcal{E}_4 \simeq \log \hat{\mathcal{E}}_{d_4}$. Similarly, the energy in the ith scale can be estimated by the $i+1th, i+2th, ...$ scales as shown in Equation 3.14.

$$\log \hat{\mathcal{E}}_3 = \begin{bmatrix} \log \hat{\mathcal{E}}_4 & \log \hat{\mathcal{E}}_5 & 1 \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha}_3}_1 \\ \underline{\hat{\alpha}_3}_2 \\ \hat{\beta}_3 \end{bmatrix} \tag{3.13}$$

$$\log \hat{\mathcal{E}}_i = \begin{bmatrix} \log \hat{\mathcal{E}}_{i+1} & \log \hat{\mathcal{E}}_{i+2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha}_i}_1 \\ \underline{\hat{\alpha}_i}_2 \\ \vdots \\ \hat{\beta}_i \end{bmatrix} \tag{3.14}$$

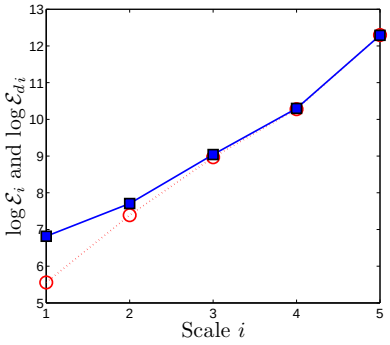
Hence the noise variance in each subband can be estimated Equation 3.15.

$$\hat{\sigma_D^2} = \begin{cases} \mathcal{E}_{di} - e^{\log \hat{\mathcal{E}}_i} & \text{if} \quad \mathcal{E}_d > e^{\log \hat{\mathcal{E}}_i} \\ 0 & \text{else} \end{cases} \tag{3.15}$$

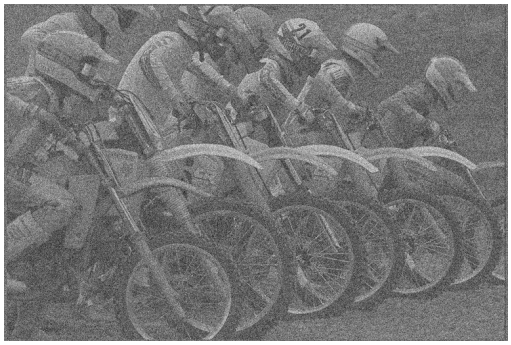
[1]If more white noise is applied, the energy at the low frequency of the noisy image might be much larger than the energy in the original image. Then it will be improper to use the approximation. According to the subjective quality evaluation results provided by the LIVE database [99], the examples in Figure 3.5(b) and 3.5(d) have the visual quality score of 35.15 ("fair quality" [99]) and 46.24 ("poor quality" [99]). If more noise is applied, then the visual quality will be worse than "poor quality". It is less meaningful for people to evaluate the images with bad quality. Hence, the energy in the high scales of the noisy image can be approximately taken as the energy in the original image, if the visual quality is acceptable.

(a)

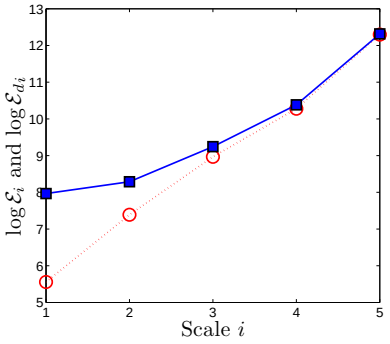
(b)



(c)



(d)



(e)

Figure 3.5: The energy in each scale. The images are provided by the LIVE database [99]. Figure (a) shows the original image in 8-bit gray scale. Figure (b) and (d) are generated by adding white noise. PSNR are 20dB and 13.9dB respectively. The figures in the right column display the energy of the noisy images in the db2 wavelet domain compared with the original image. ○ indicates the energy $\mathcal{E}_i$ in the original image, and ■ indicate the energy in the noisy image $\mathcal{E}_{di}$.

	White noise variance σ_D^2		
	[0, 0.05)	[0.05, 0.1)	[0.1, 0.2]
1H	3.841e-3	8.114e-3	1.646e-2
1V	**3.575e-3**	**8.460e-3**	1.628e-2
1D	6.297e-3	1.204e-2	1.945e-2
2H	2.326e-2	9.162e-3	1.278e-2
2V	2.172e-2	1.244e-2	**1.135e-2**
2D	4.912e-3	8.184e-3	1.644e-2
3H	2.076e-2	1.091e-2	1.309e-2
3V	2.102e-2	1.234e-2	1.353e-2
3D	2.098e-2	1.067e-2	1.402e-2

Table 3.3: The white noise variance estimation using the energy consistency model. The noisy images are provided by the LIVE database [99]. The variance of the white noise varies from 0 to 0.20, while the maximum pixel value of the original images is 1.0. The white noise variance is derived by the estimated variance in the first three scales in three directions. "1H" means the first scale in the horizontal direction. The table lists the mean squared error between the true variances and the estimation results. The least error in each variance zone is listed in bold.

In theory, the white noise variance can be estimated with any wavelet scale, i.e. $\hat{\sigma_D^2} = \mathcal{E}_{d1} - e^{\log \hat{\mathcal{E}}_1}$, $\mathcal{E}_{d2} - e^{\log \hat{\mathcal{E}}_2}$. Table 3.3 lists the noise variance estimation using different scale. In the experiment, the noisy images are provided by the LIVE database [99]. 174 noisy images were generated by adding the white noise to 29 natural images. The variance of the white noise varies from 0 to 0.20, while the maximum pixel value of the original image is 1.0. Our method is tested under three noise variance zones: $\sigma_D^2 \in [0, 0.05]$, $\sigma_D^2 \in [0.05, 0.1)$ and $\sigma_D^2 \in [0.1, 0.2]$. Each noisy image is decomposed by the db2 wavelet into 5 scales in three directions. The white noise variance is derived with the estimated energy in the first three scales in three directions, such as the first scale in the horizontal direction ("1H"). The test results in Table 3.3 show that the noise variance estimation using the first scale has the best performance for small noise variance. To estimate the large white noise variance, using the higher scale provides better performance (less estimation error). We also compare our white noise variance estimation method with other existing methods: the classic noise variance estimation method by Donoho and Johnstone [36] and the color information based noise estimation method by Staelin and Nachlieli [106]. Table 3.4 lists the mean squared error between the true variance and the estimation results. Our method uses the first scale in the horizontal direction. Compared with the other two existing methods, our method has

White noise variance σ_d^2	[0, 0.05)	[0.05, 0.1)	[0.1, 0.2]
Our method [36]	3.841e-3	**8.114e-3**	**1.646e-2**
Donoho and Johnstone's method [36]	**1.434e-3**	8.422e-3	2.870e-2
Staelin and Nachlieli's method [106]	6.031e-3	2.619e-2	3.679e-2

Table 3.4: The performance of white noise variance estimation. The noisy images are provided by the LIVE database [99]. The variance of the white noise varies from 0 to 0.20, while the maximum pixel value of the original images is 1.0. Each method is tested under three noise variance zones. The table lists the mean squared error between the true variances and the estimation results. The least error in each variance zone is listed in bold.

the best performance when $\sigma_D^2 \in [0.05, 0.2]$ (medium noise). Staelin and Nachlieli's method [106] outperforms the others when $\sigma_D^2 \in [0, 0.05]$ (small noise).

The energy consistency model can also be applied to estimate Gaussian blur distortion variance. In Chapter 6, it is revealed that the visual quality of an image with the Gaussian blur has strong connection to the distortion variance at the high frequency. The Gaussian blur can be modeled in the wavelet domain by Equation 3.16 [97], where X_i is the original wavelet coefficient in the ith subband and X_{di} is the wavelet coefficient of the distorted (blurred) image. The parameter k controls the degree of blur. Since the Gaussian blur has a low pass filter behavior, it affects wavelet coefficients less and less as the scale level increases.

$$X_{d_i} = kX_i \tag{3.16}$$

Let D_i donate the difference between the original image X_i and the distorted image X_{d_i} in the $i$the scale. With the energy consistency model, the estimation of the blur distortion variance $\sigma_{D_i}^2$ in the high frequency is feasible as shown in Equation 3.21.

$$\begin{aligned}
\sigma_{D_i}^2 &= E|X_i - X_{di}|^2 = E|X_i - kX_i|^2 && (3.17)\\
&= (1-k)^2 E|X_i|^2 && (3.18)\\
&= (1-k)^2 \sigma_{X_i}^2 && (3.19)\\
&= (1-k)^2 \mathcal{E}_i && (3.20)
\end{aligned}$$

$$\hat{\sigma^2_{D_i}} = (1-\hat{k})^2\hat{\mathcal{E}}_i \tag{3.21}$$

$$\hat{k^2} = \frac{\sigma^2_{X_{di}}}{\hat{\sigma^2_{X_i}}} = \frac{\mathcal{E}_{X_{di}}}{\hat{\mathcal{E}}_{X_i}} \tag{3.22}$$

Equation 3.19 comes from the assumption that the wavelet coefficient in a subband has zero mean [70]. Since the energy of a subband is defined as the coefficient variance, $\sigma^2_{X_i}$ is substituted by $\mathcal{E}_i$ in Equation 3.20.

The energy consistency model considers the energy of coefficients over the entire subband. This model is the foundation for an approach to blind quality assessment, especially for distortions that have great impact on the image energy spectra, such as white noise or Gaussian blur. But this model has some limitation. It is not suitable to evaluate fastfading distortion, which is a common distortion when the communication is interfered. An image with the fast fading distortion is shown in Figure 6.16. Figure 3.6(a) displays the first and second scale in the horizontal direction of the distorted image (with fastfading distortion). Figure 3.6(b) displays the first and second scale in the horizontal direction of the reference image (the original image). Figure 3.6(c) and Figure 3.6(d) display the distorted image and the reference image in the spatial domain. It is clear that fastfading distortion causes the wavelet coefficients in the first scale to be circularly shifted. But the circular shift does not change the energy in the first scale. It means that using the energy consistency model will not be helpful to estimate the distortion in the first scale. We notice that fastfading distortion undermines on the wavelet second property [70]: large/small coefficients tend to have large/small children. In the rest part of this chapter, we propose the local energy consistency model in order to capture the distortion on the wavelet second property.

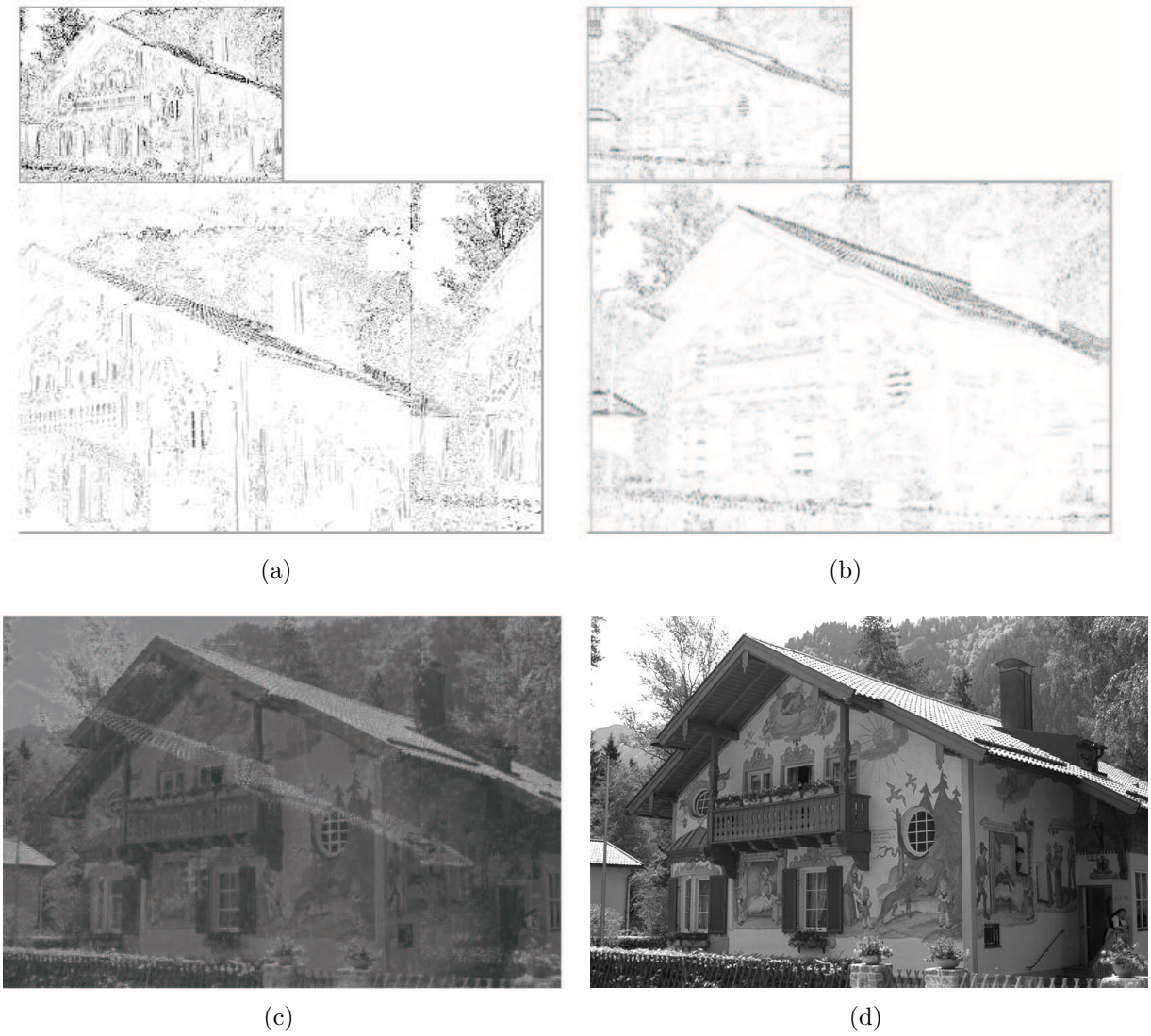
Definition 3.3.3 *Local energy $\mathcal{E}_i(m_i, n_i)$ in the ith scale is defined by Equation 3.23*

.

$$\mathcal{E}_i(m_i, n_i) = \frac{\sum_{(p,q)\in\mathcal{N}} [X_i(p,q) - \mu_{m_i,n_i}]^2}{M \times N} \tag{3.23}$$

where $X_i(p,q)$ is the coefficient in the $i$the scale at (p,q). $\mathcal{N}$ is the defined neighborhood, such as a 3×3 window. M and N are the width and height of the defined neighborhood $\mathcal{N}$. μ_{m_i,n_i} is the average within the neighborhood $\mathcal{N}$.

The correlation between the local energy in two adjacent scales is tested in the fol-



(a) (b)

(c) (d)

Figure 3.6: The image with fastfading distortion and its representation in the db2 wavelet domain. Figure (a) displays the first and second scale in the horizontal direction of the distorted image (with fastfading distortion). Figure (b) displays the first and second scale in the horizontal direction of the reference image (the original image). Figure 3.6(c) and Figure 3.6(d) display the distorted image and the reference image in the spatial domain. It is clear that fastfading distortion causes the wavelet coefficients in the first scale to be circularly shifted.

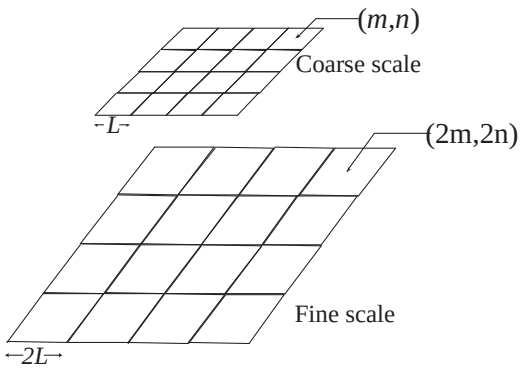


Figure 3.7: Local energy consistency experiment

lowing experiment. 50 natural images with the medium size[1] were involved in the experiment. The details are listed as follows.

1. Define the neighborhood in the parent scale. The neighborhood can be defined as a square area with side L, as shown in Figure 3.7.

2. Divide the parent scale, i.e. the $i+1$th scale in the horizonal direction, into non-overlapped neighborhoods centered at (m_{i+1}, n_{i+1}), i.e. $(L/2 + pL, L/2 + qL)$, where p and q are natural numbers.

3. Define the neighborhood in the child scale, i.e. the ith scale in the horizonal direction. The neighborhood is defined as a square area with side $2L$.

4. Divide the child scale into non-overlapped neighborhoods, with the centers at (m_i, n_i), i.e. $(L + 2pL, L + 2qL)$, where p and q are natural numbers. Such arrangement will make sure that each neighborhood in the parent scale has a corresponding neighborhood in the child scale.

5. Compute local energy $\mathcal{E}_{i+1}(m_{i+1}, n_{i+1})$ and $\mathcal{E}_i(m_i, n_i)$ by Equation 3.23.

6. Compute the linear correlation between the local energy in the parent scale and its child scale. If the local energy $\mathcal{E}_i(m_i, n_i)$ is treated as a random variable, the linear correlation of the local energy is defined in Equation 3.24, which is also called the local energy correlation.

$$\rho_{\log \mathcal{E}_i(m_i,n_i),\log \mathcal{E}_{i+1}(m_{i+1},n_{i+1})} = \frac{\mathrm{cov}\,(\ \log \mathcal{E}_i(m_i, n_i), \log \mathcal{E}_{i+1}(m_{i+1}, n_{i+1})\ )}{\sigma_{\log \mathcal{E}_i(m_i,n_i)}\ \sigma_{\log \mathcal{E}_{i+1}(m_{i+1},n_{i+1})}} \quad (3.24)$$

[1]Image size varies from 480×625 to 512×768 [99].

7. Choose the size of the neighborhood in the parent scale from 3×3 to 9×9.

Table 3.5 shows the experiment results: the linear correlation between the local energy in the ith and $i+1$th scales. We notice that local energy has strong correlation between scales, and this behavior is stronger when the neighborhood size increases. Figure 3.8 displays the two dimensional histogram of the local energy in the log-log coordinates under the db97 wavelet. The histogram verifies the wavelet secondary property [70]: the local energy at (m_i, n_i) in the fine scale tends to be large/small when the local energy at the corresponding location (m_{i+1}, n_{i+1}) in the coarse scale is large/small.

For the fastfading example in Figure 6.16, the linear correlation between the local energy in the first two scales are 0.23 and 0.85 for the distorted image and the reference image. Hence the linear correlation between the local energy can indicate the impact of fastfading distortion on the wavelet secondary property. If the wavelet secondary property is only distorted within certain areas, how to identify these areas? A new concept, the local energy set, will help us.

Definition 3.3.4 *The* ***local energy set*** $\{\mathcal{E}_i(m_i, n_i)\}$ *in the ith scale is defined within a neighborhood entered at (m_i, n_i) and its surrounding neighborhoods. The local energy set is composed of the local energy of the above neighborhoods, as shown in Figure 3.9. The size of the local energy set could be 3×3 (9 neighborhoods), 4×4 or 5×5.*

As shown in Figure 3.9, the local energy set $\{\mathcal{E}_i(m_i = a, n_i = b)\}$ in the ith scale is indicated with the shaded area, as well as its corresponding local energy set $\{\mathcal{E}_{i+1}(m_{i+1} = 2a, n_{i+1} = 2b)\}$ in the $i+1$th scale. According to the experiment details (Part 2 and Part 4), each local energy set in the ith scale has the corresponding local energy set in the $i+1$th scale.

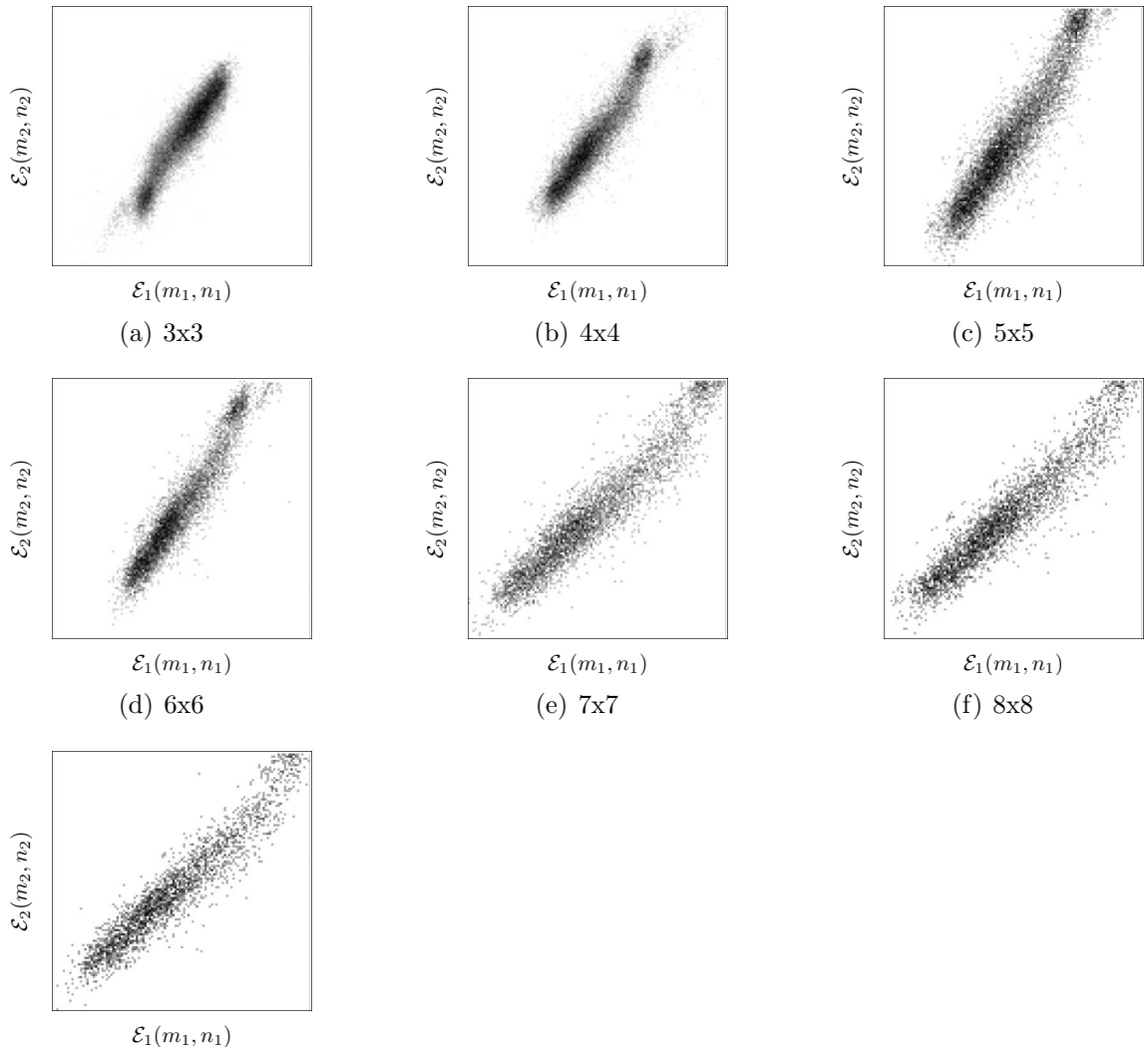
The correlation between two local energy sets is defined as the linear correlation between the local energy in each set, as shown in Equation3.25. To simplify the notation, let x_j and y_j denote the log of the local energy in the local energy set $\{\mathcal{E}_i(m_i, n_i)\}$ and its corresponding local energy set $\{\mathcal{E}_{i+1}(m_{i+1}, n_{i+1})\}$. From the definition in Equation 3.25, $|\rho(m_i, n_i; m_{i+1}, n_{i+1})| \leq 1$. If $|\rho(m_i, n_i; m_{i+1}, n_{i+1})|$ is close to 1, it means the two local energy sets have strong correlation. According to the wavelet secondary property, a local energy set should have strong correlation with its corresponding local energy set in the adjacent scales.

$$\rho(m_i, n_i; m_{i+1}, n_{i+1}) = \frac{n \sum_j x_j y_j - \sum x_j y_j}{\sqrt{n \sum x_j^2 - (\sum x_j)^2} \sqrt{n \sum y_j^2 - (\sum y_j)^2}} \tag{3.25}$$

Wavelet	Consistency between 1*st* Scale and 2*nd* Scale						
	3×3	4×4	5×5	6×6	7×7	8×8	9×9
db1	0.8941	0.9016	0.9071	0.9083	0.9128	0.9122	0.9147
db2	0.8682	0.8845	0.8886	0.8928	0.8925	0.8911	0.8912
db3	0.8983	0.9027	0.9083	0.9093	0.9047	0.9034	0.9030
bior1.1	0.9068	0.9084	0.9119	0.9118	0.9140	0.9126	0.9161
bior1.3	0.9011	0.8906	0.8823	0.8775	0.8726	0.8753	0.8701
bior1.5	0.8769	0.8744	0.8764	0.8853	0.8869	0.8913	0.8935
rbio1.1	0.8982	0.9019	0.9064	0.9082	0.9107	0.9097	0.9108
rbio1.3	0.8959	0.8999	0.9041	0.9046	0.8987	0.9036	0.8957
rbio1.5	0.8822	0.8875	0.9004	0.9037	0.9020	0.9055	0.9042
db97	**0.8372**	**0.8924**	**0.9182**	**0.9110**	**0.9294**	**0.9284**	**0.9371**
Wavelet	Consistency between 2*nd* Scale and 3*rd* Scale						
	3×3	4×4	5×5	6×6	7×7	8×8	9×9
db1	0.7660	0.7790	0.7835	0.7981	0.8409	0.8537	0.8780
db2	0.7660	0.8227	0.8687	0.8684	0.9020	0.8664	0.9224
db3	0.6871	0.7602	0.8120	0.8232	0.8754	0.8817	0.8926
bior1.1	0.8873	0.9031	0.9227	0.9228	0.9399	0.9285	0.9414
bior1.3	0.7115	0.7686	0.8598	0.8100	0.9063	0.813	0.9212
bior1.5	0.7637	0.6685	0.7677	0.7650	0.8571	0.8111	0.8607
rbio1.1	0.8873	0.9031	0.9227	0.9228	0.9399	0.9285	0.9414
rbio1.3	0.7124	0.7586	0.8355	0.7994	0.8901	0.8050	0.8969
rbio1.5	0.7521	0.6552	0.7486	0.7251	0.8498	0.7685	0.8540
db97	**0.7660**	**0.7790**	**0.7835**	**0.7981**	**0.8409**	**0.8537**	**0.8780**

Table 3.5: Local energy consistency

The table lists the linear correlation between two energy sets $\{\log \mathcal{E}_{i+1}(m_{i+1}, n_{i+1})\}$ and $\{\log \mathcal{E}_i(m_i, n_i)\}$, where m_i, n_i are the valid coordinators. In Chapter 6, fastfading distortion is evaluated in the db97 wavelet domain. We are interested the energy consistency in the db97 wavelet domain.



(a) 3x3 (b) 4x4 (c) 5x5

(d) 6x6 (e) 7x7 (f) 8x8

(g) 9x9

Figure 3.8: 2-D histogram of the local energy in the first and second scale as the neighborhood size changes.

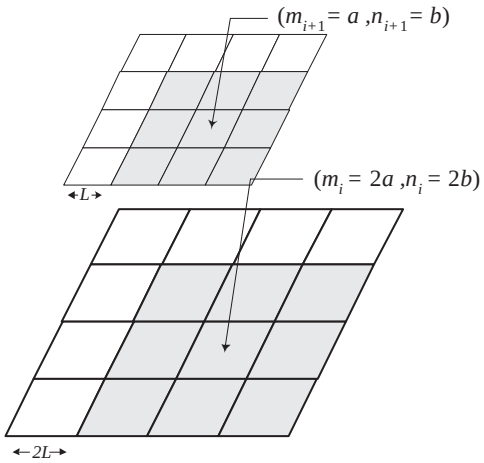


Figure 3.9: Local energy set. An example of the local energy set $\{\mathcal{E}_i(m_i = a, n_i = b)\}$ in the ith scale is indicated with the shaded area, as well as its corresponding local energy set $\{\mathcal{E}_{i+1}(m_{i+1} = 2a, n_{i+1} = 2b)\}$ in the $i + 1$th scale. The size of the local energy set is 3×3.

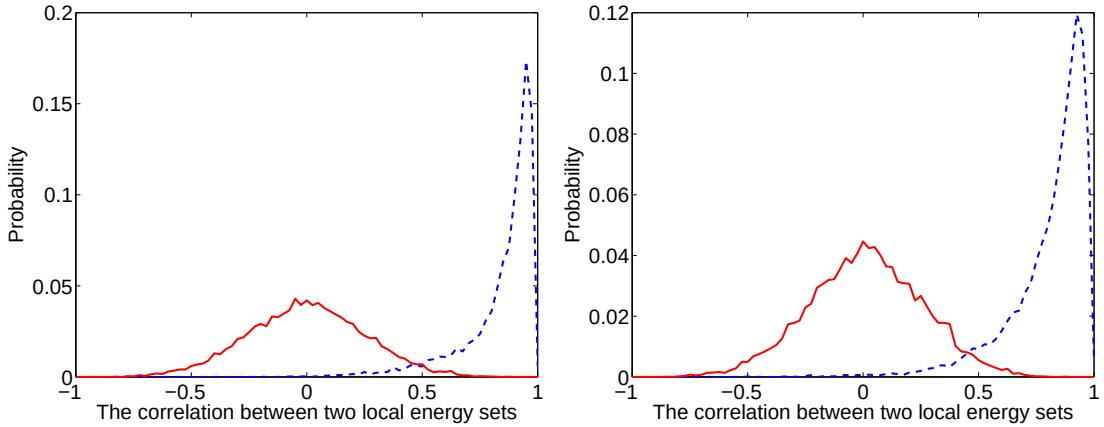
where n is the number of the local energy contains in the local energy set. For the local energy set with the size of 3×3, $n = 9$.

Table 3.6 lists the correlation between the local energy sets in the ith and $i + 1$th scales ($i = 1, 2$). For a local energy set and its corresponding energy set, the average correlation is greater than 0.75, the standard deviation is less than 0.3. Please note that the distribution of the local energy set correlation is not Gaussian, but more close to a Poisson distribution, as shown in Figure 3.10. From Table 3.6, we also notice that the average correlation increases and the standard deviation decreases as the local energy set size increases. For two random local energy sets, the correlation average is very close to zero and the distribution is similar to a Gaussian distribution.

To be clear, ***the wavelet secondary property is distorted*** if the wavelet coefficients are shifted from their original position due to some distortion. To identify whether there is a distortion on a position, our solution is to compute the correlation between the local energy set centered at that position and the corresponding local energy set in the adjacent scale. If the correlation ρ is less than certain threshold T_c, the position is classified to have the wavelet secondary property distorted (the first category c_1), otherwise the position is classified to be normal (the second category c_2). Figure 3.10 plots the probability ($p(\rho|c_1)$ and $p(\rho|c_2)$) of the local energy set correlation where there is/ is not distortion. There are two ways in which a classification

Wavelet Framework	Local Energy Set Size	A local energy set & its corresponding set		A local energy set & a random set	
		average	standard deviation	average	standard deviation
The 1st and 2nd scales	3×3	0.7809	0.2473	1.168e-4	0.3877
	4×4	0.8118	0.1968	2.169e-4	0.3024
	5×5	0.8295	0.1743	7.700e-3	0.2551
	6×6	0.8436	0.1587	1.113e-4	0.3887
The 2nd and 3rd scales	3×3	0.7530	0.2817	2.810e-3	0.3942
	4×4	0.7886	0.2291	5.912e-3	0.3062
	5×5	0.8169	0.1936	4.253e-3	0.2600
	6×6	0.8346	0.1726	8.124e-3	0.2322

Table 3.6: The correlation between local energy sets. The table lists the average and standard deviation of the correlation between two local energy sets. The test results show that there is strong correlation (the large average and small deviation) between the local energy set and its corresponding set. This verifies the wavelet secondary property. The table also shows very weak correlation between the local energy set in the ith scale and a random local energy set in the $i+1$th scale. The test uses 200 natural images and the db1 wavelet. The neighborhood of the local energy is 5×5 and 10×10 in the $i+1$th and ith scale respectively.



(a) (b)

Figure 3.10: The probability of the local energy set correlation under the db1 wavelet. Figure (a) plots the probability of the local energy set correlation between the first and second scales. Figure (b) plots the probability of the local energy set correlation between the second and third scales. The solid lines (red) indicate the correlation between two random local energy sets, while the dashed lines (blue) indicate the correlation between the local energy set and its corresponding set in the adjacent scale.

error can occur; either an observation $\rho < T_c$ and the true category is c_2, or $\rho > T_c$ and the true category is c_2. If we assume the probability of each category is equal, the classification error probability is:

$$\begin{aligned} P(error) &= p(\rho > T_c, c_1) + p(\rho < T_c, c_2) && (3.26) \\ &= \int_{\rho > T_c} p(\rho|c_1)p(c_1) + \int_{\rho < T_c} p(\rho|c_2)p(c_2) && (3.27) \end{aligned}$$

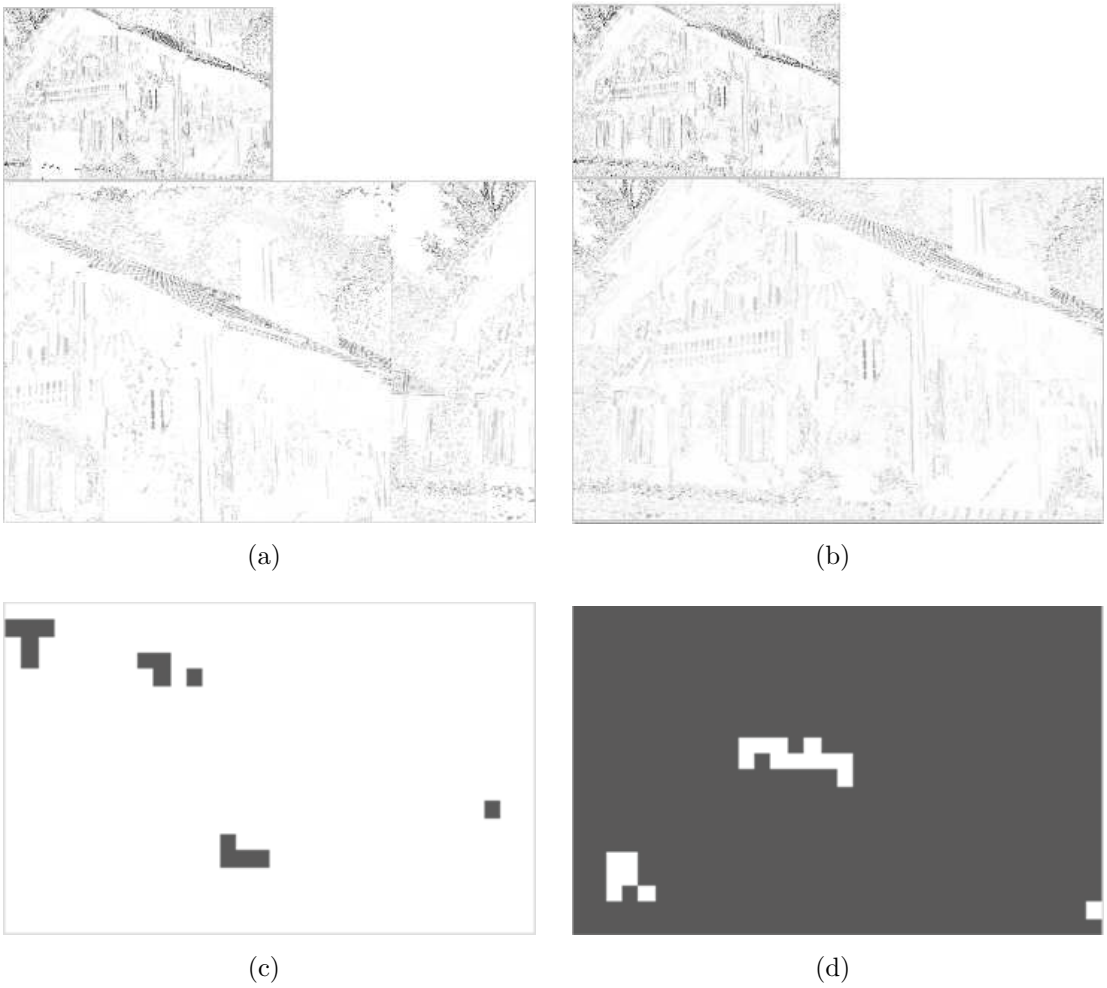
The threshold T_c is chosen that the error probability in Equation 3.27 is minimized. According to the probability of the local energy set correlation in Figure 3.10, T_c can be determined where $p(\rho|c_1)$ and $p(\rho|c_2)$ cross each other. To identify the wavelet secondary property distortion in the first scale and second scales, $T_c = 0.4825$ based on Figure 3.10(a) and $P(error) = 0.09$. To identify the wavelet secondary property distortion in the second scale and third scales, $T_c = 0.4695$ based on Figure 3.10(b) and $P(error) = 0.14$.

Back to the example in Figure 6.16, the wavelet secondary property is distorted within the entire first scale due to fastfading distortion. In Figure 3.11(c), the distortion is identified by comparing the correlation ρ and T_c. Figure 3.11(d) displays the identification result on a reference image. The white areas in the figures indicate the damaged areas determined by our method.

Definition 3.3.5 *Based on the strong correlation among the* ***local energies*** *of different scales in the same direction, the local energy in the ith scale can be estimated with the local energy in the higher scales, as shown Equation 3.28, where β_i is a constant and $\underline{\alpha_i}$ is a constant vector for the ith scale. Since the estimation is based on strong correlation, this method is called the* ***local energy consistency model****.*

$$\log \hat{\mathcal{E}}_i(m_i, n_i) = \begin{bmatrix} \log \mathcal{E}_{i+1}(m_{i+1}, n_{i+1}) & \log \mathcal{E}_{i+2}(m_{i+2}, n_{i+2}) & \dots \end{bmatrix} \begin{bmatrix} \underline{\alpha_i}_1 \\ \underline{\alpha_i}_2 \\ \vdots \end{bmatrix} + \beta_i \quad (3.28)$$

The parameters $\underline{\alpha_i}$ and β_i can be estimated as with the energy consistency model in the last section. If only the distorted image is accessible, the local energy in the



(a) (b) (c) (d)

Figure 3.11: The example to illustrate the distortion on the wavelet secondary property. Figure (a) displays the first two scales of a test sample from the LIVE database [99] in the db97 wavelet [3] domain. Figure (b) displays the first two scales of the reference image of Figure (a) in the db97 wavelet domain. Figure (c) and (d) indicate the labeled area(s) (the white areas) where is(are) classified to have the wavelet secondary property damaged. The identification is to compare the local energy set correlation ρ and the threshold T_c. In Figure (a), the first scale is circularly shifted to the left, so the wavelet secondary property damaged within the entire scale. In Figure (b), no distortion occurs. Ideally, Figure (c) should be all white and Figure (d) should be all dark.

reference image (the original image) is estimated by Equation 3.29.

$$\log \hat{\mathcal{E}}_i(m_i, n_i) = \begin{bmatrix} \log \hat{\mathcal{E}}_{i+1}(m_{i+1}, n_{i+1}) & \log \hat{\mathcal{E}}_{i+2}(m_{i+2}, n_{i+2}) & \dots \end{bmatrix} \begin{bmatrix} \underline{\hat{\alpha_i}}_1 \\ \underline{\hat{\alpha_i}}_2 \\ \vdots \end{bmatrix} + \hat{\beta}_i \quad (3.29)$$

The local energy consistency model can be applied to estimate the white noise variance and blur distortion in the high frequency as with the energy consistency model. The white noise variance $\sigma^2_{D_{m_i,n_i}}$ of the neighborhood centered at (m_i, n_i) in the ith scale can be estimated in Equation 3.30, where $\mathcal{E}_d(m_i, n_i)$ is the local energy of the noisy image.

$$\hat{\sigma^2}_{D_{m_i,n_i}} = \begin{cases} \mathcal{E}_d(m_i, n_i) - e^{\log \hat{\mathcal{E}}(m_i,n_i)} & \text{if} \quad \mathcal{E}_d(m_i, n_i) > e^{\log \hat{\mathcal{E}}(m_i,n_i)} \\ 0 & \text{else} \end{cases} \quad (3.30)$$

The blur distortion variance in each neighborhood can be estimated in Equation 3.32. The Gaussian blur can be modeled in the wavelet domain by Equation 3.31, where X_{m_i,n_i} is the original wavelet coefficient of the neighborhood in the ith scale and $X_{d_{m_i,n_i}}$ is the wavelet coefficient of the distorted (blurred) image. The parameter k_{m_i,n_i} controls the degree of blur.

$$X_{d_{m_i,n_i}} = k_{m_i,n_i} X_{m_i,n_i} \quad (3.31)$$

$$\hat{\sigma^2_{D_{m_i,n_i}}} = (1 - \hat{k}_{m_i,n_i})^2 \hat{\mathcal{E}}(m_i, n_i) \quad (3.32)$$

$$\hat{k^2} = \frac{\mathcal{E}_d(m_i, n_i)}{\hat{\mathcal{E}}(m_i, n_i)} \quad (3.33)$$

3.4 Summary

This chapter first reviews the background of the wavelet. Later, we reviewed the most successful statistical models: the hidden Markov tree(HMT) model and the Gaussian scale mixture(GSM) model. The HMT model emphasizes the inter-scale dependency, while the GSM model describes the intra-scale statistics more precisely. In Chapter 4, these three models are applied to derive the information that can be perceived by human observers. In Chapter 5, we are interested which model is more suitable for quality assessment. In the last part of Chapter 3, we introduce a linear relationship of the energy across scales, referred to as the energy consistency model and the local energy consistency model. These two energy models are the foundation to address

blind quality assessment in Chapter 6.

Chapter 4

Information Theoretic Framework

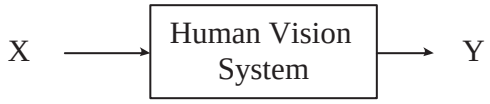
In this chapter we propose a measure of the amount of information that can perceived by humans. This chapter first reviews basic information concepts. Then the relationship between the perceived visual information and the actual image information is explained in terms of human contrast sensitivity.

4.1 Communication Channel and Perceptual Information

A communication channel refers to the medium used to convey information from a sender (or transmitter) to a receiver. We can imagine a communication channel as a water pipe. The maximum amount of water which can go through the pipe over a unit time is limited by the physical characters of the pipe and outlet, such as the diameter and shape. The other factor that influences the water flux is the pressure at the entrance of the pipe. Reliably transmitted information to the receiver through a communication channel is dominated by the channel physical characters (i.e. heating noise) and the amount of information at the sender. A communication channel could be made up of different physical media (air, water or circuits) and can be modeled physically by trying to calculate the physical processes which modify the transmitted signal. For example, Gaussian noise communication channel has the white noise added in to simulate the external interference or electronic noise during transmission.

Generally, a communication channel scan be considered as a system in Figure 4.1(a), with the transmitted signal X as the input and the received signal Y as the output. This generalized communication channel could model any pre-processing, post-

(a) A typical communication channel



(b) Human visual system model

Figure 4.1: Typical communication channel and human visual system model

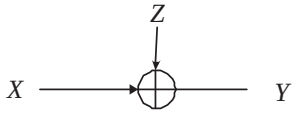


Figure 4.2: The additive Gaussian communication channel

processing and communication steps. The human vision system with complicated biological features can be described as a communication channel, with images as the input and visual signals in the brain as the output, as shown in Figure 4.1(b).

The information theory of C. E. Shannon [93] defines reliably transmitted information as the mutual information between the transmitted X and the received Y through a communication channel. Basically the mutual information describes the common information between X and Y.

Definition 4.1.1 *The information of X that can be perceived by human beings is defined as* ***perceptual information****. Mathematically, it is the mutual information between the transmitted X and the received Y through a communication channel which simulates the human vision behavior(s).*

To prepare the future discussion on perceptual information, let us review some common communication channels and the corresponding mutual information.

Definition 4.1.2 *The* ***additive Gaussian channel*** *[28]. $Y = X + Z$, where $Z \sim g(0, \sigma_Z^2)$ and independent of X.*

If the input signal $X \sim g(0, \sigma_X^2)$, then the reliably transmitted information that can be delivered by this Gaussian channel is [28]

$$I(X;Y) = \frac{1}{2}\log(\frac{\sigma_X^2 + \sigma_Z^2}{\sigma_Z^2}) \tag{4.1}$$

Definition 4.1.3 *The **parallel Gaussian channel** [28]. $\underline{Y} = \underline{X} + \underline{Z}$, where $\underline{Z} \sim g(\mathbf{0}, \mathbf{\Sigma}_{\underline{Z}})$ and independent of $\underline{X}$.*

If the input $\underline{X} \sim g(\mathbf{0}, \mathbf{\Sigma}_{\underline{X}})$, then the reliably transmitted information that can be delivered by this parallel Gaussian channel is [28]

$$I(\underline{X};\underline{Y}) = \frac{1}{2}\log|\mathbf{\Sigma}_{\underline{Y}}| - \frac{1}{2}\log|\mathbf{\Sigma}_{\underline{Z}}| \tag{4.2}$$

where $\mathbf{\Sigma}_{\underline{Y}} = \mathbf{\Sigma}_{\underline{X}} + \mathbf{\Sigma}_{\underline{Z}}$.

A generalized communication channel may describe any operation, processing or system. Let us define the quantization process as a communication channel.

Definition 4.1.4 *The **quantization channel** is defined as a communication channel which quantizes the input X to the discrete output Y as follows.*

$$Y = \lfloor \frac{X}{\triangle} \rfloor \tag{4.3}$$

where $\triangle$ is the constant quantization step.

The reliably transmitted information delivered by this quantization channel is can be computed as follows.

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) && (4.4)\\ &= h(Y) && (4.5)\\ &= h(\lfloor \frac{X}{\triangle} \rfloor) && (4.6) \end{aligned}$$

Equation 4.4 is the definition of mutual information [28]. The conditional entropy $h(Y|X)$ quantifies the remained uncertainty of Y if X is known. In Equation 4.3, Y can be determined by X and $\triangle$. Hence there is no uncertainty remained in Y if X and $\triangle$ are given, which means $h(Y|X) = 0$. $h(\lfloor \frac{X}{\triangle} \rfloor)$ has no closed form expression, but it can be approximately estimated by $h(\lfloor \frac{X}{\triangle} \rfloor)$ as shown in Equation 4.7 [28].

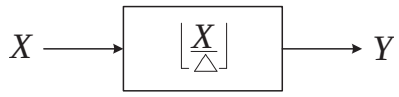


Figure 4.3: The quantization channel

$$h(Y) = h(\lfloor \frac{X}{\triangle} \rfloor) \simeq \begin{cases} h(X) - \log(\triangle) & if \ \ h(x) > \log(\triangle) \\ 0 & else \end{cases} \tag{4.7}$$

If $X \sim \mathcal{N}(0, \sigma_X^2)$, $h(X) = \frac{1}{2}\log(2\pi e \sigma_X^2)$ [28]. The approximation error can be found in Appendix B.

4.2 The Contrast Sensitivity Model

In Section 4.1, we explain that the human vision system may be described by a communication channel. Constructing such a communication channel is essential to derive the perceptual information and information theoretic framework for quality assessment.

The current consensus seems to be that contrast sensitivity is a primary component when people detect the difference between images. Over 60% of the quality metrics reviewed in Chapter 2 adopted the contrast sensitivity in different forms. Our solution also adopts this vision characteristic as the key component for quality assessment, and the test results in Chapter 5 shows that the quality assessment results have excellent consistency with human judgments.

In this section, we will introduce two models to simulate contrast sensitivity. The first model is intuitive. The second model is not as straightforward as the first one. We will explain the equivalency between the two models. Furthermore, the perceptual information is simplified by using the second model, which makes analysis easier.

A representative model for contrast sensitivity is the quantization with different quantization steps at different frequencies. This model is used by some quality metrics [17, 130] and most compression algorithms [1, 3]. In our solution, the contrast sensitivity at a certain frequency is modeled by the quantization channel with the quantization step "$\triangle$" indicating human visual sensitivity at that frequency. Hence the perceptual in-

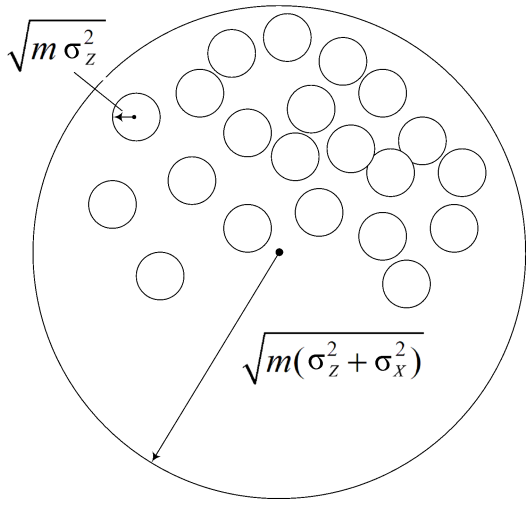


Figure 4.4: Sphere decoding for Gaussian channel

formation of "X" at a frequency defined by the quantization channel is:

$$I(X;Y) \simeq \begin{cases} h(X) - \log(\triangle) & if\ \ h(X) > \log(\triangle) \\ 0 & else \end{cases} \tag{4.8}$$

In Equation 4.8, if the quantization step $\triangle$ gets larger, the perceptual information gets less. It means the human observer is less sensitive to the information at this frequency and results in the less information perceived.

The second solution to model contrast sensitivity is the Gaussian additive channel. In the following section, we will explain that the second solution is equivalent to the quantization channel. Consider a Gaussian signal X sent through a Gaussian channel. The received signal $Y = X + Z$, where Z is the additive Gaussian noise and independent of X. If a codeword $\underline{x}$ ($[x_1, x_2, x_3, ..., x_n]$) of length n is sent through the channel, where each element x_i is generated by X independent of x_j for $i \neq j$, the perturbed codeword $\underline{y}$ ($[y_1, y_2, y_3, ..., y_n]$) is received. The probability of the distance between two codewords is

$$p(||\underline{x} - \underline{y}||) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_Z} \exp(-\frac{|x_i - y_i|^2}{2\sigma_Z^2}) \tag{4.9}$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_Z} \exp(-\frac{z_i^2}{2\sigma^2}) \tag{4.10}$$

$$= \frac{1}{(\sqrt{2\pi}\sigma_Z)^n} \exp(-\frac{z_1^2 + z_2^2 + z_3^2 + ... + z_n^2}{2\sigma^2}) \quad (4.11)$$

$$= \frac{1}{(\sqrt{2\pi}\sigma_Z)^n} \exp(-\frac{z_1^2 + z_2^2 + z_3^2 + ... + z_n^2}{n}\frac{n}{2\sigma^2}) \quad (4.12)$$

Please note that z_i is generated by the Gaussian noise Z. As the codeword length n grows, the summation part inside exp will converge to σ_Z^2.

$$\lim_{n\to+\infty} \frac{z_1^2 + z_2^2 + z_3^2 + ... + z_n^2}{n} = \sigma_Z^2 \quad (4.13)$$

Geometrically, Equation 4.13 shows that the perturbation will almost certainly be at some point near the surface of a n dimensional sphere of radius $\sqrt{n\sigma_Z^2}$ centered at the original codeword $\underline{x}$ if n is large. More precisely, by taking n sufficiently large, the perturbation will lie within a sphere of radius $\sqrt{n(\sigma_Z^2 + \epsilon)}$ where ϵ is arbitrarily small, and the received codeword $\underline{y}$ will lie within a sphere of radius $\sqrt{n(\sigma_Z^2 + \sigma_X^2 + \epsilon)}$. If the n dimensional sphere of radius $\sqrt{n(\sigma_Z^2 + \sigma_X^2 + \epsilon)}$ is divided into M spheres of radius $\sqrt{n(\sigma_Z^2 + \epsilon)}$ which are not overlapped, as shown in Figure 4.2, the "good" codewords $\underline{x}$ can be selected as the centers of the above M spheres, and the received $\underline{y}$ can be decoded according to which sphere it lies in. All the points within a sphere will be mapped to one point. This basic idea is the same as the quantization, which maps the points within $[k\triangle - \triangle/2, k\triangle + \triangle/2]$ to $k\triangle$ (k is an integer). The maximum number of non-intersecting decoding spheres in the sphere of $\underline{y}$ is no more than:

$$\frac{(n\sigma_Y^2)^{\frac{n}{2}}}{(n\sigma_Z^2)^{\frac{n}{2}}} = \frac{[n(\sigma_X^2 + \sigma_Z^2)]^{\frac{n}{2}}}{(n\sigma_Z^2)^{\frac{n}{2}}} = 2^{\frac{n}{2}log_2(\frac{\sigma_X^2+\sigma_Z^2}{\sigma_Z^2})} \quad (4.14)$$

The right side of Equation 4.14 is the maximum of possible codewords. The reliable transmitted information is $\frac{n}{2}log_2(\frac{\sigma_X^2+\sigma_Z^2}{\sigma_Z^2})$ for each codeword with length of n. On average, each bit of a codeword contains entropy of $\frac{1}{2}log_2(\frac{\sigma_X^2+\sigma_Z^2}{\sigma_Z^2})$, which is the mutual information in Equation 4.2. This sphere pack method was proposed by Shannon [94] to accomplish reliable communication in the presence of noise. We borrow his idea to explain that the communication through the Gaussian channel is equivalent to a quantization procedure. Hence the perceptual information of "X" at a frequency defined by the Gaussian channel is shown in Equation 4.15, where the Gaussian noise Z regulates vision sensitivity at that frequency with σ_Z^2.

$$I(X;Y) = \frac{1}{2}\log(\frac{\sigma_X^2 + \sigma_Z^2}{\sigma_Z^2}) \quad (4.15)$$

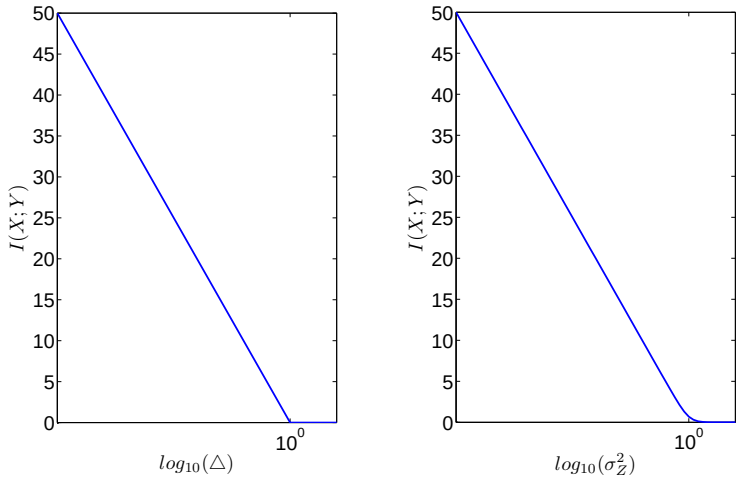


Figure 4.5: Different definitions of the perceptual information. X obeys the standard normal distribution. The perceptual information $I(X;Y)$ defined by quantization channel is plotted in Figure (a). The perceptual information defined by the additive Gaussian channel is plotted in Figure (b). It is clear that two definitions produce similar perceptual information. When the perceptual information is close to zero, the definition by Gaussian channel acts like a soft threshold while the definition by quantization acts like a hard threshold.

Figure 4.5 shows the perceptual information defined in two different ways: the quantization channel and the additive Gaussian channel. The source signal X obeys the standard normal distribution. The perceptual information $I(X;Y)$ defined by the additive Gaussian channel is plotted with different amounts of perception noise Z. The perceptual information defined by the quantization channel is plotted in 4.5(b). It is clear that two definitions output similar perceptual information. They differ when the perceptual information is close to zero where the definition by the Gaussian channel acts like a soft threshold while the definition by a quantization channel acts like a hard threshold. Another difference between the two definitions is that the perceptual information defined by the Gaussian channel has a simpler expression. The perceptual information definition by the quantization channel in Equation 4.8 is a piecewise function. For high dimensional signals, the computation of the perceptual information by a quantization channel will be much more complicated [142]. Furthermore such a definition (by a quantization channel) will make the perceptual information expression extremely complicated [142] when the non-Gaussian image statistics is described by the

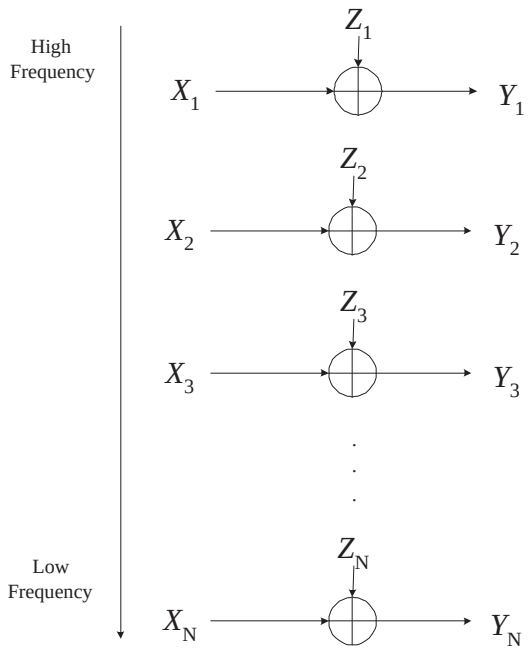


Figure 4.6: The contrast sensitivity model. Each Gaussian channel interprets the image on certain frequency and the noise level (σ_{Z_i}) indicates the contrast sensitivity on that frequency.

HMT [29] or GSM [84] model. Hence the perceptual information defined by the additive Gaussian channel is preferable and used in the rest part of the thesis.

Considering that the human eyes exhibit different contrast sensitivities to images at different frequencies [20], the contrast sensitivity model is defined as a parallel Gaussian noise channel as shown in Figure 4.6(a), which is first proposed by Sheikh and Bovick [97] [1]. Each channel interprets the image on certain scale and the noise level indicates the contrast sensitivity on that frequency (scale). Consider an image in the frequency (wavelet) domain, let the random variable X_i represent the image in the ith frequency (scale). Mathematically, the contrast sensitivity model can be written in Equation 4.16, where the noise Z_i is Gaussian and independent of X_j for any i and j.

$$Y_i = X_i + Z_i \quad \forall i \tag{4.16}$$

[1] Ruderman [87] first described the information in the retina using a Gaussian noise channel. Sheikh and Bovick [97] improved the model by expanding the single Gaussian channel to the parallel channel.

According to the perceptual information definition, the perceptual information of an image is the mutual information between this image and the received image through the parallel Gaussian channel. The perceptual information of an image at a certain frequency is the mutual information between this image at that frequency and the received image through the single Gaussian channel, i.e. $I(X_i; Y_i)$.

4.2.1 Perceptual Information

Perceptual information depends not only on the contrast sensitivity model but also on how image statistics is modeled. In Chapter 3 Section 2, three statistical models are reviewed. Table 4.1 briefly illustrates the differences of the perceptual information when different statistical models are applied. In this section, we will derive the expression of the perceptual information using different statistical models.

Consider an image in the wavelet domain and let the random variable X_i represent the image in the ith scale. The random vector $\underline{X}$ is composed of the image X_i in each scale. Then the perceptual image of Y_i in the ith scale is computed in Equation 4.17, where the noise Z_i is Gaussian and independent of X_i.

$$Y_i = X_i + Z_i \quad \forall i \tag{4.17}$$

If the statistical model (the simple Gaussian model or the HMT model [29]) describes the marginal statistics of the wavelet coefficients in a scale, Equation 4.17 is suitable to compute the perceptual information. If the statistical model (the GSM model [84]) describes the intra-scale joint statistics of the wavelet coefficients, Equation 4.17 needs a modification. The basic idea of the contrast sensitivity model is the Gaussian additive channel. Mathematically, each coefficient in the ith scale is independently added with the Gaussian noise, which obeys $g(0, \sigma^2_{Z_i})$. Let $\underline{X_i}$ denote the random vector of the ith scale. In the GSM model, $\underline{X_i}$ could be formed by a 3×3 window. The corresponding contrast sensitivity model of the ith scale can be written as follows.

$$\underline{Y_i} = \underline{X_i} + \underline{Z_i} \tag{4.18}$$

	Statistics in the *ith* scale	Perceptual information of the *ith* scale
The simple Gaussian Model	marginal statistics, Gaussianity	the mutual information between two random **variables**
The HMT model	marginal statistics, non-Gaussianity	the mutual information between two random **variables**
The GSM model	joint statistics, non-Gaussianity	the mutual information between two random **vectors**

Table 4.1: Perceptual information using different statistical model in the *ith* scale

where $\underline{Z_i}$ is joint Gaussian and $\mathbf{\Sigma}_{\underline{Z_i}} = \begin{bmatrix} \sigma^2_{Z_i} & 0 & ... & 0 \\ 0 & \sigma^2_{Z_i} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \sigma^2_{Z_i} \end{bmatrix}$

1. ***The perceptual information using the simple Gaussian model***

The simple Gaussian model assumes the image X_i in the *ith* scale to be Gaussian and independent of other scales. Since Z_i is independent of Z_j if $i \neq j$, Y_i is independent of Y_j. Hence the perceptual information of $\underline{X}$ across all scales can be computed as:

$$I(\underline{X};\underline{Y}) = \sum_i I(X_i;Y_i) \tag{4.19}$$

$$= \sum_i I(X_i;X_i+Z_i) \tag{4.20}$$

$$= \frac{1}{2}\sum_i \log(1+\frac{\sigma^2_{X_i}}{\sigma^2_{Z_i}}) \tag{4.21}$$

Equation 4.19 comes from the independency assumption of each scale. Equation 4.21 comes from Equation 4.2 in the first section of this chapter. The perceptual informa-

tion of the image X_i in the ith scale can be computed as follows:

$$I(X_i; Y_i) = \frac{1}{2}\log(1 + \frac{\sigma^2_{X_i}}{\sigma^2_{Z_i}}) \tag{4.22}$$

2. ***The perceptual information using the HMT model***

From Chapter 3 Section 2, the HMT model [29] describes the non-Gaussian marginal statistics of image at each frequency (scale). Also the HMT model describes the joint statistics across all frequencies (scales). The probability density of the image $\underline{X}$ across all frequencies (scales) is depicted as the mixture of the joint Gaussian distributions in Equation 4.23, where $\underline{X}|\underline{\theta} \sim g(\mathbf{0}, \mathbf{\Sigma}_{\underline{X}|\underline{\theta}})$ and $\underline{\theta}$ is the vector formed by the hidden state θ_i in each scale.

$$f(\underline{X}) = \sum_{\underline{\theta}} f(\underline{X}|\underline{\theta}) f(\underline{\theta}) \tag{4.23}$$

To derive the perceptual information of $\underline{X}$ under the HMT model, it is necessary to derive the probability density of $\underline{Y} = \underline{X} + \underline{Z}$. Since $\underline{Z}$ is Gaussian and independent of $\underline{X}$, $\underline{Y}|\underline{\theta} = \underline{X}|\underline{\theta} + \underline{Z}$ and $\underline{Y}|\underline{\theta} \sim g(\mathbf{0}, \mathbf{\Sigma}_{\underline{X}|\underline{\theta}} + \mathbf{\Sigma}_{\underline{Z}})$, the probability density of $\underline{Y}$ is the mixture of the joint Gaussian distributions. The mutual information between $\underline{X}$ and $\underline{Y}$ is derived as follows. Equation 4.24 comes from the definition of the mutual information [28]. There is no closed form for the entropy of the Gaussian mixture random vector. Huber *et al.* [57] proposed the approximation for $h(\underline{Y})$ as the conditional entropy $h(\underline{Y}|\underline{\theta})$ in Equations 4.26 and 4.27. Hence the perceptual information $I(\underline{X}; \underline{Y})$ of $\underline{X}$ can be computed in Equation 4.28, where N is the length of the vector $\underline{X}$.

$$\begin{aligned}
I(\underline{X}; \underline{Y}) &= h(\underline{Y}) - h(\underline{Y}|\underline{X}) && (4.24)\\
&= h(\underline{Y}) - h(\underline{Z}) && (4.25)\\
&\simeq h(\underline{Y}|\underline{\theta}) - h(\underline{Z}) && (4.26)\\
&= \frac{1}{2}\sum_{\underline{\theta}} p(\underline{\theta}) \log(2\pi e)^N |\mathbf{\Sigma}_{\underline{Y}|\underline{\theta}}| - \frac{1}{2}\log(2\pi e)^N |\mathbf{\Sigma}_{\underline{Z}}| && (4.27)\\
&= \frac{1}{2}\sum_{\underline{\theta}} p(\underline{\theta}) \log |\mathbf{\Sigma}_{\underline{Y}|\underline{\theta}}| - \frac{1}{2}\log |\mathbf{\Sigma}_{\underline{Z}}| && (4.28)
\end{aligned}$$

where $\mathbf{\Sigma}_{\underline{Z}} = \begin{bmatrix} \sigma^2_{Z_1} & 0 & ... & 0 \\ 0 & \sigma^2_{Z_2} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \sigma^2_{Z_N} \end{bmatrix}$

The HMT model describes the marginal intra-scale statistics with the Gaussian mixture model: $f(X_i) = p(\theta_i = s)g(X_i; 0, \sigma^2_{s;i}) + p(\theta_i = \mathcal{L})g(X_i; 0, \sigma^2_{\mathcal{L};X_i})$. Since X_i is independent of Z_i, the probability density of Y_i can be written as: $f(Y_i) = p(\theta_i = s)g(Y_i; 0, \sigma^2_{s;i} + \sigma^2_{Z_i}) + p(\theta_i = \mathcal{L})g(Y_i; 0, \sigma^2_{\mathcal{L};i} + \sigma^2_{Z_i})$. The computation of $I(X_i; Y_i)$ is very similar to $I(\underline{X}; \underline{Y})$ and is shown as follows. $h(Y_i)$ is approximated by the conditional entropy $h(Y_i|\theta_i)$ suggested by Huber *et al.* [57].

$$\begin{aligned} I(X_i; Y_i) &= h(Y_i) - h(Y_i|X_i) && (4.29) \\ &= h(Y_i) - h(Z_i) && (4.30) \\ &\simeq h(Y_i|\theta_i) - h(Z_i) && (4.31) \\ &= \frac{1}{2}\sum_{\theta_i} p(\theta_i) \log \sigma^2_{Y_i|\theta_i} - \frac{1}{2} \log \sigma^2_{Z_i} && (4.32) \end{aligned}$$

3. ***The perceptual information using the GSM model***

The GSM model [84] is different from the above two statistical models. The simple Gaussian model and the HMT model describe the marginal statistics in each subband, while the GSM model describes the intra-scale joint statistics in the *ith* scale as shown in Equation 4.33, where $\underline{X_i}$ is the random vector formed by the coefficients in the defined neighborhood (such as a 3×3 window) in the *ith* scale, where s_i is the hidden state and a positive discrete random scalar.

$$f(\underline{X_i}) = \sum_{s_i} p(s_i) g(\underline{X_i}; \mathbf{0}, s_i \mathbf{\Sigma}_{\underline{X_i}}) \quad (4.33)$$

The perceptual information in the *ith* scale using the GSM model is the mutual information between two random vectors $\underline{X_i}$ and $\underline{Y_i}$. The computation of $I(\underline{X_i}; \underline{Y_i})$ is similar to the computation of the perceptual information using the HMT model. $\mathbf{\Sigma}_{\underline{Y_i}|s_i}$ $= \mathbf{\Sigma}_{\underline{X_i}|s_i} + \mathbf{\Sigma}_{\underline{Z_i}}$, since Z_i is independent of the image coefficients.

$$\begin{aligned} I(\underline{X_i}; \underline{Y_i}) &= h(\underline{Y_i}) - h(\underline{Y_i}|\underline{X_i}) && (4.34) \\ &= h(\underline{Y_i}) - h(\underline{Z_i}) && (4.35) \\ &\simeq h(\underline{Y_i}|s_i) - h(\underline{Z_i}) && (4.36) \end{aligned}$$

$$= \frac{1}{2}\sum_{s_i} p(s_i)\log|\mathbf{\Sigma}_{\underline{Y_i}|s_i}| - \frac{1}{2}\log|\mathbf{\Sigma}_{\underline{Z_i}}| \tag{4.37}$$

where $\mathbf{\Sigma}_{\underline{Z_i}} = \begin{bmatrix} \sigma^2_{Z_i} & 0 & ... & 0 \\ 0 & \sigma^2_{Z_i} & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \sigma^2_{Z_i} \end{bmatrix}$

Under the GSM model, each scale is assumed to be independent of other scales. The total perceptual information across all scales is simply the summation of the information in each scale, as shown in Equation 4.38.

$$I(\underline{X_1}, \underline{X_2}, ..., \underline{X_N}; \underline{Y_1}, \underline{Y_2}, ..., \underline{Y_N}) = \sum_i I(\underline{X_i}; \underline{Y_i}) \tag{4.38}$$

As shown above, the expression of the perceptual information greatly depends on the statistical models. To avoid possible misunderstanding, let us summarize the perceptual information in Table 4.2. Also it is necessary to assign an uniform symbol for the perceptual information, to help clarity the basic idea of our quality assessment method in the next chapter. Let $\mathcal{P}(X_i)$ denote the perceptual information in the *ith* scale and let $\mathcal{P}(\underline{X})$ denote the perceptual information across all scales, regardless which statistical model is applied.

	The perceptual information of the *ith* scale $\mathcal{P}(X_i)$	The total perceptual information across all scales $\mathcal{P}(\underline{X})$
The simple Gaussian model	$\frac{1}{2}\log(1+\frac{\sigma^2_{X_i}}{\sigma^2_{Z_i}})$	$\frac{1}{2}\sum_i \log(1+\frac{\sigma^2_{X_i}}{\sigma^2_{Z_i}})$
The HMT model [29]	$\frac{1}{2}\sum_{\theta_i} p(\theta_i)\log\sigma^2_{Y_i\|\theta_i} - \frac{1}{2}\log\sigma^2_{Z_i}$	$\frac{1}{2}\sum_{\underline{\theta}} p(\underline{\theta})\log\|\mathbf{\Sigma}_{\underline{Y}\|\underline{\theta}}\| - \frac{1}{2}\log\|\mathbf{\Sigma}_{\underline{Z}}\|$
The GSM model [84]	$\frac{1}{2}\sum_{s_i} p(s_i)\log\|\mathbf{\Sigma}_{\underline{Y_i}\|s_i}\| - \frac{1}{2}\log\|\mathbf{\Sigma}_{\underline{Z_i}}\|$	$\frac{1}{2}\sum_i \left(\sum_{s_i} p(s_i)\log\|\mathbf{\Sigma}_{\underline{Y_i}\|s_i}\| - \log\|\mathbf{\Sigma}_{\underline{Z_i}}\|\right)$

Table 4.2: The perceptual information expression using different statistical models

4.3 Summary

Chapter 4 first reviews fundamental information theory, which provides the necessary explanation about the basic concept of entropy, conditional entropy and mutual information. In the second part of this chapter, we propose two solutions to model contrast sensitivity behavior. The first model simulates the contrast sensitivity by a quantization channel, whereas the second model adopts a Gaussian channel. The test results show that two models have very similar responses, but the latter model has more advantages over the theoretical analytics. Finally, the perceptual information is derived if different statistical models are applied. In the next chapter, we will test which statistical model will be more suitable for quality assessment.

Chapter 5

Full-reference Quality Assessment

In this chapter, we introduce a quality metric for images with full reference access. we also discuss several issues which reveal the key points for a successful image quality assessment method. In the last part of this chapter, our method is tested by five subjective databases from different research groups and compared with other existing quality metrics.

5.1 Image Quality Assessment

Image quality assessment based on information theory can be traced back to 1994. Ruderman [87] described the information in the retina. At a specific frequency, Ruderman modeled the response of the human vision system as an additive noise channel. Later, Sheikh *et al.* [97] enhanced this framework by adopting a complicated statistical model, the Gaussian scale mixture model [84].

In Chapter 4, we proposed an information theoretic framework to describe the perceptual information of an image. This idea leads to a simple solution for full reference quality assessment. Basically, the quality assessment is a comparison between the test image and its reference, and the visual quality of the test image is defined as the information of the image differences (between the test image and its reference) that can be perceived by human viewers.

$\underline{X}$ and $\underline{X}_d$ are the test image and its reference in the wavelet domain with each element for each subband. Let $\underline{D}$ denote the image difference in the wavelet domain, as computed in Equation 5.1. The visual quality can be simply defined as the perceptual

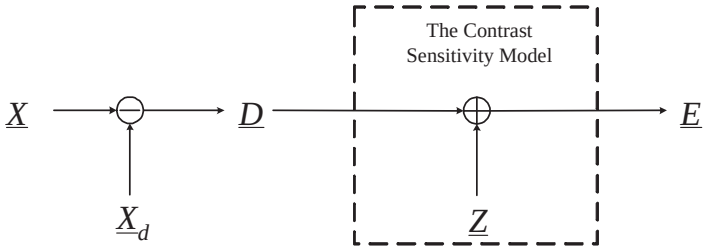


Figure 5.1: The simple quality assessment criterion

difference $\mathcal{P}(\underline{D})$ in Equation 5.3.

$$\underline{D} = \underline{X} - \underline{X}_d \tag{5.1}$$

$$\underline{E} = \underline{D} + \underline{Z} \tag{5.2}$$

$$PD = \mathcal{P}(\underline{D}) \tag{5.3}$$

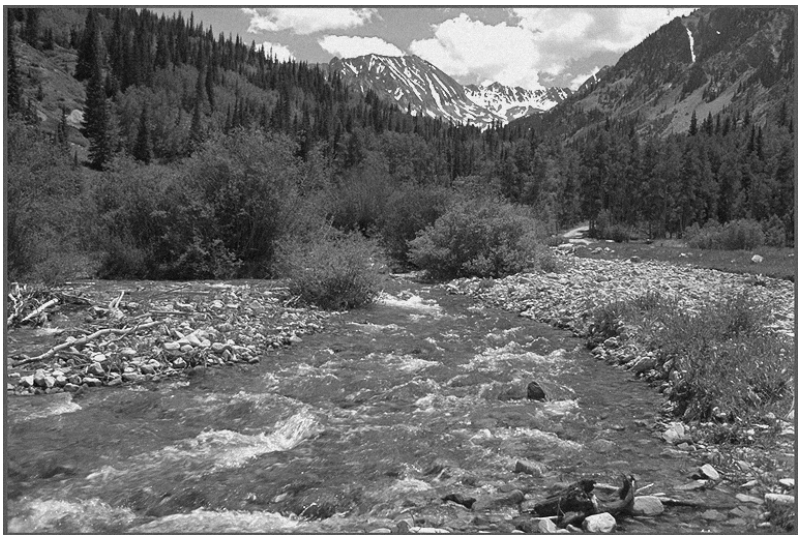
From the PD definition, PD will output the same quality score for the same pixel intensity change. But the same pixel intensity change may attract different visual attention and result in different quality judgments. As shown in Figure 5.2, the same amount of white noise was added to two images. The peak noise signal ratio is 30.45dB. The noise in Figure 5.2(a) is easier to notice and people are more likely to assign a bad quality score for Figure 5.2(a). But PD provides the same quality score for Figures 5.2(a) and 5.2(b).

The example in Figure 5.2 illustrates that human eyes are less likely to detect the distortions in texture-rich areas (Figure 5.2(b)). This means that the visual quality assessment is dependent on image content. Our second solution to quality assessment is defined as the perceptual difference normalized by the perceptual information of the reference image in Equation 5.4.

$$NPD = \frac{\mathcal{P}(\underline{D})}{\mathcal{P}(\underline{X})} \tag{5.4}$$

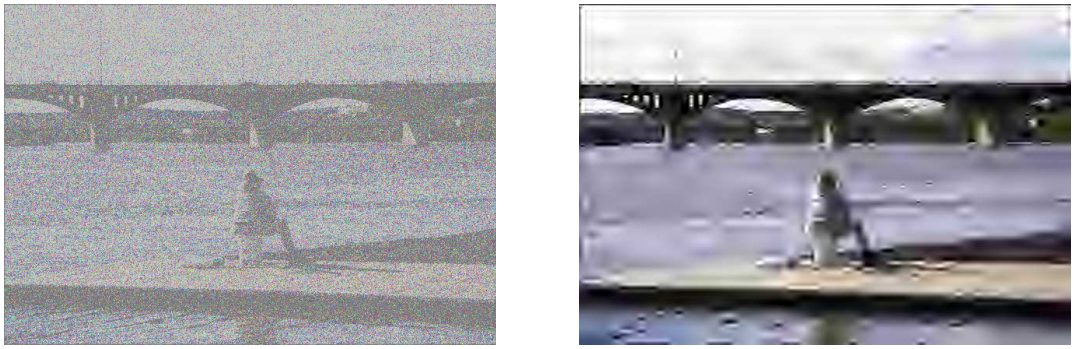
Since texture-rich images have greater uncertainty, they have more perceptual information compared with the texture-poor images. According to visual quality defined by the normalized perceptual difference, the same distortion will cause less visual degradation for texture-rich images. Using NPD, the visual quality scores for Figures 5.2(a) and 5.2(b) are 0.4521 and 0.3212 respectively. The smaller score means better visual quality. Figures 5.4(a) and 5.4(b) display the performance of PD and NPD on the entire LIVE database [99]. It is clear that NPD has much better consistency with human

(a)



(b)

Figure 5.2: The different visual quality degradations caused by the same noise. The same white noise was added to the original images in Figure (a) and (b). The peak noise signal ratio is 30.45dB. The noise in Figure (a) is easier to notice and people are more likely to assign a bad quality score for it. PD provides the same quality score for both images, since PD only considers visual quality as the image difference that can be perceived. NPD evaluates the quality scores of Figure (a) and Figure (b) to be 0.4521 and 0.3212. The smaller score means the better visual quality.



(a) (b)

Figure 5.3: The shortcoming of NPD. Figure (a) contains white noise distortion while Figure (b) contains JPEG2000 distortion. Both images are provided by the LIVE database [99] with almost the same subjective visual quality score. NPD gives two different quality scores: 2.745 and 0.86 for Figure (a) and (b) respectively.

judgment. The Spearman correlation between the NPD evaluations and human judgments is 0.8876, which is much larger than the Spearman correlation (0.7743) between the PD evaluations and human judgments. But there is one problem with the second solution. Figure 5.3 displays two test images provided by the LIVE database [99]. Both images have almost the same subjective visual quality score. But NPD gives two different quality scores: 2.745 for Figure 5.3(a) and 0.86 for Figure 5.3(b). It means that NPD does not deal with different distortions on the same scale, especially when the distortion gets larger, as shown in Figure 5.4(b). The reason is that the compression distortion will eliminate the image high frequency components but white noise distortion will add the irrelevant information to the high frequency. When the distortion reaches a certain level, white noise distortion in the high frequency will exceed the original information. This causes NPD greater than 1. For compression distortion, the large distortion happens when the high frequency information is mostly eliminated and the corresponding NPD is close to 1 but less than 1.

To solve this problem, the normalization step in NPD is modified. The information theoretic criterion (ITC) for image quality is proposed in Equation 5.5, where the max operator returns the larger one of the perceptual information of the test image and its reference on the ith scale.

$$ITC = \frac{\mathcal{P}(\underline{D})}{\sum_i \max\{\mathcal{P}(X_i), \mathcal{P}(X_{d_i})\}} \tag{5.5}$$

Let us analyze the impact of the max operator on our quality assessment metric. For instance, the image is contaminated by white noise in the spatial domain. Under the orthogonal wavelet framework, the wavelet coefficient X_i in the ith scale is disturbed by the white noise D_i [70]. If the simple Gaussian model is applied to describe the image statistics, the perceptual information in ITC can be derived as follows. Since the noise D_i is independent of X_i, the test image X_{d_i} in the ith scale is Gaussian with variance $\sigma^2_{X_i} + \sigma^2_{D_i}$. Equation 5.6, 5.7 and 5.9 can be easily derived from the expression of the perceptual information in Table 4.2. The larger perceptual information between $\mathcal{P}(X_i)$ and $\mathcal{P}(X_{d_i})$ is $\mathcal{P}(X_{d_i})$. Also $\mathcal{P}(X_{d_i}) > \frac{1}{2}\log(1 + \sigma^2_{D_i}/\sigma^2_{Z_i})$. Then the summation of the perceptual information $\mathcal{P}(X_{d_i})$ across all scales will be larger than $P(\underline{D})$. Since $ITC = \mathcal{P}(\underline{D})/\sum_i \max\{\mathcal{P}(X_i), \mathcal{P}(X_{d_i})\}$, it is not difficult to see that ITC for white noise is limited to the range of 0 and 1.

$$\mathcal{P}(\underline{D}) = \frac{1}{2}\sum_i \log(1 + \frac{\sigma^2_{D_i}}{\sigma^2_{Z_i}}) \tag{5.6}$$

$$\mathcal{P}(X_i) = \frac{1}{2}\log(1 + \frac{\sigma^2_{X_i}}{\sigma^2_{Z_i}}) \tag{5.7}$$

$$\mathcal{P}(X_{d_i}) = \frac{1}{2}\log(1 + \frac{\sigma^2_{X_{d_i}}}{\sigma^2_{Z_i}}) \tag{5.8}$$

$$= \frac{1}{2}\log(1 + \frac{\sigma^2_{X_i} + \sigma^2_{D_i}}{\sigma^2_{Z_i}}) \tag{5.9}$$

From the ITC definition, a smaller ITC score means better quality. If ITC is 0, the visual quality is perfect. ITC is usually less than 1. The exception could happen when the distortion is extremely large. If $ITC > 1$, one possible reason is that $\mathcal{P}(D_i) > \max\{\mathcal{P}(X_i), \mathcal{P}(X_{d_i})\}$, which means that the distorted image may have no correlation with the reference image. In practice, the visual quality of such an image is usually not acceptable at all. From the test results in Figure 5.4(c), if the ITC score is greater than 1, the corresponding subjective quality score is larger than 60, which means "bad quality" [99].

Figure 5.4 displays the performance of PD, NPD and ITC. Three quality metrics evaluate the test samples with white noise or JPEG2000 distortion. When evaluating the images with large distortions, the test results show that NPD exhibits different behaviors for white noise and JPEG2000 distortions. This results in an inconsistency with subjective judgment. On the other hand, ITC can deal with different distortions on the same scale.

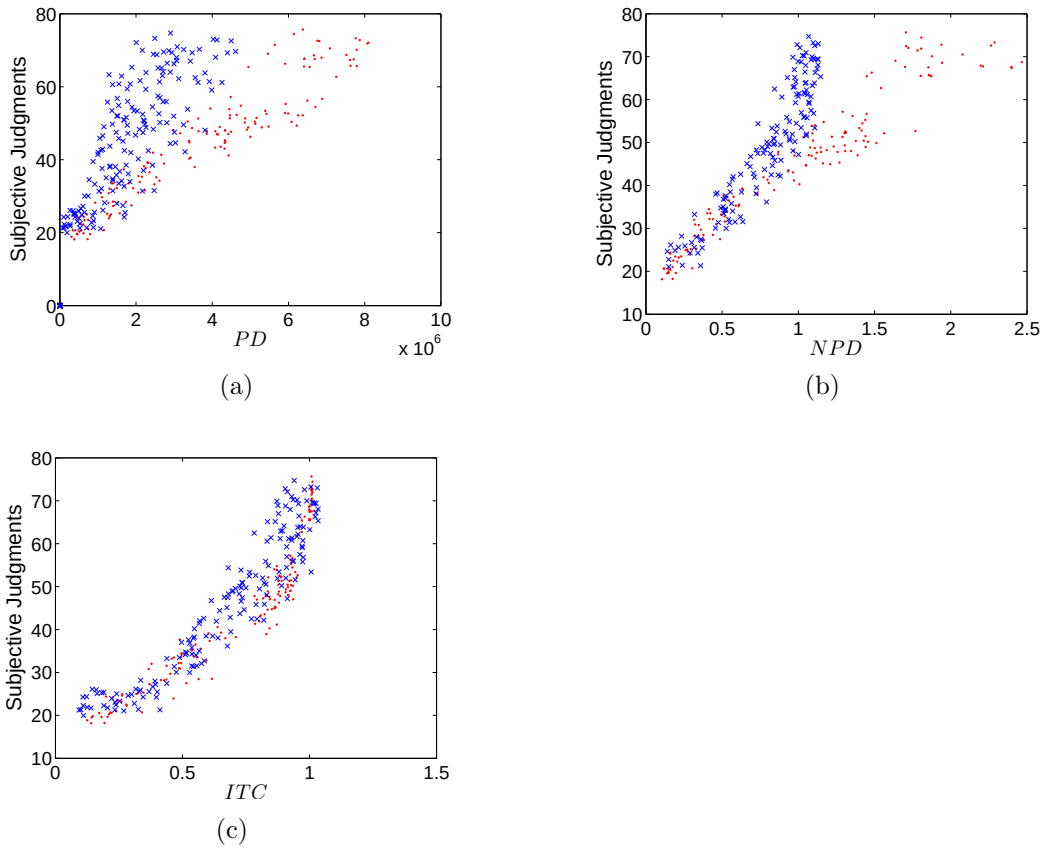


Figure 5.4: The performances of PD, NPD and ITC. Three quality metrics evaluate the test samples with white noise or JPEG2000 distortion provided by the LIVE database [99]. The test samples with white noise are plotted with dots and the test samples with the JPEG2000 distortion are plotted with crosses. The Spearman correlation between the objective evaluations and the subjective judgments are 0.7743, 0.8876 and 0.9168 for PD, NPD and ITC respectively.

Our information theoretic criterion can be summarized in four steps as follows.

1. Compute the test images $\underline{X}_d$ and its reference $\underline{X}$ in the wavelet domain.
2. Compute the image difference in the wavelet domain as $\underline{D} = \underline{X} - \underline{X}_d$.
3. Compute the perceptual information according to which statistical model is used. The perceptual information expression is listed in Table 4.2.
4. Compute ITC score for the test image according to Equation 5.5.

To implement the ITC, there are several issues to be determined. In the first step, images are transformed into the wavelet domain, but what kind of wavelet framework is suitable for quality assessment? In the third step, images in the wavelet domain can be described by the simple Gaussian model, HMT or GSM model. Which model should be used by ITC to provide the best quality evaluation? The contrast sensitivity model ($\sigma_{Z_i}^2$) controls the perceptual information in each frequency. The question is how to set the noise strength of each channel in the contrast sensitivity model. In the next three sections, we will focus on the above three questions.

5.2 Different Multi-resolution Frameworks

In this section, we are going to test the impact of different wavelet frameworks on the performance of ITC. The purpose of this test is to figure out the wavelet that is most suitable for ITC. In our experiment, ITC will be tested under 38 different wavelet frameworks, which covers the orthogonal wavelets and non-orthogonal wavelets. The complete wavelet list can be found in Appendix J.

Before the testing, a statistical model and a set of σ_{Z_i} for the contrast sensitivity model need to be determined to accomplish ITC. Since 38 wavelets will be tested with thousands of test images, ITC with less computation complexity is preferable. According to the ITC definition, the computation complexity largely depends on the perceptual information computation. The last section in Chapter 4 provides the expressions of the perceptual information using different statistical models in Table 4.2. The table implies that the simple Gaussian model needs the least computation. Hence in our experiment, we choose the simple Gaussian model[1]. The parameters $\sigma_{Z_i}^2$ in the

[1]Theoretically, all three statistical models should be tested under different wavelets. But the next section provides evidence that the advanced statistical models do not guarantee performance

contrast sensitivity model are selected[2] such that the ITC has the best performance, i.e. the largest Spearman correlation with the subjective quality scores.

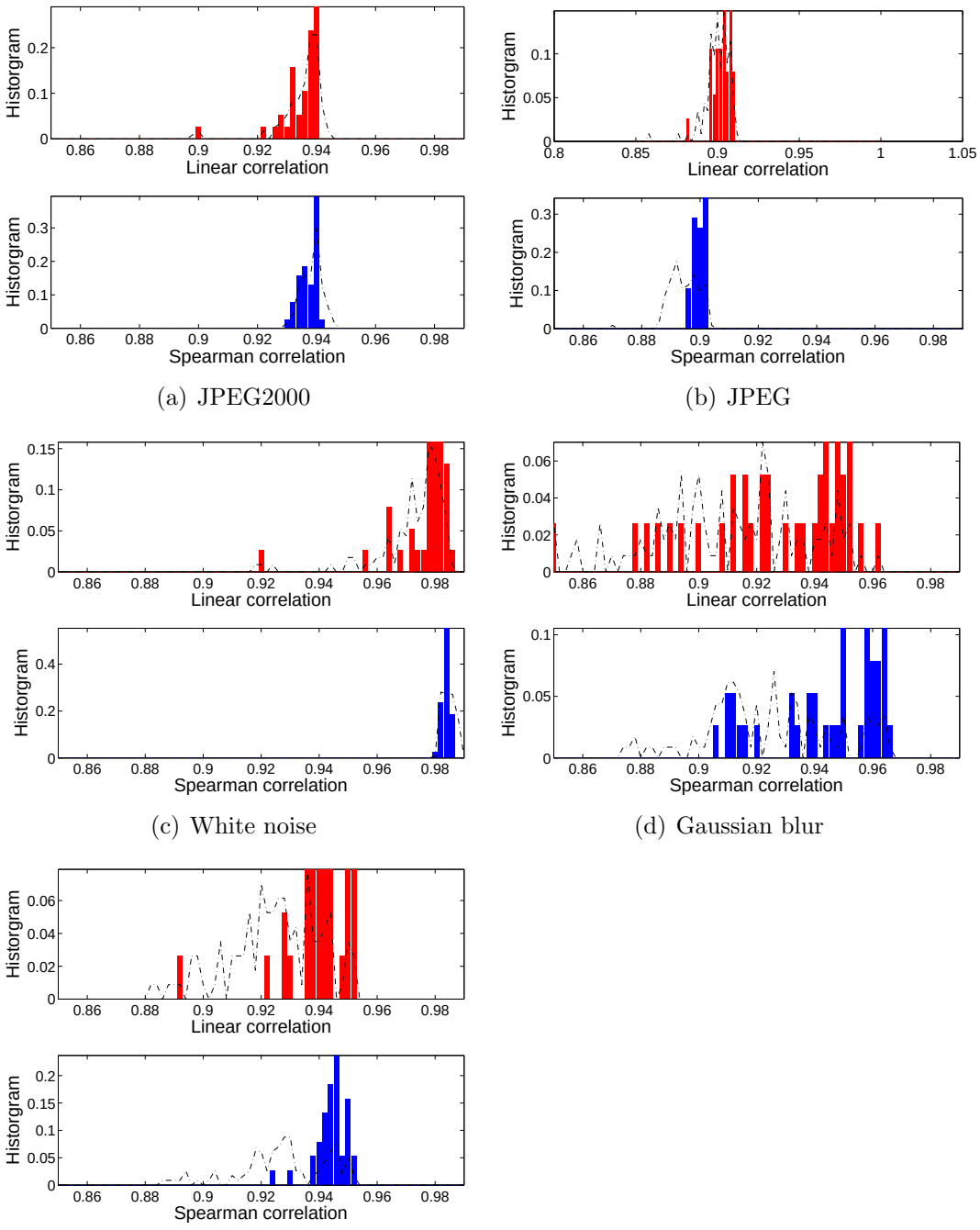
Table 5.2 lists the test results of ITC under 7 different wavelet frameworks. Some of them are orthogonal wavelets [70], such as "db1", "db3", "db6". The rest of them are non-orthogonal [70], such as "bior1.1", "bior2.4" and the steerable pyramid [104]. Table 5.2 illustrates that ITC performs consistently under different multi-resolution frameworks. The linear correlation and Spearman correlation vary within a very small interval [0.95, 0.99]. Also the test results show that there is no significant improvement with non-orthogonal wavelets, such as biorthogonal wavelets or steerable pyramid.

We test 38 possible multi-resolution frameworks in total, which cover Daubechies [70], biorthogonal [70] and reverse biorthogonal [70] wavelets. Linear correlation and Spearman correlation between the subjective judgments and objective evaluations are computed for each wavelet. Instead of listing many correlation values in a huge table, we plot the histogram of linear correlation and Spearman correlation. Please note: to illustrate the performance of ITC under different multi-resolution frameworks, the correlation is computed with test images which do have actual distortions. As shown in Figure 5.5, the histogram of linear correlation coefficients is plotted in the top half of each figure, and the histogram of Spearman correlation coefficients is plotted in the bottom half. It is clear that most multi-resolution frameworks provide excellent quality judgments: the linear correlation coefficients vary within a small interval [0.88 0.99]. For JPEG2000, JPEG, white noise and fastfading distortion, the correlation coefficients concentrate on 0.93, 0.9, 0.98 and 0.94. For Gaussian blur distortion, the correlation spreads among [0.88 0.97]. Based on the above observation, we can conclude that ITC provides the consistent evaluations with human observers and ITC is not significantly affected by the particular multi-resolution framework.

We also find that ITC using image statistics in the horizontal direction yields the better performance. The bars in Figure 5.5 indicate the histogram of the correlation of ITC using the wavelet statistics in the horizonal direction. The dashed lines in each figure indicate the correlation histogram of ITC results using statistics in three directions. For JPEG2000 and white noise distortions, the dashed lines do not shift from the color bars. For JPEG, Gaussian blur and fastfading distortions, the dashed lines tend to shift to the left side of the color bars. This means the performance of ITC statistically degrades when using statistics in vertical and diagonal directions.

improvement on quality assessment. Hence this section focuses on the ITC using the simple Gaussian model.

[2]The parameters are selected by sampling all possible choices.



(a) JPEG2000

(b) JPEG

(c) White noise

(d) Gaussian blur

(e) Fastfading

Figure 5.5: The impact of different multi-resolution frameworks. Figure (a) - (e) plot the histogram of correlation coefficients. The histogram of linear correlation coefficients is plotted in the top half of each figure, and the histogram of Spearman correlation coefficients is plotted in the bottom half. It is clear that most multi-resolution frameworks provide excellent quality judgments: the linear correlation coefficients vary within a small interval [0.88 0.99]. The dashed lines indicates the histogram of the correlations of ITC using the wavelet statistics in three directions. The bars indicates the histogram of the correlations of ITC using the wavelet statistics in the horizonal direction.

Framework	JPEG2000		JPEG		White Noise	
	CC	SC	CC	SC	CC	SC
db1	0.9807	0.9741	0.9707	0.9566	0.9919	0.9916
db3	0.9809	0.9740	0.9702	0.9568	0.9910	0.9900
db6	0.9798	0.9726	0.9694	0.9553	0.9908	0.9905
bior1.1	0.9807	0.9741	0.9707	0.9566	0.9919	0.9916
bior2.4	0.9806	0.9738	0.9716	0.9571	0.9879	0.9897
bior3.7	0.9767	0.9718	0.9709	0.9560	0.9847	0.9891
Steerable Pyramid	0.9791	0.9755	0.9693	0.9574	0.9937	0.9896
Framework	Blur		Fastfading		Overall	
	CC	SC	CC	SC	CC	SC
db1	0.9778	0.9702	0.9777	0.9704	0.9790	0.9696
db3	0.9768	0.9778	0.9729	0.9666	0.9778	0.9690
db6	0.9714	0.9704	0.9708	0.9634	0.9761	0.9665
bior1.1	0.9778	0.9702	0.9777	0.9704	0.9790	0.9696
bior2.4	0.9750	0.9784	0.9712	0.9669	0.9771	0.9688
bior3.7	0.9549	0.9604	0.9713	0.9682	0.9723	0.9655
Steerable Pyramid	0.9792	0.9715	0.9380	0.9026	0.9594	0.9502

Table 5.2: Linear correlation of subjective quality and objective quality under different multi-resolution frameworks. "CC" means linear correlation. "SC" means Spearman correlation.

5.3 The Impact of Different Statistical Models on Quality Assessment

In Chapter 3, we reviewed three statistical models: the simple Gaussian model, the hidden Markov tree model [29] and the Gaussian scale mixture model [84]. In this section, we are going to test the impact of three different statistical models. In the literature [29, 39, 51, 82, 84, 85], the HMT and GSM models have significant advantage on image restoration. We are expecting a better performance of ITC using the HMT or GSM model.

In this section, ITC using the GSM model is tested with different neighborhoods (3×3 and 5×5). Also ITCs using the basic GSM model and advanced GSM model are tested. The basic GSM model simulates intra-scale statics only, while the advanced GSM model includes the parent coefficient in the neighborhood ($3 \times 3 + 1$ or $5 \times 5 + 1$). Table 5.3 listed the Spearman correlation coefficients of subjective quality and objective quality. It is quite difficult to notice any performance difference by changing the neighborhood. For JPEG2000 distortion, the largest Spearman correlation is 0.9759 and the smallest Spearman correlation is 0.9687. The difference is only

0.0072. It seems that different multi-resolution frameworks do not have great impact on the performance of ITC when the GSM model is applied.

Wavelet	JPEG2000		JPEG		White Noise		Blur		Fastfading	
	3×3	$3\times3+1$	3×3	$3\times3+1$	3×3	$3\times3+1$	3×3	$3\times3+1$	3×3	$3\times3+1$
db1	0.9751	0.9753	0.9593	0.9603	0.9801	0.9865	0.9830	0.9836	0.9707	0.9732
db2	0.9730	0.9729	0.9526	0.9542	0.9801	0.9849	0.9847	0.9841	0.9699	0.9709
db3	0.9727	0.9725	0.9499	0.9516	0.9801	0.9846	0.9830	0.9833	0.9697	0.9706
bior1.1	0.9751	0.9753	0.9593	0.9603	0.9801	0.9865	0.9830	0.9836	0.9707	0.9732
bior1.3	0.9745	0.9746	0.9588	0.9586	0.9801	0.9861	0.9840	0.9845	0.9714	0.9731
bior1.5	0.9749	0.9735	0.9527	0.9586	0.9887	0.9851	0.9807	0.9827	0.9578	0.9745
	5×5	$5\times5+1$	5×5	$5\times5+1$	5×5	$5\times5+1$	5×5	$5\times5+1$	5×5	$5\times5+1$
db1	0.9759	0.9763	0.9599	0.9600	0.9865	0.9865	0.9846	0.9849	0.9727	0.9733
db2	0.9742	0.9742	0.9543	0.9548	0.9852	0.9852	0.9867	0.9865	0.9715	0.9720
db3	0.9734	0.9735	0.9516	0.9523	0.9844	0.9846	0.9845	0.9841	0.9711	0.9714
bior1.1	0.9759	0.9763	0.9599	0.9600	0.9865	0.9865	0.9846	0.9849	0.9727	0.9733
bior1.3	0.9757	0.9758	0.9597	0.9594	0.9867	0.9867	0.9858	0.9860	0.9735	0.9739
bior1.5	0.9753	0.9754	0.9602	0.9600	0.9862	0.9863	0.9849	0.9848	0.9744	0.9751

Table 5.3: Spearman correlation of subjective quality and objective quality with the Gaussian scale mixture model

ITC using the HMT model is also tested with different multi-resolution frameworks. The test results are listed in Table 5.4. All of the Spearman correlation coefficients are larger than 0.9, which means the objective quality evaluation is consistent with the human judgments. Also ITC based on the HMT model is not sensitive to different multi-resolution frameworks. However the HMT model does have a shortcoming. To estimate the parameters of the HMT model, Crouse *et al.* proposed an expectation-maximization (EM) algorithm [29], which usually takes a long time to converge. Also we find the EM algorithm does not converge on 10% of the test samples.

Now we would like to compare the performance of ITC using three statistical models (the simple Gaussian model, GSM model and HMT model). Figure 5.6 illustrates the Spearman correlation coefficients. Three wavelets (db1, db2 and db3) are used. For each distortion, the performance difference is quite small. For JPEG2000 distortion, the correlation is 0.9726 (the simple Gaussian model), 0.9727 (the GSM model) and 0.9601 (the HMT model). For Gaussian blur distortion, ITC using the simple Gaussian model performs worst (Spearman correlation is 0.9702). But the best Spearman correlation is 0.9830, and the difference is less than 0.015. We also notice that ITC using the simple Gaussian model outperforms the other two on JPEG and white noise distortion. The result is interesting: the advanced models are not guaranteed to generate a better quality evaluation. In other words, we could choose any statistical model for ITC, or the one with the best performance on the target distortion. When the computation complexity is a concern, ITC with the simple Gaussian model could be

Wavelet	JPEG2000	JPEG	White Noise	Blur	Fastfading
db1	0.9707	0.9478	0.9820	0.9725	0.9531
db2	0.9670	0.9472	0.9818	0.9753	0.9541
db3	0.9601	0.9493	0.9833	0.9733	0.9566

Table 5.4: Spearman correlation of subjective quality and objective quality with the hidden Markov tree model

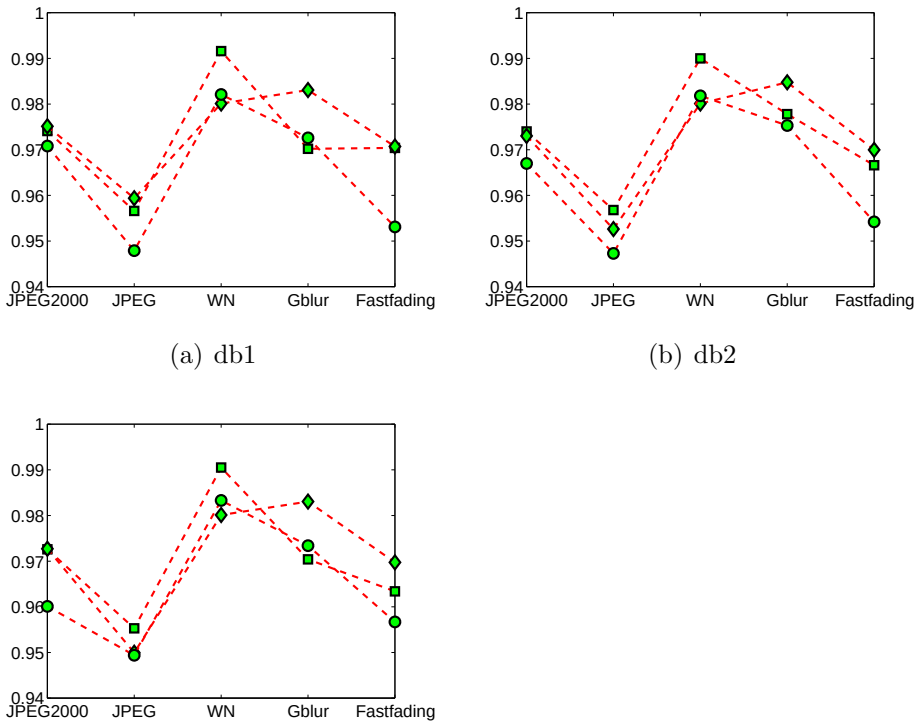
a proper choice.

Although advanced statistical models bring no improvements on ITC performances, we find another statistical model improves the ITC performance especially on fastfading distortion. Among the distortions we are dealing with, fastfading is distinct. It is caused by the JPEG2000 compression distortion (globally applied to the entire image) and data lost (locally applied, due to the compression standard). Usually when the distortion is small, JPEG2000 compression distortion is less noticeable. Hence small fastfading distortion illustrates local characteristics. This inspires us to model image statistics locally.

The basic idea is that each scale is divided into non-overlapped neighborhoods and the coefficients in each neighborhood are modeled independently. In this thesis, local modeling only adopts the simple Gaussian model. According to our description, the computation of the perceptual information needs a little modification. Let X_{m_i,n_i} denote the wavelet coefficient of the neighborhood centered at (m_i, n_i) in the ith scale. X_{m_i,n_i} is assumed to obey the Gaussian distribution. According to the perceptual information expression in Table 4.2, the perceptual information of X_{m_i,n_i} can be computed in Equation 5.10, where $\sigma^2_{X_{m_i,n_i}}$ is the variance of the wavelet coefficients within that neighborhood.

$$I(X_{m_i,n_i}; Y_{m_i,n_i}) = \frac{1}{2}\log(\frac{\sigma^2_{X_{m_i,n_i}} + \sigma^2_{Z_i}}{\sigma^2_{Z_i}}) \tag{5.10}$$

With the independence assumption among the neighborhoods and scales, the perceptual information within the entire scale can be computed in Equation 5.11 and Equation 5.12. Similarly, we can derive $\mathcal{P}(\underline{D})$ and $\mathcal{P}(\underline{X}_d)$. Follow the ITC strategy in Page 78, ITC based on local model can be applied to evaluate visual quality with Equation 5.5. Please note, the local modeling has strong connection with local energy consistency model in Chapter 3 Section 3. In Chapter 6, both are essential to address

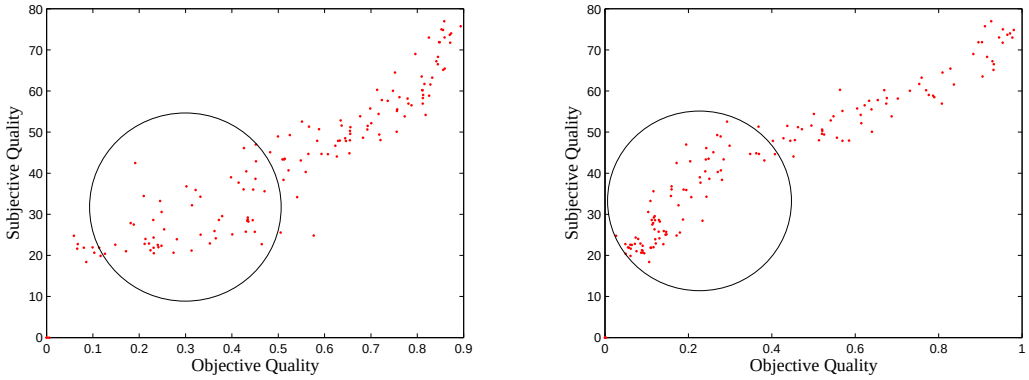


(a) db1

(b) db2

(c) db3

Figure 5.6: Linear correlation of subjective quality and objective quality under different statistical models. □: *ITC* using the simple Gaussian model. ♢: *ITC* using the GSM model. ○: *ITC* using the HMT model



(a) (b)

Figure 5.7: The performance of ITC on fastfading distortion based on local modeling

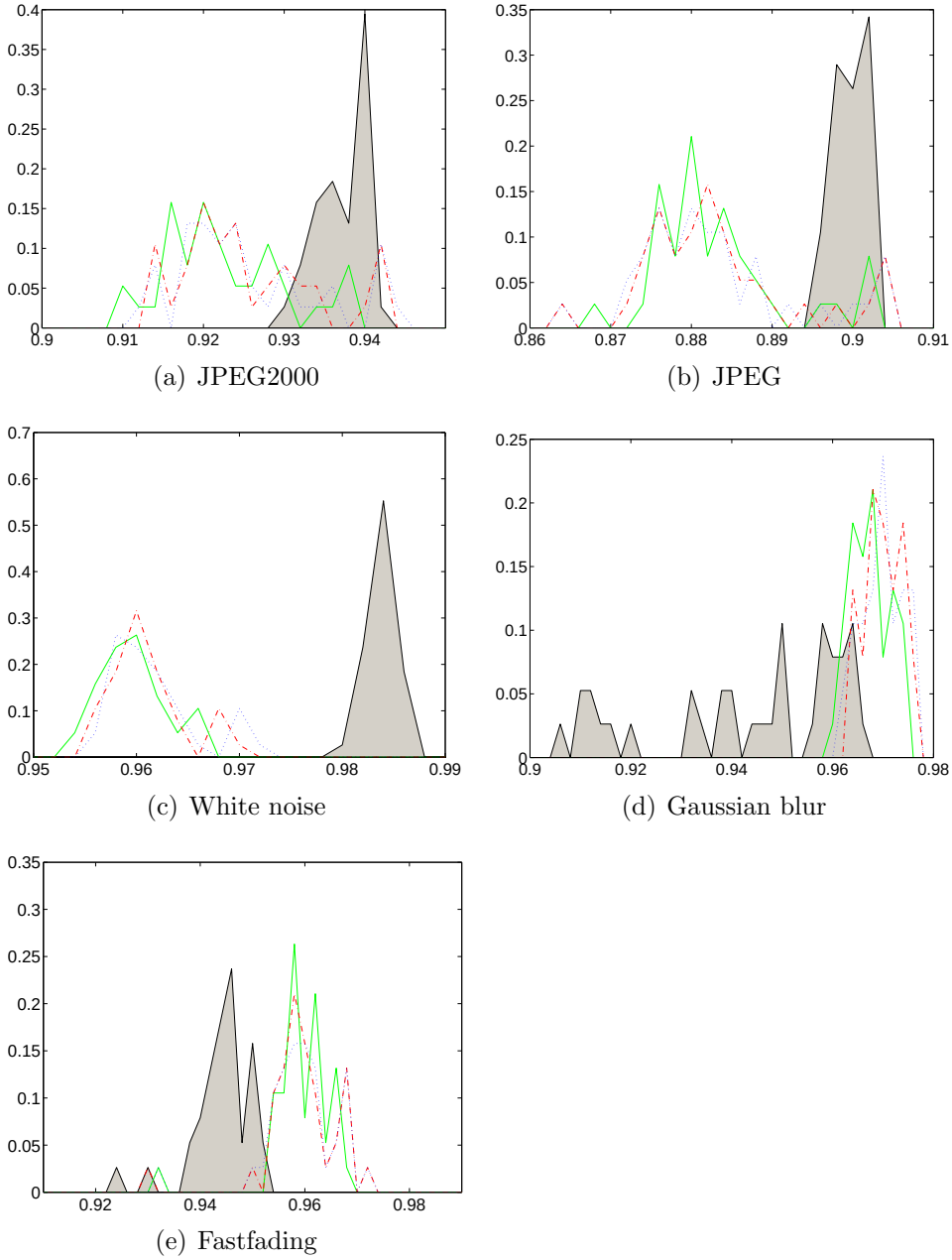
blind quality assessment.

$$\mathcal{P}(X_i) = \sum_{m_i,n_i} I(X_{m_i,n_i}; Y_{m_i,n_i}) \tag{5.11}$$

$$\mathcal{P}(\underline{X}) = \sum_{i} \sum_{m_i,n_i} I(X_{m_i,n_i}; Y_{m_i,n_i}) \tag{5.12}$$

The test results show that local modeling greatly improves ITC performance on fastfading distortion, especially for images with small distortions, as shown in the circled area in Figure 5.7. We can conclude by visual observation that objective quality in Figure 5.7(b) (neighborhood size is 5×5) has better consistency to subjective quality. Linear correlations are 0.9594 and 0.9740 respectively for global modeling and local modeling. Spearman correlations are 0.9502 and 0.9705.

More test results are plotted in Figure 5.8. Please note: to illustrate the performance of local modeling, the correlation is computed with test images which do have actual distortions. In this experiment, we test the ITC performance using local modeling with different multi-resolution frameworks. The neighborhood sizes are 3×3, 5×5 and 7×7. The histograms of Spearman correlation are plotted. The solid (green) lines indicate the results of ITC using 3×3 neighborhood, while the dashed (red) lines for 5×5 neighborhood and the dot (blue) lines for 7×7 neighborhood. The results of ITC using the entire subband statistics are plotted in black solid lines with shadow areas beneath. First, we notice that ITC with different neighborhood sizes performs consistently. It is difficult to conclude that smaller or larger neighborhood provides better evaluation. Second, ITC based on local modeling has much better performance on



(a) JPEG2000 (b) JPEG

(c) White noise (d) Gaussian blur

(e) Fastfading

Figure 5.8: The impact of different neighborhood sizes. The *ITC* performances using local modeling with different multi-resolution frameworks are plotted. The neighborhood sizes are 3×3, 5×5 and 7×7. The histograms of Spearman correlation are plotted. The solid (green) lines indicate the results of *ITC* using 3×3 neighborhood, while the dashed (red) lines for 5×5 neighborhood and the dot (blue) lines for 7×7 neighborhood. The results of *ITC* using the entire subband statistics are plotted in black solid lines with shadow areas beneath. Image statistics in the horizonal direction are used in the experiment.

	JPEG2000	JPEG	White Noise	Gaussian Blur	Fastfading
Steerable Pyramid	0.9736	0.9552	0.9824	0.9837	0.9817
db1	0.9461	0.9312	0.9642	0.9797	0.9740
db3	0.9367	0.9272	0.9774	0.9761	0.9642
db6	0.9310	0.9250	0.9692	0.9735	0.9583
bior2.4	0.9373	0.9272	0.9803	0.9743	0.9594
bior3.7	0.9315	0.9249	0.9820	0.9750	0.9544

Table 5.5: The performance of ITC using local modeling. The neighborhood size is 5×5. The table lists Spearman correlation between the subjective judgments and objective evaluations by ITC. The steerable pyramid [104] and the wavelets with small vanishing moments [70] have the better performance with local modeling is applied.

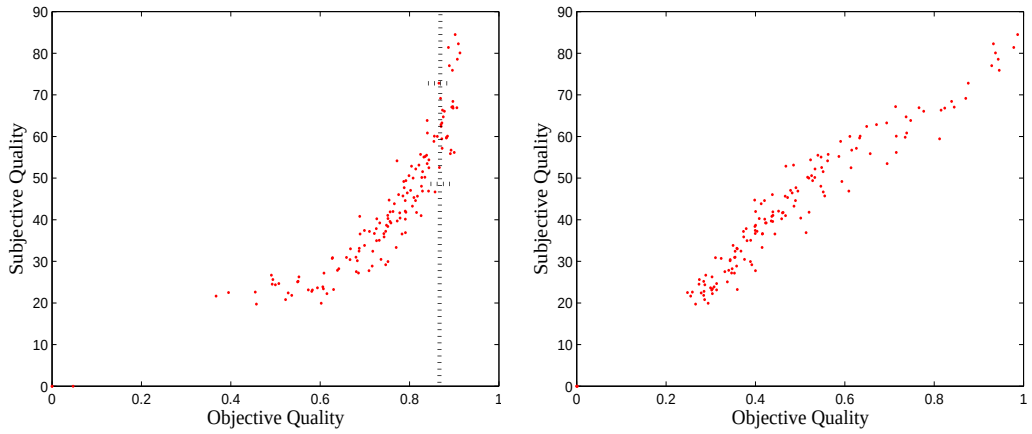
Gaussian blur and fastfading distortions. In Figures 5.8(d) and 5.8(e), the histograms of ITC using local modeling shift to the right side of ITC using global modeling. This means introducing local modeling improves the ITC performance statistically. The improvement does not only when a certain wavelet is used. Third, the local modeling is not suitable for white noise distortion, as shown in Figure 5.8(c). Forth, ITC using local modeling achieves the better performance on JPEG2000 and JPEG distortion. Consider the overall performance on different distortions, ITC based on local modeling can achieve the better performance. Among 38 wavelets, the steerable pyramid [104] and the wavelets with small vanishing moments [70] have the better performance with local modeling is applied, as shown in Table 5.5. In our ITC implementation, the steerable pyramid is chosen to transform images into the wavelet domain since it provides the best performance when local modeling is used.

5.4 Human Visual Contrast Sensitivity Model Settings

Chapter 4 explains that the noise level in each channel of the contrast sensitivity model indicates the vision sensitivity at certain frequency. It is also widely known that the human visual system has different contrast sensitivity at different frequencies. This implies that the setting ($\sigma^2_{Z_i}$) of the contrast sensitivity model could be important to quality assessment.

Figure 5.9(a) and 5.9(b) illustrate the importance of setting proper noise level. In Figure 5.9(a), the noise level in each scale (frequency) is set equally. Apparently, such a setting causes two problems. First, the objective quality and subjective quality do not have a linear relationship which make it inconvenient to use objective quality directly. More importantly the wrong setting greatly degrades the accuracy of ITC in the poor quality range. In Figure 5.9(a), if the evaluated quality is 0.86 where the dotted line is, the corresponding subjective quality varies from 45 to 70, which is almost one third of the subjective quality range. This is unacceptable for a successful quality metric. There are some studies [97, 100, 103, 124] that use nonlinear mapping to calibrate the objective quality scores with the subjective quality scores. But determining the proper nonlinear mapping is extra work.

In our thesis, the noise variances in the contrast sensitivity model are selected by testing sampling all possible choices. The selected noise variance setting has the best performance, i.e. Spearman correlation between the subjective judgments and objective quality evaluations by ITC. If the test image is 8-bit gray scale, the noise variance $\sigma^2_{Z_i}$ in the first four scales are 1.1314, 0.8000, 0.5657 and 0.4000. The noise is strong at high frequency but weak at low frequency. This means the contrast sensitivity model is sensitive at low frequency but less sensitive at high frequency, which matches the psychophysical experiment conclusion of contrast sensitivity.



(a) Improper setting

(b) Proper setting

Figure 5.9: The impact of noise level setting. Figure (a) and (b) illustrate the importance of setting proper noise level. In Figure (a), the noise level in each scale (frequency) is set equally. Apparently, such a setting causes two problems. First, the objective quality and subjective quality do not have a linear relationship which make it inconvenient to use objective quality directly. More importantly the wrong setting greatly degrades the accuracy of ITC in the poor quality range. If the evaluated quality is 0.86 (the dashed line) where the dotted line is, the corresponding subjective quality varies from 45 to 70, which is almost one third of the subjective quality range.

5.5 Implementation of ITC

---

0. Convert the test image into 8 bit gray scale, if needed.

1. Transform the test image and the reference image into four scales with the Steerable Pyramid [104]

2. Compute the difference between the test image and the reference image in the wavelet domain, i.e. $D_i = X_i - X_{d_i} \quad \forall i$

3. Compute the perceptual information

   a. Select the neighborhood size (5×5) for local modeling.

   b. Divide each scale into non-overlapped neighborhoods.

   c. Compute the variance ($\sigma^2_{D_{m_i,n_i}}$, $\sigma^2_{X_{m_i,n_i}}$ and $\sigma^2_{X_{d_{m_i,n_i}}}$) of the wavelet coefficients in each neighborhood for the test image, reference image and difference image.

   d. Compute the perceptual information (using the simple Gaussian model) in the horizontal direction according to Equation 5.11 and Equation 5.12.

4. Compute the visual quality for the test image according to its reference with

   ITC definition in Equation 5.5.

5.6 Performance and Discussion

This section will first discuss the performance of ITC on different subjective quality databases and then compare the test results for different quality metrics.

ITC is tested with the LIVE [99], IVC [64], TOYAMA [55], Cornell A57 [25] and TID2008 [81] databases. The LIVE, IVC and TOYAMA use the DSCQS subjective quality assessment method [2]. The Cornel A57 and TID2008 use two different customized subjective quality assessment methods [25, 81].

Different subjective evaluation methods result in different quality assessment performances. For the LIVE, IVC, TOYAMA and Cornell A57 databases, our quality metric performance is excellent using image statistics across all four scales. For the TID2008

database, our metric performs consistently on most distortions using low frequency subbands (except for high frequency noise distortion).

Table 5.7 lists the ITC performance on five databases. For the LIVE database, ITC has the best performance on white noise distortion and Gaussian blur distortion. The worst performance happens on JPEG distortion. For IVC database, ITC performs best on Gaussian blur distortion and worst on JPEG distortion. For TOYAMA database, ITC keeps performing consistently. Since Cornell A57 database only has 56 test samples, we do not list different distortions separately. For the above four subjective databases, the setting of ITC does not need to change to achieve consistent performance.

For TID2008 database, ITC using low frequency subband (the fourth scale) provides the consistent performance, as listed in Table 5.7. ITC has consistent performance on 11 out of 17 distortions with linear correlation and Spearman correlation greater than 0.85. Since the ITC metric is developed to evaluate the quality in a gray scale, it focuses on differentiating the structural distortions. ITC is not designed to address quality degradation caused by color distortions, such as mean shift and contrast change distortions. For JPEG or JPEG2000 transmission error distortion, ITC also does not perform very well.

Now we compare ITC with the other 17 quality metrics. All detailed test results are displayed in Appendix E with Tables E.1, E.2, E.3, E.4 and E.5. It is quite difficult to check all of the test results in five tables and figure out the most reliable quality metrics. We propose a simple metric that describe the overall performance of quality metrics over all kinds of distortions. For a certain distortion, we rank the performances of quality metrics according to the linear correlation or Spearman correlation. For instance, Metric A has 0.99 linear correlation with subjective quality judgments and Metric B has 0.9 linear correlation. Then the rank for Metric A is 1 and for Metric B is 2. If subjective database contains N types of distortions and M quality metrics are tested. We denote the rank of ith metric for jth distortion as $r_{i,j}$. Then the overall rank R_i of ith metric can be computed as:

$$R_i = \frac{\sum_j r_{i,j}}{N} \tag{5.13}$$

From the definition, the overall rank R_i varies from 1 to M. The higher overall rank means the quality metric exhibits better consistency with human judgment. We plot the overall rank for each metric in Figure 5.6. ITC has the highest overall ranks according to Spearman correlation, which is 3.8. Overall ranks for other

quality metrics can be found in Table 5.6. Based on the test results on five subjective databases, there are five quality metrics that illustrate good compatibility to difference databases. They are information theoretic criterion, visual information fidelity [97], structural similarity [122], multi-scale structural similarity [128] and visible difference predictor [17].

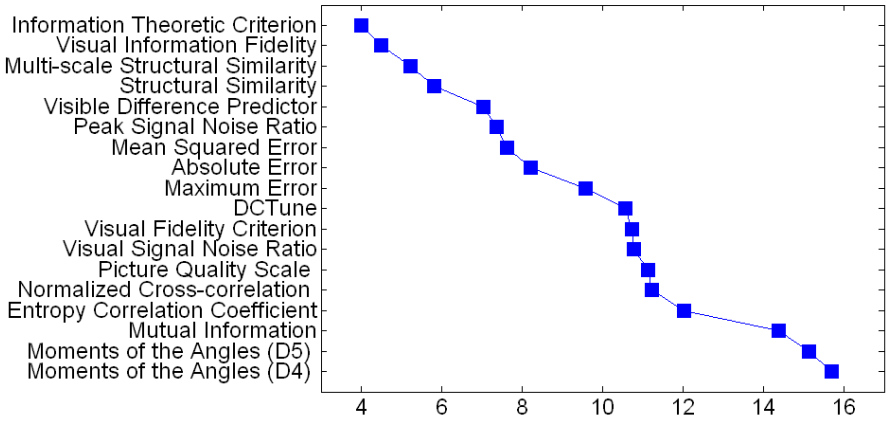


Figure 5.10: Overall rank of image quality metric. ■ Overall rank according to Spearman correlation

5.7 Summary

In this chapter, we first propose an information theoretic criterion ITC for image quality. We study ITC from different perspectives, test ITC with different subjective databases, and compare it with several quality metrics. The chapter is summarized below.

- ITC simulates contrast sensitivity with a parallel Gaussian channel. Mathematically, ITC logarithmically responds to the image distortion $\underline{D}$ between a distorted image and its reference image.
- ITC is not sensitive to different multi-resolution frameworks. It can be implemented as a computationally efficient quality assessment metric.
- Complicated statistical models will not provide noticeable benefit compared with simple statistical models.

QA methods	Overall rank R_i
Mean Squared Error	7.6
Absolute Error	8.2
Maximum Error	9.6
Moments of the Angles (D_4)	15.7
Moments of the Angles (D_5)	15.1
Normalized Cross-correlation	11.2
Mutual Information	14.4
Entropy Correlation Coefficient	12.0
Peak Signal Noise Ratio	7.4
Visual Fidelity Criterion	10.7
DCTune	10.6
Picture Quality Scale	11.1
Visible Difference Predictor	7.0
Visual Signal Noise Ratio	10.8
Structural Similarity	5.8
Multi-structural Similarity	5.2
Visual Information Fidelity	4.5
Information Theoretic Criterion	**4.0**

Table 5.6: Overall rank of image quality metric

- Local modeling has the potential to improve the performance of ITC. Even local modeling might not improve quality assessment on certain distortions. We still consider local modeling as an essential component for ITC.

- The noise level in human visual system model is important for excellent performance of ITC.

JPEG2000[1]	JPEG[1]	White Noise[1]	Blur[1]	Fastfading[1]
0.9674	0.9729	0.9722	0.9700	0.9670
JPEG2000[2]	JPEG[2]	LAR coding[2]	Blur[2]	JPEG gray[2]
0.9387	0.9507	0.9022	0.9653	0.9179
JPEG2000[3]	JPEG[3]	Cornell A57[4]	Gaussian noise[5]	Noise in color components[5]
0.9581	0.9178	0.8454	0.9184	0.9092
Spatially correlated noise[5]	Masked noise[5]	High frequency noise[5]	Impulse noise[5]	Quantization noise[5]
0.9080	0.8479	0.9362	0.8841	0.8938
Gaussian blur[5]	Image denoising[5]	JPEG[5]	JPEG2000[5]	JPEG errors[5]
0.9192	0.9330	0.9318	0.9639	0.6466
JPEG2000 errors[5]	Non-eccentricity pattern noise[5]	Local block-wise distortions[5]	Mean shift (intensity shift)[5]	Contrast change[5]
0.5628	0.7946	0.7542	0.7079	0.6367

Table 5.7: The performance of quality metrics for LIVE database. The table lists linear correlation "CC" and Spearman correlation "SC" between the subjective quality and objective quality evaluations by *ITC*. The superscript (1 to 5) of each distortion indicates where the test images come from. [1]: the LIVE database [99]. [2]: the IVC database [64]. [3]: the TOYAMA database [55]. [4]: the Cornell A57 database [25]. [5]: the TID2008 database [81].

Chapter 6

Blind Quality Assessment

In this chapter, we propose an information theoretic framework for blind quality assessment. Without reference information, it is much more difficult to evaluate image quality [98, 127]. Compared with full-reference image quality metrics, few blind quality metrics [90, 91, 98, 100, 103, 124] have been proposed. Chapter 2 reviewed most of those. Generally, there are two types of approaches to blind assessment. In the first family [90, 91, 124], researchers focus on the visual distortion in spatial domain, such as discontinuity, block effect, or small local variance (due to blur or quantization). By combining zero crossing rate, local variance, and other spatial features, people are able to predict image quality. The other family [98, 100, 103] utilizes the statistics of images with excellent quality. Image quality can be computed as the deviation from the empirical statistics [100]. The blind quality metrics proposed in this chapter belong to the second family. Before we start introducing blind quality assessment methods, one important issue needs to be addressed. It is a complex procedure for people to evaluate image quality without access to the reference image. In this chapter, we propose several methods for image statistics estimation. Although these methods are not perfectly accurate, they are able to provide a good approximation for the purpose of blind quality assessment.

In this chapter, we focus on the five distortions in the LIVE database [99]. For each distortion, there are 170-233 test samples. The validation of the blind quality metrics is not limited to the LIVE database. For JPEG and JPEG2000 distortion, the blind quality metrics are tested by four subjective databases (the LIVE [99], TOYAMA [55], IVC [64] and TID2008 [81] databases). For white noise and Gaussian blur distortion, the blind quality metrics are tested by two databases (the LIVE and TID2008 databases).

6.1 The Basic Idea of Blind Quality Assessment

Chapter 5 proposes a successful full reference quality assessment method (FRQA). Our goal is to utilize this framework to achieve blind quality assessment. The basic idea of our FRQA is to measure the perceptual difference between the test image and its reference and compare it with the perceptual information of the test image and its reference, as shown in Equation 6.1.

$$ITC = \frac{\mathcal{P}(\underline{D})}{\sum_i \max\{\mathcal{P}(X_i), \mathcal{P}(X_{d_i})\}} \tag{6.1}$$

If the simple Gaussian model is applied to describe the image statistics in each scale, the perceptual information of $\underline{D}$ is computed as Equation 6.2.

$$\mathcal{P}(\underline{D}) = \sum_i \mathcal{P}(D_i) = \sum_i \frac{1}{2}\log(1 + \frac{\sigma_{D_i}^2}{\sigma_{Z_i}^2}) \tag{6.2}$$

To follow the FRQA strategy to evaluate the test image quality without its reference, the key step is to estimate the perceptual difference. From Equation 6.2, the problem becomes how to estimate the variance of the difference in each scale. The following five sections develop the approaches to estimate the image difference variance under different circumstances.

6.2 White Noise Distortion

White noise distortion is the simplest distortion among the five distortions. To estimate the white noise variance is a classic problem. Two methods are already proposed by Donoho *et al.* [36] and Staelin *et al.* [106]. These two methods estimate the white noise variance σ_D^2 in the spatial domain. In Chapter 3 Section 3, we propose a white noise variance estimation method (Equation 3.15) based on the energy consistency model. We take it as our third method. Hence we have three estimation methods.

Under the orthogonal wavelet framework, the noise in each scale is still white noise and $\sigma_{D_i}^2$ in each scale is noise variance σ_D^2 in the spatial domain [70]. With the estimated

noise variance ready, it is straight forward to estimate the perceptual information of difference $\hat{\mathcal{P}}(D_i)$ (but not $\mathcal{P}(\hat{D}_i)$) in each scale as shown in Equation 6.3.

$$\hat{\mathcal{P}}(D_i) = \frac{1}{2}\log(1+\frac{\hat{\sigma^2}_{D_i}}{\sigma^2_{Z_i}}) \tag{6.3}$$

Under the orthogonal wavelet framework, the variance of the wavelet coefficients of the reference image in the ith scale can be estimated with Equation 6.4, which comes from the independence between white noise and images. The perceptual information of the original image in each scale can be computed as Equation 6.5. Now we are ready to estimate image quality with the ITC strategy in Equation 6.1.

$$\hat{\sigma^2}_{X_i} = \sigma^2_{X_{d_i}} - \hat{\sigma^2}_{D_i} \tag{6.4}$$

$$\hat{\mathcal{P}}(X_i) = \frac{1}{2}\log(1+\frac{\hat{\sigma^2}_{X_i}}{\sigma^2_{Z_i}}) \tag{6.5}$$

In Chapter 5, the performance of ITC is improved by local modeling. This inspires us to approach blind quality assessment using local modeling. The key step is to estimate the distortion variance locally. For white noise distortion, Chapter 3 Section 3 proposes the local noise variance estimation method based on the local energy consistency model. With Equation 3.30, the noise variance $\sigma^2_{D_{m_i,n_i}}$ of the neighborhood entered at (m_i, n_i) in the ith scale can be estimated. Since local modeling assumes each neighborhood is independent of others, the perceptual difference in the ith scale is the summation of the estimated perceptual difference in every neighborhood, as shown in Equation 6.6.

$$\hat{\mathcal{P}}(D_i) = \sum_{m_i,n_i} \hat{\mathcal{P}}(D_{m_i,n_i}) = \sum_{m_i,n_i} \frac{1}{2}\log(1+\frac{\hat{\sigma^2}_{D_{m_i,n_i}}}{\sigma^2_{Z_i}}) \tag{6.6}$$

The local energy consistency model (Equation 3.29) in Chapter 3 enables us to estimate the variance $\sigma^2_{X_{m_i,n_i}}$ of each neighborhood of the reference image. The perceptual information of the reference image in the ith scale can be estimated by Equation 6.7 according to the perceptual information expression in Table 4.2.

$$\hat{\mathcal{P}}(X_i) = \sum_{m_i,n_i} \hat{\mathcal{P}}(X_{m_i,n_i}) = \sum_{m_i,n_i} \frac{1}{2}\log(1+\frac{\hat{\sigma^2}_{X_{m_i,n_i}}}{\sigma^2_{Z_i}}) \tag{6.7}$$

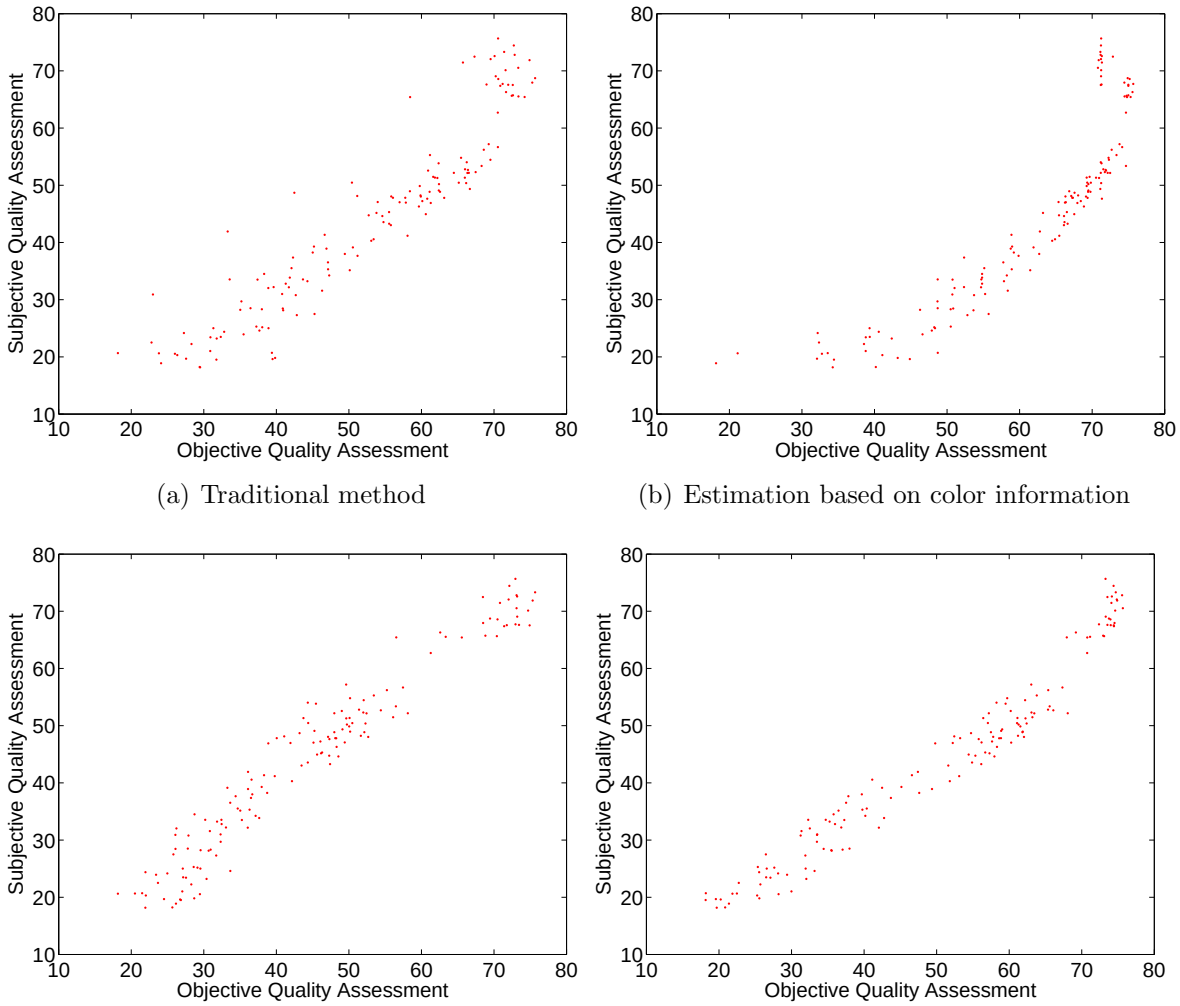
We have the perceptual information computed based on four estimation models. Method 1 uses the traditional noise variance estimation [36]. Method 2 estimates the white noise variance using color information [106]. Method 3 adopts the energy consistency model and Method 4 applies the local energy consistency model and local modeling. ITC in Equation 6.1 can be applied to evaluate the visual quality. When the first four scales are considered during quality evaluation, Spearman correlation between the evaluation results and human judgments is 0.8539. But if the visual quality of a image with white noise distortion is defined in Equation 6.8, a much better performance can be achieved, i.e. Spearman correlation can reach 0.9695.

$$BITC_{\text{wn}} = \frac{\hat{\mathcal{P}}(D_1)}{\hat{\mathcal{P}}(X_1)} \tag{6.8}$$

In the following section, the blind ITC metric for noise distortion is thoroughly tested and the key properties are summarized. The results listed in Figure 6.1 illustrate the performance of $BITC_{wn}$ based on different noise variance estimation methods. In Figure 6.1(a), the traditional noise estimation method [36] is applied. The linear correlation between the subjective judgments and objective prediction is 0.8948. In Figure 6.1(b), the color information based noise estimation method [106] is applied. The linear correlation is 0.8833. This estimation works well when the noise distortion is small, whereas it is not suitable for large noise estimation. In Figure 6.1(c), the energy consistency based noise estimation method is used. The linear correlation is 0.9695. In Figure 6.1(d), the local energy consistency based noise estimation method is used. The linear correlation is 0.9719. By visual inspection, all four methods work fine. But in Figure 6.1(b), we notice that $BITC_{\text{wn}}$ (using Method 2) is not suitable for visual quality evaluation when large noise degradation happens.

Since our energy consistency models are implemented in the wavelet domain, we are interested in which wavelet framework is suitable for $BITC_{\text{wn}}$ (using Method 3 and Method 4). The experiment includes 38 wavelets[1]. The linear correlation between the subjective judgments and objective evaluations are computed for each wavelet. Figure 6.2 shows the linear correlation between subjective and objective judgments. Instead of listing the correlation values in a big table. Figure 6.2(a) - 6.2(c) plot the histogram of linear correlation when $BITC_{\text{wn}}$ uses image statistics in three directions under different wavelet frameworks. The red bars indicate the linear correlation density of $BITC_{\text{wn}}$ using energy consistency. From the correlation histogram plotted in Figure 6.2, most linear correlation is larger than 0.9. The largest linear correlation is 0.9695 when the db4 wavelet is used. The blue bars indicate the linear correla-

[1] 38 wavelets are listed in Appendix J



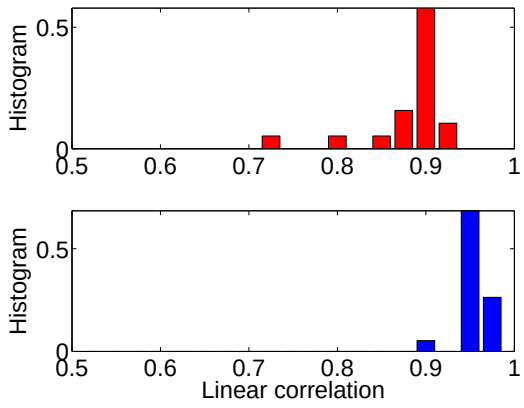
(a) Traditional method

(b) Estimation based on color information

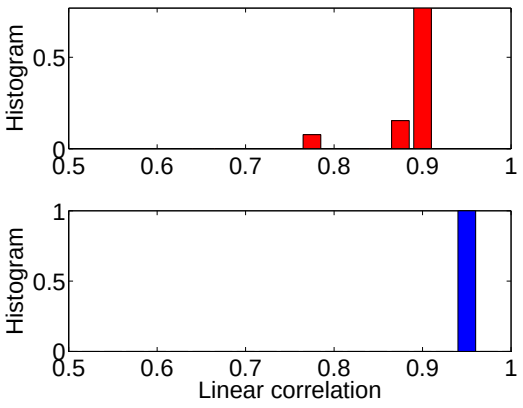
(c) Estimation based on energy consistency

(d) Estimation based on local energy consistency

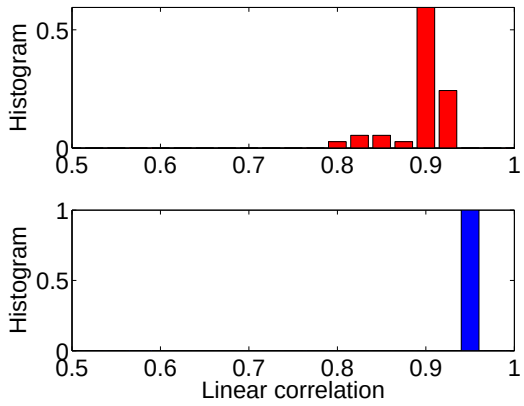
Figure 6.1: Blind quality assessment on white noise distortion. In Figure (a), the traditional noise estimation method [36] is applied. In Figure (b), the color information based noise estimation method [106]. In Figure (c), the energy consistency based noise estimation method. In Figure (d), the local energy consistency based noise estimation method. The linear correlation between the subjective judgments and objective prediction is 0.8948, 0.8833, 0.9695 and 0.9719 respectively. The db4 wavelet is used in Figures (c) and (d).



(a) Horizontal



(b) Vertical



(c) Diagonal

Figure 6.2: The performance of $BITC_{\text{wn}}$ using energy/local energy consistency. $BITC_{\text{wn}}$ is tested under 38 different wavelets. The linear correlation between the subjective judgments and objective evaluations are computed for each wavelet. Figure (a) - (c) plot the histogram of linear correlation when $BITC_{\text{wn}}$ uses image statistics in three directions under different wavelets. The red bars indicate the linear correlation histogram of $BITC_{\text{wn}}$ using energy consistency. The blue bars indicate the linear correlation histogram of $BITC_{\text{wn}}$ using local energy consistency.

Noise in color components		Spatially correlated noise	
CC	SC	CC	SC
0.8001	0.8604	0.8080	0.8640
High frequency noise		Impulse noise	
CC	SC	CC	SC
0.8346	0.8757	0.7630	0.8101

Table 6.1: The performance of blind ITC using local modeling

tion histogram of $BITC_{\text{wn}}$ using local energy consistency. The test results show that the most linear correlation is greater than 0.95 for vertical and diagonal directions. For the test results Spearman correlation, please refer to Appendix F. These results show that $BITC_{\text{wn}}$ metrics have very stable performance under most multi-resolution frameworks.

Our blind quality metric based on the local energy consistency for white noise is not limited to white noise distortion. We found that our blind quality metrics are also able to deal with other distortions: additive noise in color components [81], spatially correlated noise [81], high frequency noise [81] and impulse noise [81]. These four distortions change image spectra greatly. Table 6.2 lists the test results of $BITC_{\text{wn}}$ using local modeling (Method 4). $BITC_{\text{wn}}$ has the best performance when evaluating quality of images with high frequency noise. This performance is predictable. Since images high frequency noise will leave the low frequency components untouched, this makes the energy estimation more accurate.

The implementation of $BITC_{\text{wn}}$ is listed below. Please note that the perceptual information computation is slightly different when local modeling is used.

---

1. Estimate the perceptual information of white noise $\mathcal{P}(D_1)$ in the first scale.

   a. if using local modeling
   Estimate the local noise variance $\sigma^2_{D_{m_1,n_1}}$ with Equation 3.30. The perceptual information of white noise is computed as follows.

   $$\hat{\mathcal{P}}(D_1) = \sum_{m_1,n_1} \hat{\mathcal{P}}(D_{m_1,n_1}) = \sum_{m_1,n_1} \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_{m_1,n_1}}}{\sigma^2_{Z_1}})$$

   b. if not using local modeling
   Estimate the noise variance $\sigma^2_{D_1}$ in the first scale. The perceptual information of white noise is computed as follows.

$$\hat{\mathcal{P}}(D_1) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_1}}{\sigma^2_{Z_1}})$$

2. Estimate the perceptual information of the original image $\mathcal{P}(X_1)$ in the first scale.

   a. if using local modeling

   $$\hat{\mathcal{P}}(X_1) = \sum_{m_1,n_1} \hat{\mathcal{P}}(X_{m_1,n_1}) = \sum_{m_1,n_1} \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_{m_1,n_1}}}{\sigma^2_{Z_1}})$$

   where $\hat{\sigma^2}_{X_{m_1,n_1}}$ can be estimated by Equation 3.29.

   b. if not using local modeling

   $$\hat{\mathcal{P}}(X_1) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_1}}{\sigma^2_{Z_1}})$$

   where $\hat{\sigma^2}_{X_1}$ can be estimated by Equation 6.4.

3. Compute $BITC_{\text{wn}}$ as follows.

$$BITC_{\text{wn}} = \frac{\hat{\mathcal{P}}(D_1)}{\hat{\mathcal{P}}(X_1)}$$

6.3 Gaussian Blur Distortion

As with blind quality assessment for white noise distortion, we follow the ITC strategy to approach blind quality assessment for Gaussian blur distortion. Chapter 3 Section 3 introduces a method based on the energy consistency model to estimate the variance $\sigma^2_{D_i}$ of the image difference in each scale. Briefly speaking, Gaussian blur is modeled as in the wavelet domain [97] by Equation 6.9. The parameter k controls the degree of blur. The variance $\sigma^2_{D_i}$ of the difference in each scale can be simply estimated by Equation 6.10. The energy consistency model in Chapter 3 (Equation 3.14) provides the estimation of $\sigma^2_{X_i}$. Then k^2 can be estimated by Equation 6.11. The detailed derivation can be found in Chapter 3 Section 3 Page 45-46.

$$X_{d_i} = kX_i \tag{6.9}$$

$$\hat{\sigma^2}_{D_i} = (1-\hat{k})^2\hat{\sigma^2}_{X_i} \tag{6.10}$$

$$\hat{k^2} = \frac{\sigma^2_{X_{di}}}{\hat{\sigma^2}_{X_i}} = \frac{\mathcal{E}_{d_i}}{\hat{\mathcal{E}}_i} \tag{6.11}$$

The perceptual difference $\mathcal{P}(D_i)$ and perceptual information of test image $\mathcal{P}(X_{d_i})$ and its reference $\mathcal{P}(X_i)$ can be computed according to the perceptual information expression in Table 4.2. We are ready to estimate the quality of an image with Gaussian blur by ITC in Equation 6.1.

$$\hat{\mathcal{P}}(D_i) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_i}}{\sigma^2_{Z_i}}) \tag{6.12}$$

$$\hat{\mathcal{P}}(X_i) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_i}}{\sigma^2_{Z_i}}) \tag{6.13}$$

$$\mathcal{P}(X_{d_i}) = \frac{1}{2}\log(1 + \frac{\sigma^2_{X_{d_i}}}{\sigma^2_{Z_i}}) \tag{6.14}$$

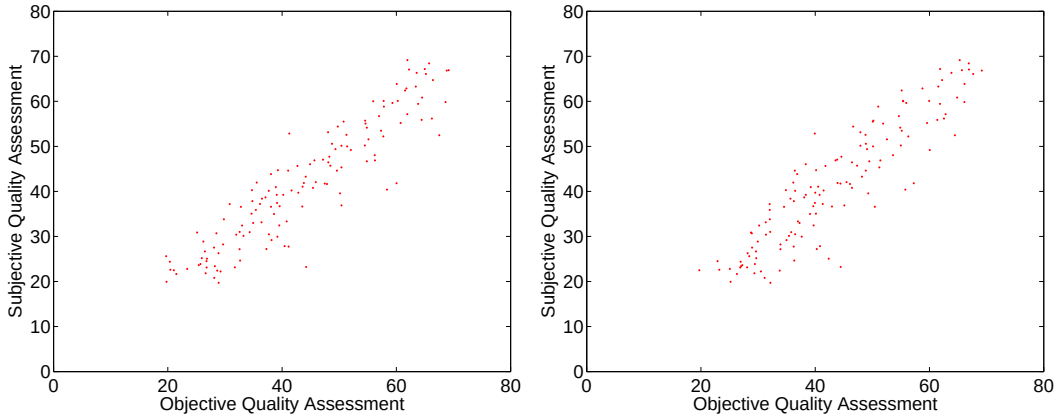
As with blind quality assessment for white noise distortion, the local modeling may benefit the blind quality assessment for Gaussian blur distortion. Chapter 3 Section 3 proposes the local energy consistency model (Equation 3.29) to estimate the variance $\sigma^2_{X_{m_i,n_i}}$ of the reference image in each neighborhood. Follow the above estimation method for the difference variance, it is not difficult to estimate $\sigma^2_{D_{m_i,n_i}}$ in each neighborhood. The detailed derivation can be found in Chapter 3 Section 3 Equation 3.32. The perceptual information of each neighborhood is estimated Equation 6.15. The total perceptual information (Equation 6.16) of the entire scale is the summation of the perceptual information in each neighborhood because local modeling assumes each neighborhood is independent of others. As with the last section, the perceptual information of the reference (original) image can be estimated by Equations 6.17 and 6.18.

$$\hat{\mathcal{P}}(D_{m_i,n_i}) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_{m_i,n_i}}}{\sigma^2_{Z_i}}) \tag{6.15}$$

$$\hat{\mathcal{P}}(D_i) = \sum_{m_i,n_i} \hat{\mathcal{P}}(D_{m_i,n_i}) \tag{6.16}$$

$$\hat{\mathcal{P}}(X_{m_i,n_i}) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_{m_i,n_i}}}{\sigma^2_{Z_i}}) \tag{6.17}$$

$$\hat{\mathcal{P}}(X_i) = \sum_{m_i,n_i} \hat{\mathcal{P}}(X_{m_i,n_i}) \tag{6.18}$$



(a) (b)

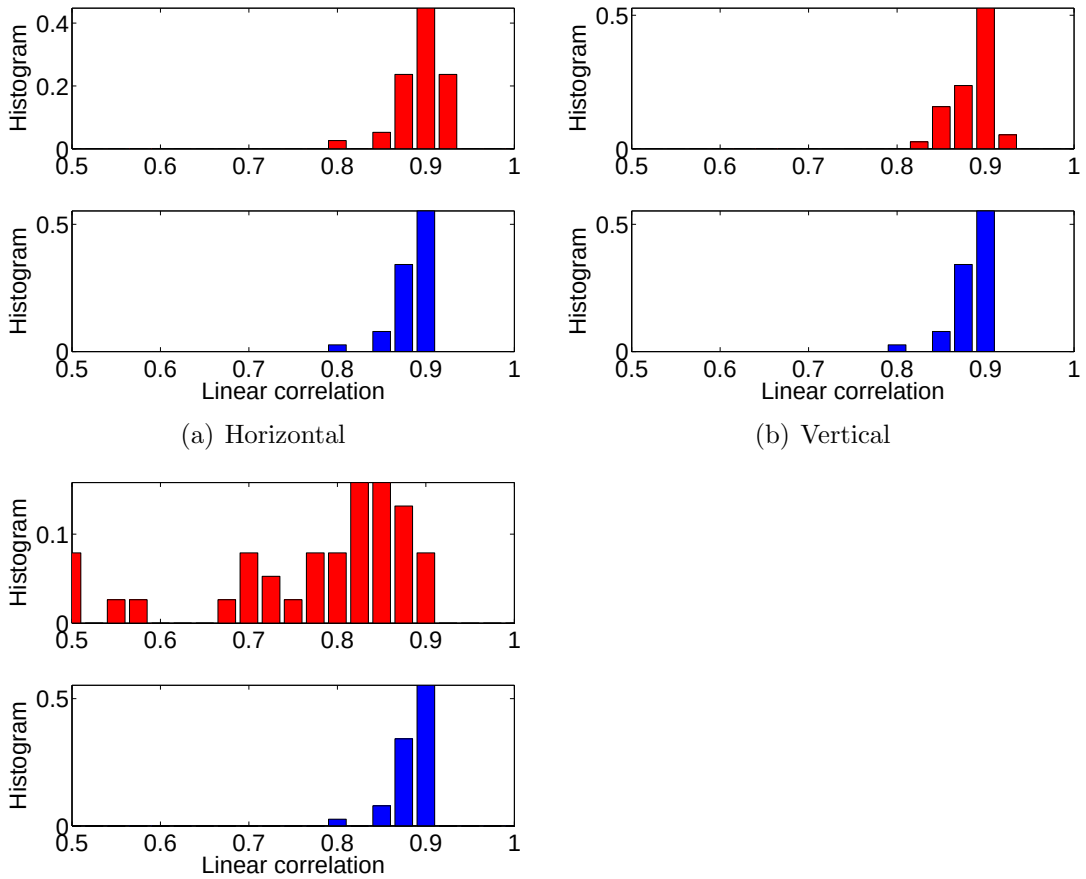
Figure 6.3: Blind quality assessment on blurring distortion. For the method based on the energy consistency model (Figure (a)), the linear correlation between subjective judgments and objective judgment is 0.9204. For the method based on the local energy consistency model (Figure (b)), the linear correlation is 0.9066.

Until now, we have the perceptual information computed based on two estimation models. ITC in Equation 6.1 can be applied to evaluate the visual quality. When the first four scales are considered during quality evaluation, Spearman correlation between the evaluation results and human judgments is 0.8245. But if the visual quality of a image with white noise distortion is defined in Equation 6.19, a much better performance can be achieved, i.e. Spearman correlation can reach 0.9314.

$$BITC_{\text{gblur}} = \frac{\hat{\mathcal{P}}(D_1)}{\hat{\mathcal{P}}(X_1)} \tag{6.19}$$

Figure 6.3 displays the blind quality assessment for Gaussian blur distortion. For the method based on the energy consistency model, the linear correlation between subjective judgments and objective judgment is 0.9204. For the method based on the local energy consistency model, the linear correlation is 0.9066.

In the following session, $BITC_{\text{gblur}}$ is thoroughly tested and the key properties are summarized. As before, $BITC_{\text{gblur}}$ is tested using different wavelet transforms. The histogram of linear correlation using $BITC_{\text{gblur}}$ based on the energy consistency model is plotted with red bars in Figure 6.4. Figures 6.4(a), 6.4(b) and 6.4(c) illustrate the histogram of linear correlation in three different directions. The largest linear correlation is 0.9204 when the db4 wavelet is used. Figure 6.4 also displays the linear cor-



(a) Horizontal

(b) Vertical

(c) Diagonal

Figure 6.4: The performance of $BITC_{\text{gblur}}$ using the energy/local energy consistency. Figure (a) - (c) plot the histogram of correlation coefficients. The histogram of linear correlation coefficients using the energy consistency model in $BITC_{\text{glur}}$ is plotted with red bars, and the histogram of linear correlation coefficients using the local energy consistency model in $BITC_{\text{glur}}$ is plotted with blur bars.

relation histogram of $BITC_{\text{gblur}}$ using local energy consistency under different multi-resolution frameworks. For blur distortion, local energy consistency model does not improve the performance, which is different from what we observed on noise distortion.

The implementation of $BITC_{\text{gblur}}$ is listed below. Please note that the perceptual information computation is slightly different when local modeling is used.

---

1. Estimate the perceptual information of Gaussian blur $\mathcal{P}(D_1)$ in the first scale.

   a. if using local modeling
   Estimate the local distortion variance $\sigma^2_{D_{m_1,n_1}}$ with Equation 3.32. The perceptual information of Gaussian blur is computed as follows.

   $$\hat{\mathcal{P}}(D_1) = \sum_{m_1,n_1} \hat{\mathcal{P}}(D_{m_1,n_1}) = \sum_{m_1,n_1} \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_{m_1,n_1}}}{\sigma^2_{Z_1}})$$

   b. if not using local modeling
   Estimate the distortion variance $\sigma^2_{D_1}$ in the first scale with Equation 6.10. The perceptual information of the distortion is computed as follows.

   $$\hat{\mathcal{P}}(D_1) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{D_1}}{\sigma^2_{Z_1}})$$

2. Estimate the perceptual information of the original image $\mathcal{P}(X_1)$ in the first scale.

   a. if using local modeling

   $$\hat{\mathcal{P}}(X_1) = \sum_{m_1,n_1} \hat{\mathcal{P}}(X_{m_1,n_1}) = \sum_{m_1,n_1} \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_{m_1,n_1}}}{\sigma^2_{Z_1}})$$

   where $\hat{\sigma^2}_{X_{m_1,n_1}}$ can be estimated by Equation 3.29.

   b. if not using local modeling

   $$\hat{\mathcal{P}}(X_1) = \frac{1}{2}\log(1 + \frac{\hat{\sigma^2}_{X_1}}{\sigma^2_{Z_1}})$$

   where $\hat{\sigma^2}_{X_1}$ can be estimated by Equation 3.14.

3. Compute $BITC_{\text{gblur}}$ as follows.

   $$BITC_{\text{gblur}} = \frac{\hat{\mathcal{P}}(D_1)}{\hat{\mathcal{P}}(X_1)}$$

6.4 JPEG2000 Distortion

Blind quality assessment of compression distortion has been of interest for many years, but limited progress has been achieved [103]. This can be noticed by the number of metrics that have been proposed. Until now few quality assessment methods have been proposed and only 2 of them have acceptable performance. The first method is introduced by Sheikh *et al.* [96] who addressed blind quality assessment by inspecting the hidden states of wavelet coefficients in adjacent scales. The other method is proposed by Sazzad *et al.* [90] who assumed the visual quality of JPEG2000 picture is relevant to the edge information. For the existing blind quality metrics, two procedures are always involved in the implementation. First, a training set must be chosen. Usually the training set comprises representative samples which cover all possible situations. Second, a training method will be applied to determine the parameter setting for the quality metric. The selection of a proper method from various candidates is not easy. The training method is selected by carefully considering its strengths and weaknesses. Even if a blind quality metric is implemented with the above two problems properly solved, the metric might not be compatible for a different data set. In [90], Sazzad tried to test his method with two different subjective databases. The parameter settings listed in his paper show significant differences in Table 6.4. This leads to a question of which setting should be used when a JPEG picture is evaluated. It might be one of the reasons that blind quality metrics are not widely used. In this section, we propose a simple but reliable quality evaluation method. As we know, the success of full reference ITC depends on the perceptual information $\mathcal{P}(\underline{D})$ of the distortion. Our blind quality metric ($BITC_{\text{JPEG2000}}$) for JPEG2000 also follows the ITC strategy. $BITC_{\text{JPEG2000}}$ has several advantages. First, it does not need a training procedure. Second, it is validated by four databases from different research groups. Third, the basic idea introduced below can also be applied to blind quality assessment for JPEG compression.

The LIVE database [99]					
γ_1	34.5354	γ_2	37.5732	γ_3	42.9897
γ_4	1.1934	γ_5	6.0552	γ_6	6.3377
γ_7	6.834	γ_8	6.8069	γ_9	0.8304
The TOYAMA database [55]					
γ_1	2.8507	γ_2	-3.4735	γ_3	22.1784
γ_4	2.2957	γ_5	0.0096	γ_6	0.3619
γ_7	-0.3168	γ_8	0.0452	γ_9	2.7841

Table 6.4: The parameters in Sazzad's method. γ_1 ... γ_9 are parameters.

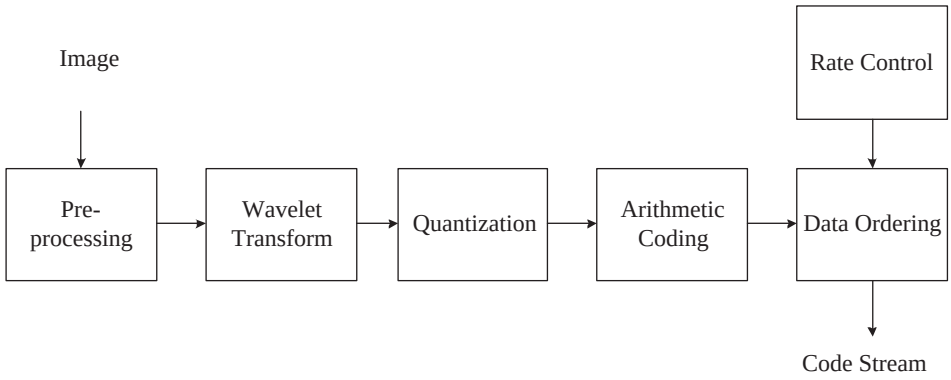


Figure 6.5: JPEG2000 compression framework

6.4.1 Blind Quality Assessment for JPEG2000

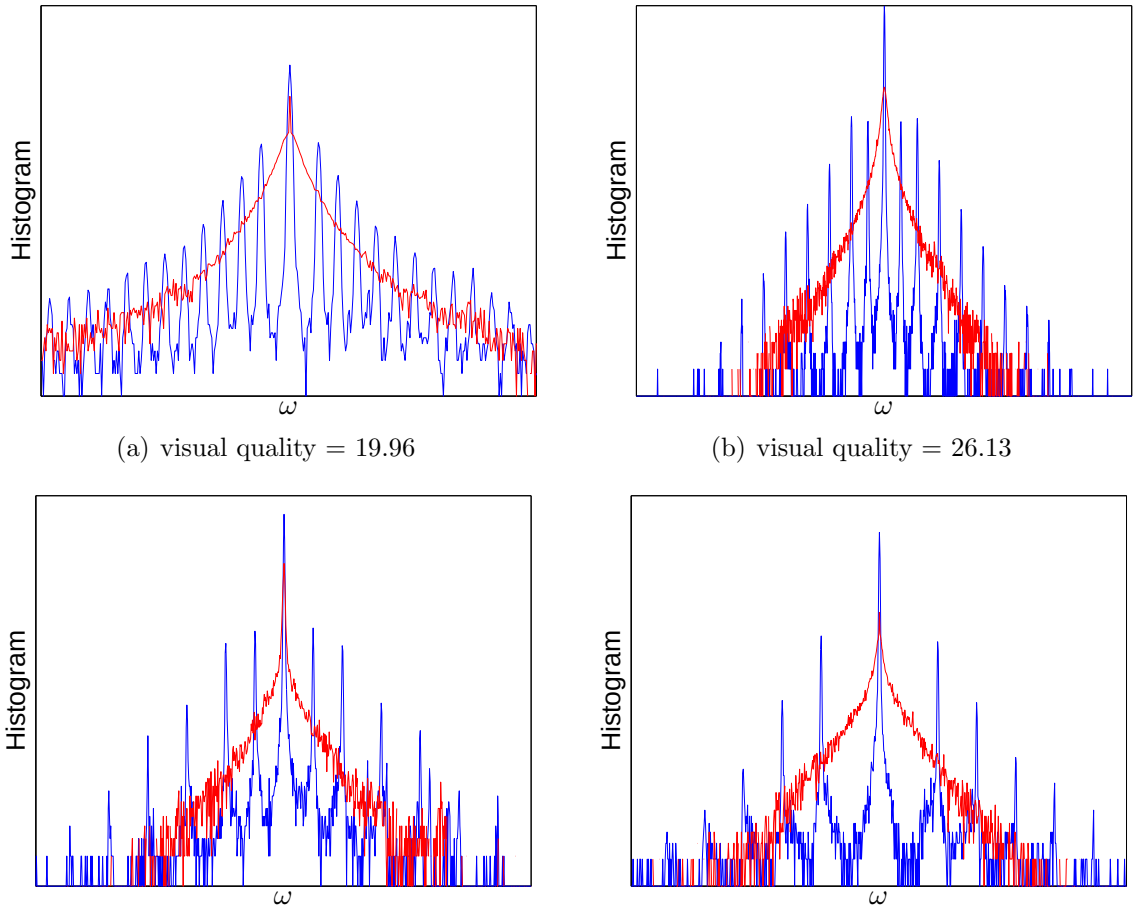
The JPEG2000 compression standard (as shown in Figure 6.5) includes pre-processing (color components transformation, i.e.), wavelet transform (Cohen-Daubechies-Feauveau 9/7 wavelet or Cohen-Daubechies-Feauveau 5/3 wavelet), quantization, arithmetic coding, typical rate control and data packing. Among all operations, quantization is the primary operation that causes the information loss. Hence the key step of blind quality assessment on the JEPG2000 picture is to estimate the lost information during the quantization.

In order to follow the ITC strategy, we need to estimate the quantization distortion in each scale (frequency). But the energy estimation model is not helpful, because the information at each frequency is degraded simultaneously. It is not guaranteed that the low frequency is preserved. It appears that the estimation of difference variance $\sigma^2_{D_i}$ is not practical. But the strong characteristic pattern of compression distortion will be helpful. Figure 6.6 shows the wavelet coefficient histograms of reference images (red line) and distorted (compressed) images (blue line). Each image is transformed to the wavelet domain by Cohen-Daubechies-Feauveau 9/7 wavelet. We notice that the histogram of the distorted (compressed) image shows a strong quantization pattern. Due to computation error, the patterns usually are not ideal, but we are still able to estimate the quantization step for each distorted (compressed) image. The quantization procedure in the ith scale can be written as in Equation 6.20.

$$X_{d_i} = \lfloor \frac{X_i}{\triangle_i} \rfloor \tag{6.20}$$

where $\triangle_i$ is the quantization step in the ith scale.

In order to estimate the perceptual information of the quantization distortion, we



(a) visual quality = 19.96

(b) visual quality = 26.13

(c) visual quality = 37.59

(d) visual quality = 48.84

Figure 6.6: Wavelet coefficients histogram. Each figure plots the wavelet coefficients histograms of reference image (thick line) and distorted image (thin line). Notice that as the image quality gets worse (quality score gets bigger), the quantization step (the distance between histogram peaks) gets bigger.

should investigate the statistics of the quantization distortion. The quantization distortion $D_i = X_i - X_{d_i}$ obeys the uniform distribution in Equation 6.21, if the characteristic function $\Phi_{X_i}(u)$ of the probability density function $f(X_i)$ is bandlimited [133],

i.e. $\Phi_{X_i}(u) = 0$ for $|u| > \frac{\pi}{\triangle}$.

$$f(D_i) = \begin{cases} \dfrac{1}{\triangle_i} & \text{for } |D_i| < \dfrac{1}{2\triangle_i} \\ 0 & \text{else} \end{cases} \tag{6.21}$$

If X_i is assumed to be Gaussian, D_i is approximately uniformly distributed as shown in Figure 6.7(a), as the ratio of $\triangle$ over σ varies from 0.1 to 0.6. The quantization distortion (of a Gaussian signal) is almost uniformly distributed within $[-\triangle_i/2, \triangle_i/2]$. The quantization distortion of the actual image wavelet coefficient is more concentrated at zero[2]. When the quantization step $\triangle_i$ is small, the probability density of the quantization distortion is close to a uniform distribution. It is reasonable to

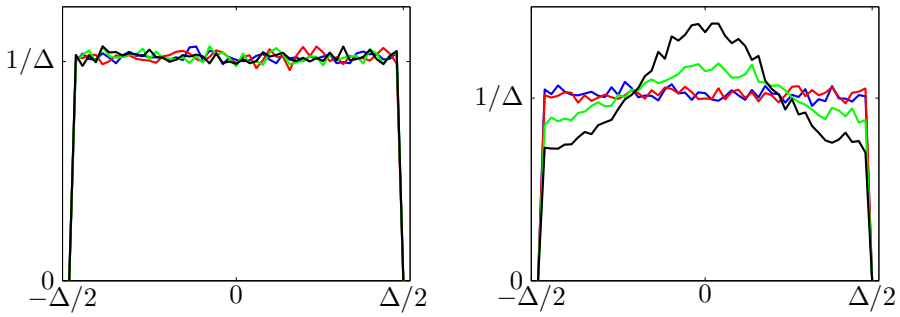


Figure 6.7: The probability density of quantization noise. Figure (a) and (b) display the probability density of the quantization noise. In Figure (a), the quantization is applied on a Gaussian source X. In Figure (b), the quantization is applied on the actual image wavelet coefficients X. The blue line: $\triangle/\sigma_X = 0.1$. The red line: $\triangle/\sigma_X = 0.2$. The green line: $\triangle/\sigma_X = 0.4$. The black line: $\triangle/\sigma_X = 0.6$.

assume that the quantization distortion in the ith scale is approximately uniformly distributed, then $\sigma^2_{D_i} = \frac{1}{12}\triangle_i^2$. According to the expression of perceptual information in Table 4.2, the perceptual difference in the ith scale can be approximately computed

[2]In the wavelet domain, many image coefficients are very close to zero. This results in the peak of the probability density of wavelet coefficients around zero, which means the large bandwidth of $\Phi_{X_i}(u)$. To ensure $\Phi_{X_i}(u) = 0$ for $|u| > \frac{\pi}{\triangle}$, the quantization step $\triangle$ should be small. In other words, if the quantization step $\triangle$ gets smaller, the quantization distortion becomes more uniformly distributed.

as Equation 6.22[3]

$$\mathcal{P}(D_i) \simeq \frac{1}{2}\log(1+\frac{\sigma_{D_i}^2}{\sigma_{Z_i}^2}) \tag{6.22}$$

$$= \frac{1}{2}\log(1+\frac{\triangle_i^2}{12\sigma_{Z_i}^2}) \tag{6.23}$$

In Chapter 5, the normalization step in ITC is designed to eliminate the impact of varied image content. To follow the ITC strategy, a simple method proposed next accomplishes the normalization. The basic idea is to derive the quantization distortion if X_i obeys the standard normal distribution, then normalization is unnecessary. As shown in Equation 6.24, the equivalent quantization step $\bar{\triangle}$ is derived. It is not difficult to derive $\bar{\triangle}$ with a simple search table.

$$\bar{\triangle}_i = \{\triangle | Pr(X_{d_i}=0) = \int_{-\triangle/2}^{\triangle/2} \frac{1}{\sqrt{2\pi}}\exp(\frac{\tau^2}{1})d\tau\} \tag{6.24}$$

The perceptual information of the "normalized" quantization distortion $\bar{D}_i$ is given in Equation 6.25. According to the ITC strategy in Equation 6.1, the visual quality can be defined in Equation 6.26 by considering the perceptual distortion in all scales.

$$\mathcal{P}(\bar{D}_i) = \frac{1}{2}\log(1+\frac{\bar{\triangle}_i^2}{12\sigma_{Z_i}^2}) \tag{6.25}$$

$$BITC_{\text{JPEG2000}} = \sum_i \mathcal{P}(\bar{D}_i) \tag{6.26}$$

The objective quality evaluation by $BITC_{\text{JPEG2000}}$ (Equation 6.26) has Spearman correlation of 0.86 with human judgment according to the LIVE database [99]. Since $BITC_{\text{JPEG2000}}$ is the linear summation of the perceptual difference in all scales, the performance will be degraded if the perceptual difference in certain scale is not consistent with human judgment. In other words, if $BITC_{\text{JPEG2000}}$ is defined within a scale, better performance might be achieved, as shown in Equation 6.27.

$$BITC_{\text{JPEG2000}} = \mathcal{P}(\bar{D}_i) = \frac{1}{2}\log(1+\frac{\bar{\triangle}_i^2}{12\sigma_{Z_i}^2}) \tag{6.27}$$

[3]Equation 6.22 computes the perceptual information of D_i if D_i obeys the Gaussian distribution, which is emphasized in Chapter 4. There is no closed form to compute the perceptual information of an uniform distribution, but Equation 6.22 is a good approximation. Please refer to Appendix H to check the approximation accuracy.

In practice, we find that better performance can be achieved by the following definition, where the $\bar{\triangle}^2$ is substituted by $\triangle\bar{\triangle}$. From the test result, Spearman correlation of $BITC_{\text{JPEG2000}}$ by Equation 6.27 is 0.89 when image statistics of the second scale in the diagonal direction is used. Spearman correlation of $BITC_{\text{JPEG2000}}$ by Equation 6.28 is 0.94. In the next section, we will validate $BITC_{JPEG2000}$ (Equation 6.28) thoroughly.

$$BITC_{\text{JPEG2000}} = \frac{1}{2}\log(1 + \frac{\triangle_i\bar{\triangle}_i}{12\sigma^2_{Z_i}}) \tag{6.28}$$

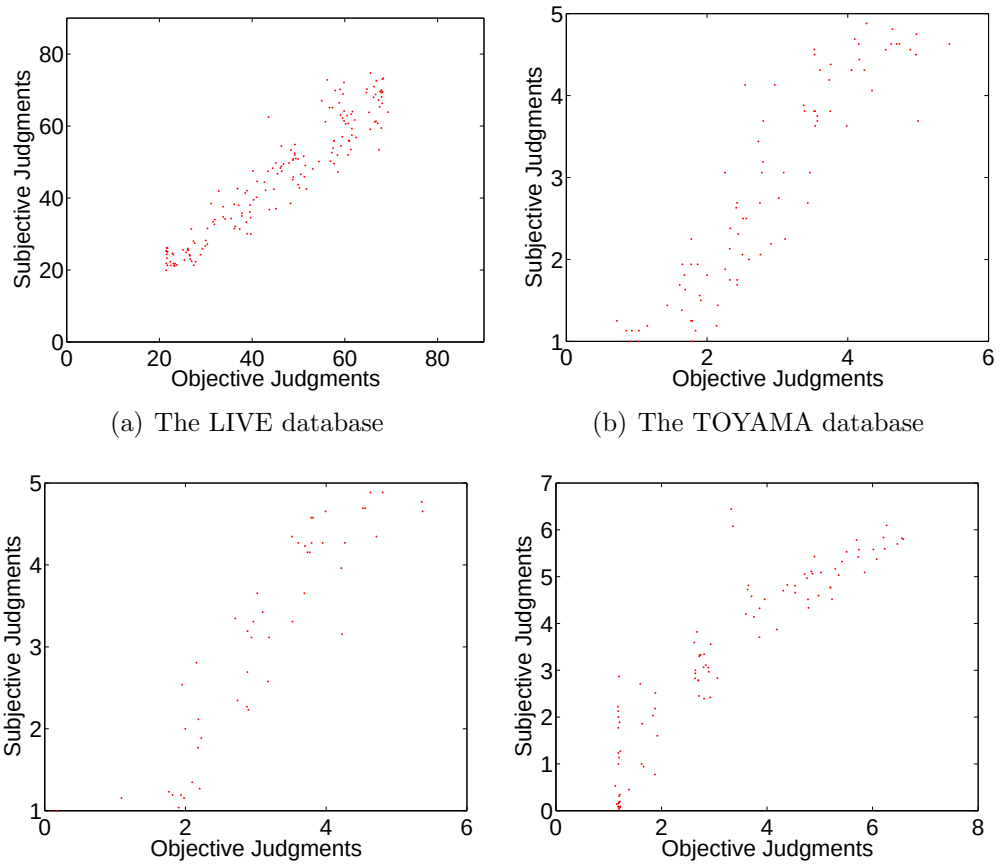
6.4.2 Implementation of $BITC_{\text{JPEG2000}}$

---

1. Transform the test image into the db97 wavelet domain.
2. Derive the quantization step $\triangle_i$ in the ith scale. A simple method (Matlab code) is given in Appendix H.
3. Compute the equivalent quantization step $\bar{\triangle}_i$ with Equation 6.24. A fast method to derive the equivalent quantization step is to use a search table.
4. Compute the visual quality $BITC_{\text{JPEG2000}}$ with Equation 6.28.

6.4.3 Performance and Discussion

As we discussed above, the blind quality metric for JPEG2000 is computed using the wavelet coefficients within certain subband. Now let us test the performance of $BITC_{\text{JPEG2000}}$ when different subbands are applied, i.e. first, second and third scales on three directions. Figure 6.9 displays the linear correlation between the subjective and objective judgments. The test results show that $BITC_{\text{JPEG2000}}$ generates consistent evaluation using the wavelet coefficients in the second scale. Since the number of the wavelet coefficients at low frequency scale are limited (for a 512×512, the size of a subband in the third scale is only 64×64), it might cause an estimation error. Also for small JPEG2000 distortion, the quantization pattern at low frequency scale is not easily recognized especially considering the roundoff error, resulting in the degradation of the $BITC_{\text{JPEG2000}}$ performance. On the other hand, wavelet coefficients in the high frequency scales are converted to zero when JPEG2000 distortion is distinct. It is not possible to estimate the quantization pattern when most coefficients are quantized to zero. Also if the texture is not rich in that scale, it might also be difficult to identity



(a) The LIVE database

(b) The TOYAMA database

(c) The IVC database

(d) The TID2008 database

Figure 6.8: The performance of blind quality metric on JPEG2000 pictures. The linear correlation between the subjective and objective judgements for the LIVE database [99] is 0.9241. The linear correlation for the TOYAMA database [55] is 0.9104. The linear correlation for the IVC database [64] is 0.9045. The linear correlation for the TID2008 database [81] is 0.9125.

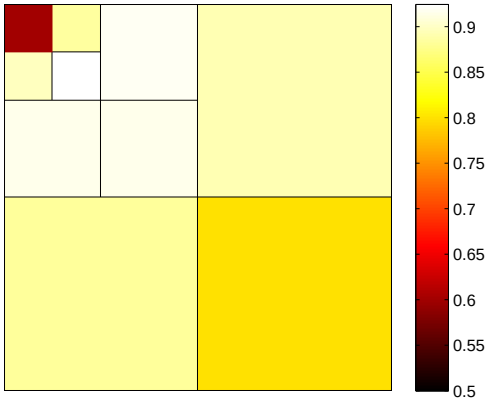


Figure 6.9: JPEG2000 blind quality metric using different subbands. The largest linear correlation is 0.9241 in the diagonal direction in the third scale. The corresponding Spearman correlation is 0.9363. The experiment is conducted with the LIVE database [99].

the quantization pattern.

In the following section, $BITC_{\text{JPEG2000}}$ is compared with two existing methods [89,98]. Sazzad *et al.* [89] considered the image quality as the combination of several image features, as shown in Equation 6.29.

$$MOSp = \frac{4}{1+e^{-1.0217(C-3)}} + 1 \tag{6.29}$$

where

$$\begin{aligned} C = & \{\gamma_1 log(S+1) + \gamma_2 log(A+1) + \gamma_3 log(Z+\gamma_4)\} \times \\ & \{\gamma_5 log(H_f+1) + \gamma_6 log(V_f+1) + \gamma_7 log(H+1) + \gamma_8 log(V+1) + \gamma_9\} \end{aligned} \tag{6.30}$$

$S, A, Z, H_f, V_F H$ and V are the image features defined in [89], and $\gamma_{1,2,\ldots,9}$ are the parameters derived by training. Compared with our method, it is clear that Sazzad used many more parameters (10 parameters). In order to determine the parameters, a training set and method is necessary. The parameters are easily overtrained because

too many parameters are used compared with size of the training set (60 samples were used for training [89]). The other problem is that Sazzad's method is sensitive to the choice of the training set. In order to achieve good performance, Sazzad chose the training set composed by the images with the best quality or worst quality. This setting results in such a method which performs pretty well when dealing with images containing small distortion or large distortion but much worse for medium distortion.

Sheikh *et al.* [98] stated that the probability of hidden states gives the indication of the quality loss, using the hidden Markov tree model [29]. He estimated image quality by converting the probability $p_i(ss)$ (the hidden states of parent and child are both small) into image features in Equation 6.31.

$$q_i = K_i \left[1 - \exp\left(-\frac{p_i(ss) - u_i}{T_i} \right) \right] \tag{6.31}$$

where K_i, u_i and T_i are the parameters in the ith scale. The image features in different scales are combined to derive the quality feature Q in Equation 6.32.

$$Q = \sum_{i=v,h,d} k_i q_i \tag{6.32}$$

where k_v, k_h and k_d are the parameters for different directions in different directions. Finally the visual quality score is given in Equation 6.33.

$$IQ = \beta_1 log_{\beta_2}(Q - \beta_3) + \beta_4 x + \beta_5 \tag{6.33}$$

where $\beta_{1,2...,5}$ are the parameters derived by training, and Q is defined by the image features in wavelet domain. Sheikh is the first to approach blind quality assessment by using inter-scale consistency under multi-resolution framework. But his method has the same problem as Sazzad's method: it is sensitive to the training set.

Based on the above description, our method involves very limited parameters compared with other existing method. Our method does not achieve the parameters by training. Also our method does not rely on the feature extraction and highly non-linear feature combinations which leaks the proper theoretical foundations. Table 6.5 compares the performances of three blind quality metrics for JPEG2000 distortion, which is plotted in Figure 6.10. For display purposes, we normalize the performances. For instance, if the linear correlations of the objective evaluation by three methods are 0.9, 0.6 and 0.3 separately. Then the normalized correlations are 1, 0.67 and 0.33. It is clear that our $BITC_{\text{JPEG2000}}$ method has the best consistency with the

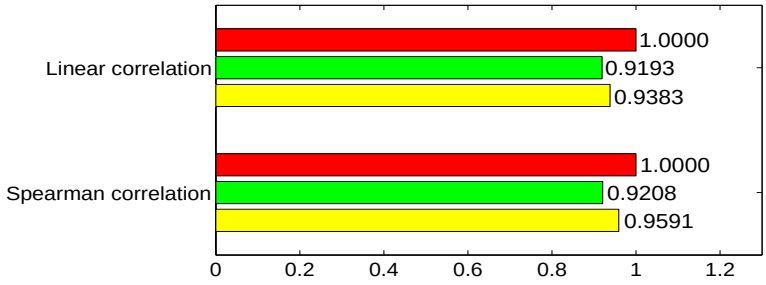
Database	$BITC_{\text{JPEG2000}}$		Sheikh's method		Sazzad's method	
	CC	SC	CC	SC	CC	SC
LIVE [99]	**0.9474**	**0.9363**	0.8709	0.8621	0.8889	0.898
Toyama [55]	**0.9104**	**0.9203**	0.7622	0.7799	0.8258	0.8169
IVC [64]	**0.9045**	**0.9305**	0.8493	0.8567	0.7290	0.7922
TID2008 [81]	**0.9125**	**0.9183**	0.3965	0.4933	0.8717	0.8750

Table 6.5: JPEG 2000 blind quality assessment metrics comparison."CC": linear correlation between subjective and objective judgments. "SC": Spearman correlation. The bold numbers indicate the best performances.

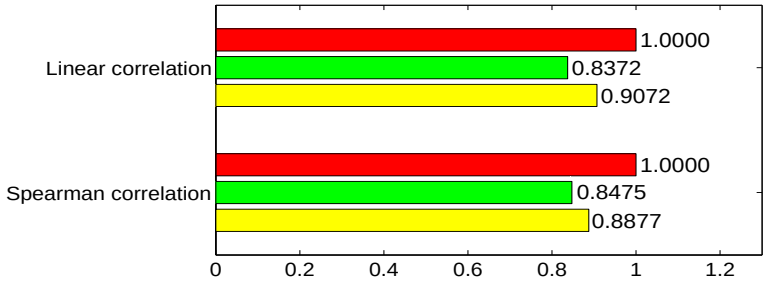
human judgments(largest linear correlation and Spearman correlation). Meanwhile, the correlation is greater than 0.9 for any database. It means the visual quality score by $BITC_{JPEG2000}$ does reflect the human opinion. Furthermore, the computation complexity of $BITC_{\text{JPEG2000}}$ is less than 1/4 of Sheikh's method and 1/6 of Sazzad's method. Last but not least, blind ITC can be easily embedded into JPEG2000 codec. After the JPEG2000 encoder computes the wavelet coefficients, $BITC_{\text{JPEG2000}}$ can take over the coefficients and estimate the image quality. With the wavelet coefficients available, $BITC_{\text{JPEG2000}}$ needs less than 5% of the computation complexity of any existing quality metrics.

From the above description of the blind quality assessment method, our proposed method has several advantages:

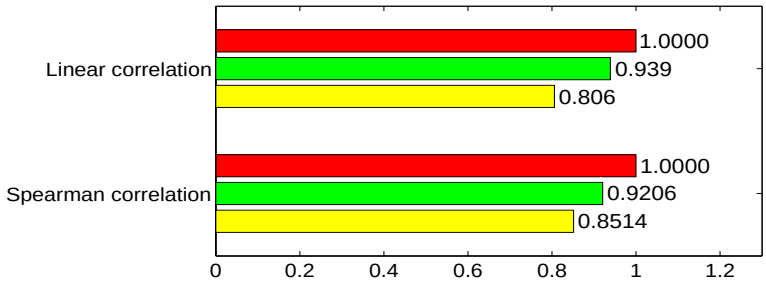
- The proposed method is based on distortion estimation, which has a theoretic foundation. Since the distortion pattern (quantization pattern) is strong, the estimation of quantization step $\triangle$ is easy to design.
- From the performance of the proposed method, we notice that the proposed metric exhibits excellent consistency with human subjective judgments and compatibility to different subjective databases. Also the proposed method does not require training data or training methods.
- The proposed method is fast and suitable for real-time application.
- The proposed method can be embedded into JPEG2000 codec and estimate the quality of the compressed image instantly.



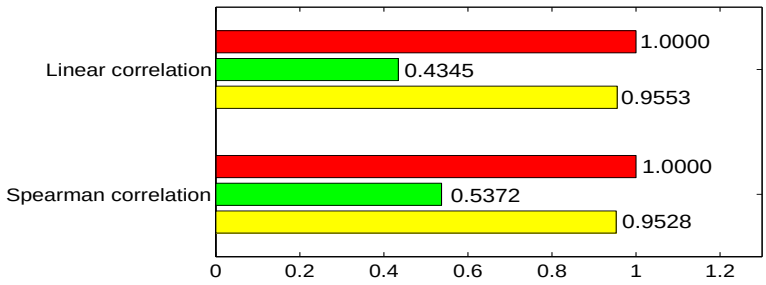
(a) LIVE database



(b) Toyama database



(c) IVC database



(d) TID2008 database

Figure 6.10: JPEG2000 blind quality assessment metrics comparison (continued). ■ *ITC* metric, ■ Sheikh's method, ■ Sazzad's method.

6.5 JPEG Distortion

6.5.1 Blind Quality Assessment on JPEG

Although JPEG distortion is visually different from JPEG2000 distortion, both compression standards share the same principle. The JPEG and JPEG2000 distortions are both due to information loss from quantization. This makes it possible to use the idea in the last section to approach blind quality assessment for JPEG image. The major difference between the two standards is the transform. The JPEG standard uses discrete cosine transform (DCT), whereas the JPEG2000 standard uses Cohen-Daubechies-Feauveau 9/7 wavelet or Cohen-Daubechies-Feauveau 5/3 wavelet. In the JPEG2000 standard, the quantization steps vary across different subbands. The quantization step in a certain subband is estimated on the entire subband. In the JPEG standard, an image is divided into many micro-blocks. Each 8×8 block is converted to a frequency-domain representation which also has size of 8×8, as shown in Equation 6.34.

$$\begin{aligned} X(u,v) &= \alpha(u)\alpha(v)\sum_{m=0}^{7}\sum_{n=0}^{7} I(m,n)\cos\left[\frac{\pi}{8}\left(m+\frac{1}{2}\right)u\right]\cos\left[\frac{\pi}{8}\left(n+\frac{1}{2}\right)v\right] \\ \alpha(n) &= \begin{cases} \sqrt{\frac{1}{8}} & \text{if } n=0 \\ \sqrt{\frac{2}{8}} & \text{otherwise} \end{cases} \end{aligned} \tag{6.34}$$

where u $(0 \le u < 8)$ is the horizontal spatial frequency, v $(0 \le v < 8)$ is the vertical spatial frequency and $I(m,n)$ is the pixel value at coordinates (m,n). The quantized DCT coefficients are computed as follows.

$$X_d(u,v) = \lfloor \frac{X(u,v)}{Q_0(u,v)} \rfloor \tag{6.35}$$

According to the JPEG standard, the quantization matrix $Q_0(u,v)$ is specified and shown in Equation 6.36. In practice, the quality of JPEG images is controlled by the scaled quantization matrix, i.e. $Q(u,v) = k \times Q_0(u,v)$, where k adjusts image quality. The quantization step cannot be simply derived from Equation 6.36. For a JPEG image, the quantization step at certain frequency $(u\frac{\pi}{8}, v\frac{\pi}{8})$ is derived with the same

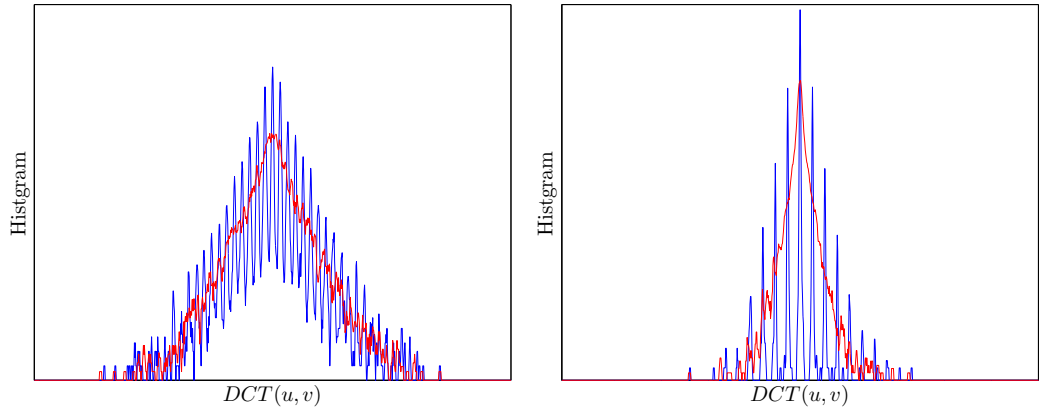
method from the last section.

$$Q_0(u,v) = \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} \overset{u}{\longrightarrow} \\ \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \end{matrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix} \Big\downarrow v \tag{6.36}$$

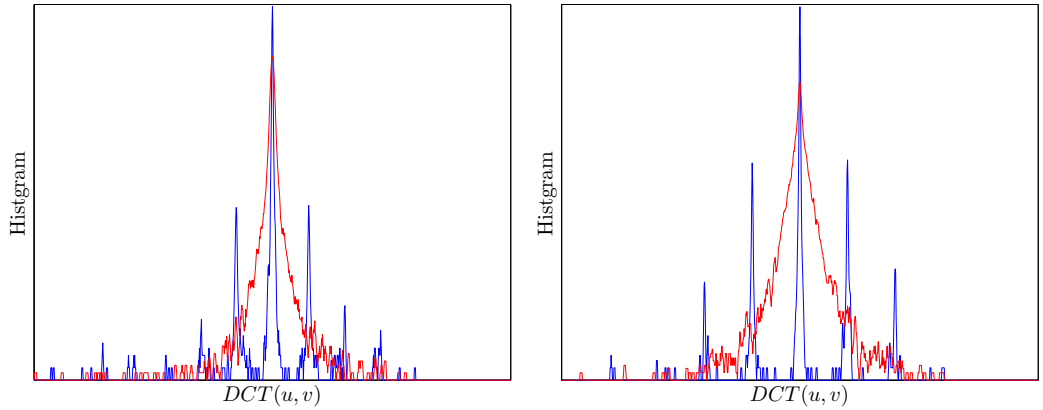
Before proposing the blind quality metric for JPEG image, consider the quantization pattern at the frequency (u_0, v_0). The histogram in Figure 6.11 is generated by the coefficients at frequency $(u_0 = 3, v_0 = 3)$ for all micro-blocks in a JPEG image. Small quantization step usually means good visual quality, and large quantization step results in poor visual quality. We notice that JPEG has a similar distortion pattern as JPEG2000. Hence we can use the same quality assessment method in the last section for JPEG distortion evaluation and derive a similar blind quality assessment metric for JPEG pictures, as shown in Equation 6.37, where $\triangle_{u,v}$ is the actual quantization step at the frequency $(u\frac{\pi}{8}, v\frac{\pi}{8})$ and the equivalent quantization step $\bar{\triangle}_{u,v}$ is derived in Equation 6.38. $X_d(u,v)$ in Equation 6.38 is the DCT coefficient of the JPEG image at frequency $(u\frac{\pi}{8}, v\frac{\pi}{8})$.

$$BITC_{\text{JPEG}} = \frac{1}{2}\log(1 + \frac{\triangle_{u,v}\bar{\triangle}_{u,v}}{\sigma^2_{Z_{u,v}}}) \tag{6.37}$$

$$\bar{\triangle}_{u,v} = \{\triangle | Pr\big(X_d(u,v) = 0\big) = \int_{-\triangle/2}^{\triangle/2} \frac{1}{\sqrt{2\pi}} \exp(-\frac{\tau^2}{2}) d\tau\} \tag{6.38}$$



(a) visual quality = 26.53 (b) visual quality = 29.32



(c) visual quality = 42.52 (d) visual quality = 52.68

Figure 6.11: DCT coefficients histogram. Each figure plots the DCT coefficients histograms of reference image (thick line) and distorted image (thin line). Notice that as the image quality gets worse (quality score gets bigger), the quantization step (the distance between histogram peaks) gets bigger.

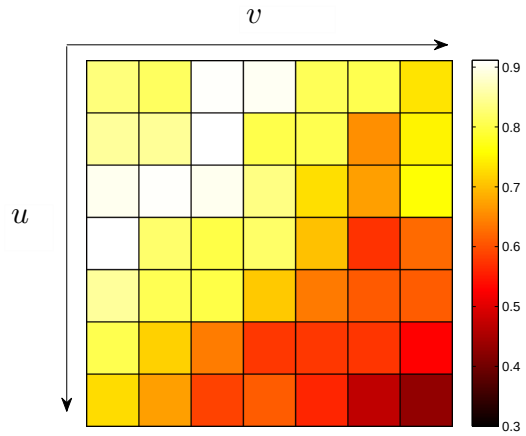


Figure 6.12: The performance of JPEG blind quality metric using DCT coefficients at $(\frac{\pi}{8}u, \frac{\pi}{8}v)$. The figure illustrates the linear correlation between the subjective and objective judgments using DCT coefficients at $(\frac{\pi}{8}u, \frac{\pi}{8}v)$, where $u, v = \{0, 1, 2, 3, 4, 5, 6, 7\}$. The maximum linear correlation is 0.9313 at the frequency of $u = 2, v = 2$

.

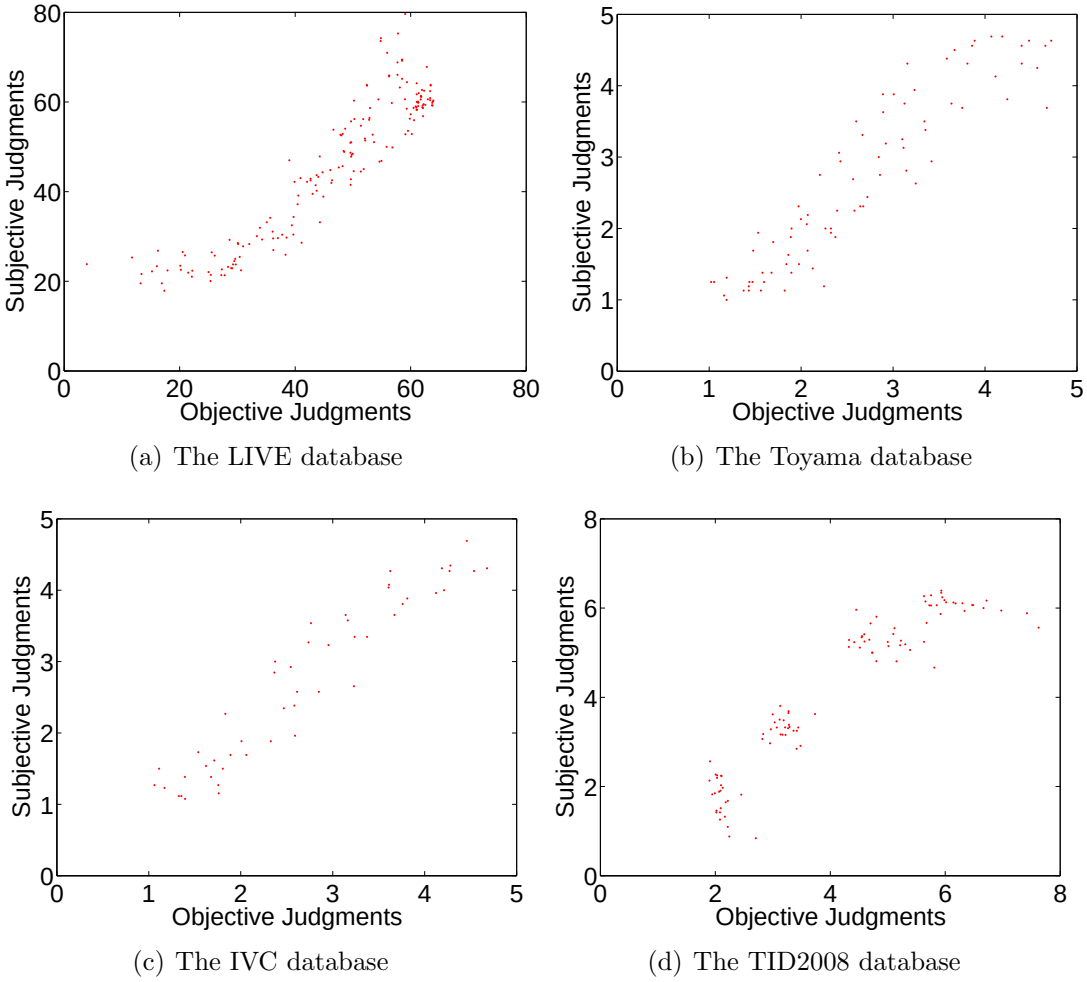
6.5.2 Implementation of $BITC_{\mathbf{JPEG}}$

---

1. Transform the test image into the DCT domain.
2. Derive the quantization step $\triangle_{u,v}$ at the frequency $(u\frac{\pi}{8}, v\frac{\pi}{8})$. A simple method (Matlab code) is given in Appendix H.
3. Compute the equivalent quantization step $\bar{\triangle}_{u,v}$ with Equation 6.38.
4. Compute the visual quality $BITC_{\text{JPEG}}$ with Equation 6.37.

6.5.3 Performance and Discussion

From the above description, we defined the blind quality metric using the statistics of DCT coefficients at $(u\frac{\pi}{8}, v\frac{\pi}{8})$. Figure 6.12 displays the linear correlation between objective and subjective judgments using DCT coefficients at different frequencies. During our test, we use the DCT coefficients at the frequencies $u, v = \{0, 1, 2, 3, 4, 5, 6, 7\}$. The best linear correlation with the human judgment is achieved at the frequency of $u = 2, v = 2$. Also we notice that blind quality metric provides consistent evaluation



(a) The LIVE database

(b) The Toyama database

(c) The IVC database

(d) The TID2008 database

Figure 6.13: The performance of blind quality assessment for JPEG distortion. The linear correlation between the prediction results and subjective judgments is 0.9089 for the LIVE database [99]. The linear correlation between the prediction results and subjective judgments is 0.8659 for the Toyama database [55]. The linear correlation between the prediction results and subjective judgments is 0.9275 for the IVC database [64]. The linear correlation between the prediction results and subjective judgments is 0.9353 for the TID2008 database [81].

when $u + v \simeq 4$, the white band as shown in Figure 6.12. The blind quality metric also performs badly at high frequency. The reason is that most DCT coefficients are quantized to zeros when large JPEG distortion happens and it is difficult to catch the quantization pattern.

The performance of $BITC_{\text{JPEG}}$ is plotted in Figure 6.13 and three methods are compared in Table 6.6. Our method shows excellent compatibility with four different subjective databases. The correlation is greater than 0.9 for any database. It means the visual quality score by $BITC_{JPEG}$ does reflect the human opinion.

In the following section, $BITC_{\text{JPEG}}$ is compared with two existing methods [89, 124]. Wang's method [124] defined the blind quality of an JPEG image as follows:

$$S = \alpha + \beta B^{\gamma_1} A^{\gamma_2} Z^{\gamma_3} \tag{6.39}$$

where α, β, γ_1, γ_2 and γ_3 are the model parameters that must be estimated with the subjective test data. A (pixel activity), B (block discontinuity) and Z (zero cross ratio) are the image features defined in [124]. Sazzad applied the same method [89] in the last section to evaluate JPEG pictures. As with the blind metric for JPEG2000, our proposed method does not require complicated feature extractions, combinations or training to accomplish the quality evaluation. Table 6.6 lists the performances of three methods. Apparently, the proposed $BITC_{\text{JPEG}}$ has the best linear correlation and Spearman correlation, while Sazzad's method does not support Toyama database well. For a better illustration, the test results are normalized and plotted in Figure 6.14.

The proposed method also does not need to decode the compressed image. Our method takes the transform coefficients as the input. Furthermore our method only needs one frequency coefficient out of 64 coefficients in a microblock. Our method also needs not to extract the features, such as (pixel activity or block discontinuity), in the spatial domain. These makes our method very fast. $BITC_{\text{JPEG}}$ only needs 1/4 time of Wang's method and less than 1/10 of Sazzad's method if the input is the decoded JPEG image in the spatial domain. To evaluate images in the JPEG format, $BITC_{\text{JPEG}}$ needs only 5% of the above computation time, since the DCT coefficients are accessible. $BITC_{\text{JPEG}}$ method can be embedded into the JPEG codec easily. When the JPEG encoder computes the DCT coefficients, $BITC_{\text{JPEG}}$ can take over the coefficients and estimate the image quality.

Compared with existing blind quality metrics, our method has three advantages:

Database	$BITC_{\text{JPEG}}$		Wang's method		Sazzad's method	
	CC	SC	CC	SC	CC	SC
LIVE [99]	**0.9206**	**0.9065**	0.8228	0.901	0.8892	0.8962
Toyama [55]	**0.9211**	**0.9403**	0.8772	0.8942	0.7193	0.7648
IVC [64]	**0.9494**	**0.9561**	0.9229	0.9333	0.9354	0.9472
TID2008 [81]	**0.9353**	0.9145	0.9308	**0.9159**	0.3182	0.5318

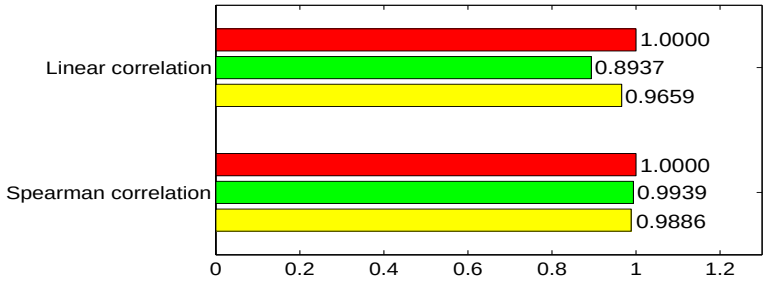
Table 6.6: JPEG blind quality assessment metrics comparison. "CC": linear correlation between subjective and objective judgments. The bold numbers indicate the best performances.

- The proposed method is based on distortion estimation, which is simple and efficient.
- The method can be applied to different databases (the LIVE [99], TOYAMA [55], IVC [64] and TID2008 [81] databases).
- The method does not rely on complicated nonlinear mapping from image features to image quality.
- The proposed method is fast and suitable for real-time application.
- The method can be embedded into JPEG codec for quality estimation.

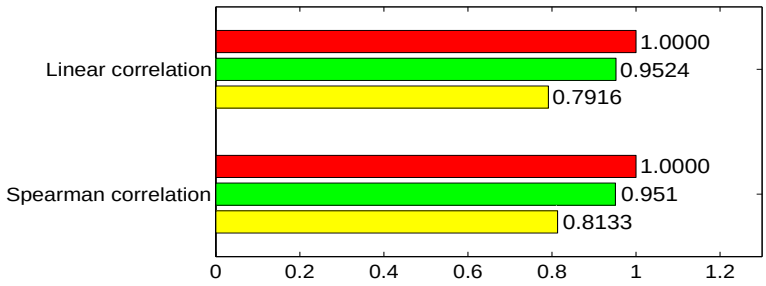
6.6 Fastfading Distortion

6.6.1 Blind Quality Assessment for Fastfading Distortion

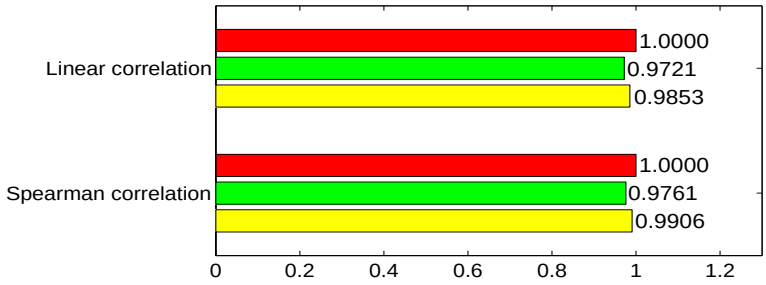
As shown in Figure 6.15, fastfading distortion will happen during source encoding and wireless transmission. The information loss during source coding is mainly quantization distortion. The information loss during wireless transmission is due to the uncertainty in the communication channel. The result is a difficulty in predicting the lost information without a reference image. Figures 6.16 displays four images with fastfading distortion. If we take a close look at the examples in Figures 6.16, we notice that fastfading distortion is quite different from the other four distortions, since it exhibits different distortion behaviors. Basically, the distortion behaviors can be summarized into three categories: *shift*, *elimination* and *quantization*.



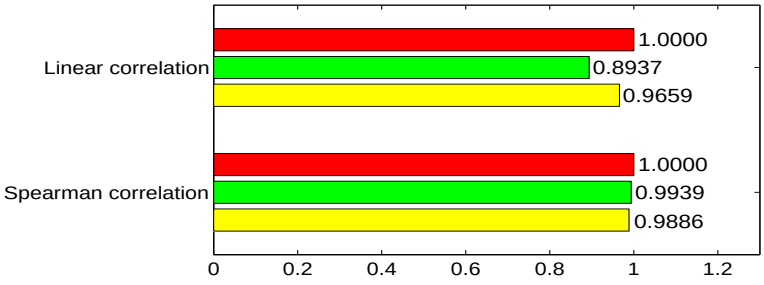
(a) LIVE database



(b) Toyama database



(c) IVC database



(d) TID2008 database

Figure 6.14: JPEG blind quality assessment metrics comparison (continued). ■ *ITC* metric, ■ Zhou's method, ■ Sazzad's method.

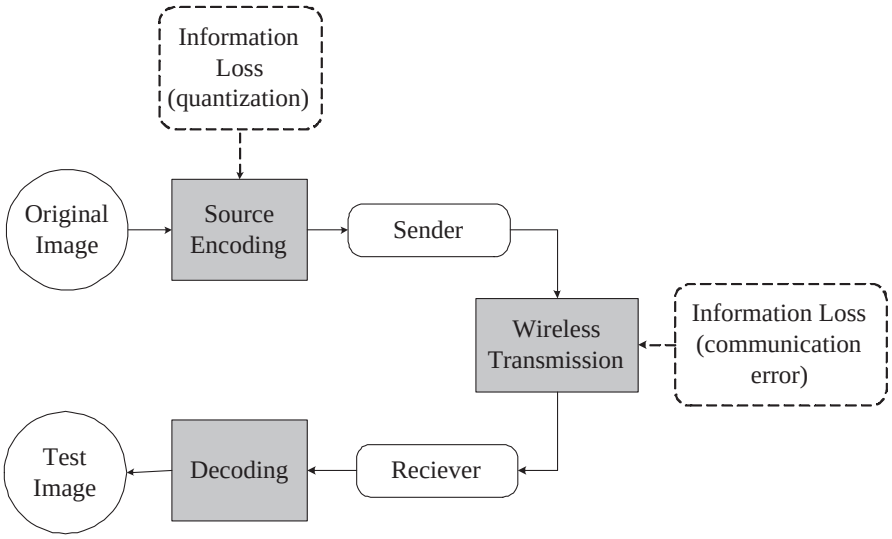


Figure 6.15: Image transmission

- ***Shift***: the image coefficients are shifted away from where they are supposed to be.
- ***Elimination***: the image coefficients are "wiped" out, due to the data loss in the transmission.
- ***Quantization***: the image details are discarded due to the quantization procedure during the compression. Since the original image is compressed first and sent through the wireless channel, the quantization distortion is always a part of the fastfading distortion.

Figure 6.17 displays three examples that exhibit the above three distortion behaviors. Compared to the reference in Figure 6.17(d), Figure 6.17(a) shows the shift behavior. The roof edge is duplicated and shifted to the left. Figure 6.17(b) shows the elimination behavior. The difference on the top left is smaller, but larger on the rest part (information is eliminated). Figure 6.17(c) shows the quantization behavior, which is not easily recognized in the spatial domain. Figure 6.18 displays the above distorted images and their reference in the wavelet domain and clearly illustrates the distortion behaviors. Figure 6.18(a) shows that the shift behavior is caused by the coefficient shift in wavelet domain. In Figure 6.18(b), the coefficients other than the top left are eliminated. Most structures in Figure 6.18(c) are preserved but detail coefficients are discarded. This suggests that the distortion behavior is quantization. Unlike the other

(a) plane

(b) student sculpture

(c) ocean

(d) flowers

Figure 6.16: Image samples with fastfading distortion. In Figure (d), the distortion happens on the top of the buildings (the white spots). Other distortions in Figure (d) becomes less noticeable due to the resizing and printing.

four distortions (white noise, Gaussian blur, JPEG2000 and JPEG distortion), fastfading distortion might not exhibit single distortion behavior. If we watch closely on Figure 6.18(c), there is a nearly blank block (elimination behavior) on the roof. Figure 6.18(b) shows that the distortion behaviors might only happen on certain places.

From the above discussion, we summarize the characteristics of fastfading distortion as below.

1. Fastfading distortion is the combination of three behaviors: shift, elimination and quantization.

2. Fastfading distortion may happen locally (the shift or elimination on certain places) or globally (the quantization on the entire wavelet scale). The other four distortions (white noise, Gaussian blur, JPEG2000 and JPEG distortion) are usually applied to the entire image.

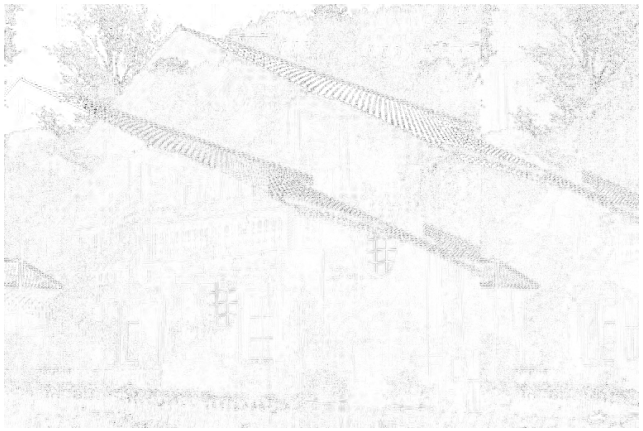
In the following section, we will follow the ITC strategy to estimate visual quality of images with fastfading distortion. As shown in Figure 6.15, the image is first compressed according to the JPEG2000 standard. The quantization distortion is applied to the entire image. Then the compressed image is transmitted through the wireless communication channel where the shift or elimination might happen. According to the above quality degradation procedure, our quality metric is designed as follows.

- Evaluate quantization distortion by the method proposed in blind quality assessment for JPEG2000.

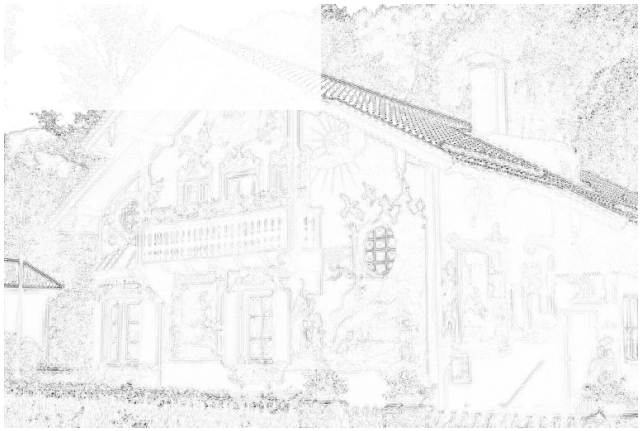
  As discussed in Chapter 6 Section 3, the perceptual information of the JPEG2000 quantization distortion in the ith scale is given in Equation 6.40 (from Equation 6.23 in Chapter 6 Section 3), where $D_{i\ \mathrm{Q}}$ denotes the quantization distortion in the ith scale.

$$\mathcal{P}(D_{i\ \mathrm{Q}}) = \frac{1}{2}\log\left(1 + \frac{\triangle_i^2}{12\sigma_{Z_i}^2}\right) \tag{6.40}$$

- Identify if the current neighborhood is degraded by shift or elimination distortion. Since shit and elimination distortion may only happen on certain positions, it is proper to divide each scales into non-overlapped neighborhoods and identify shift or elimination distortion on each neighborhoods.



(a) Shift behavior

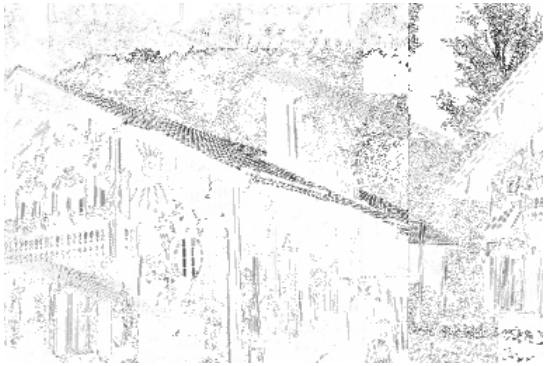


(b) Elimination behavior

(c) Quantization behavior

(d) Reference image

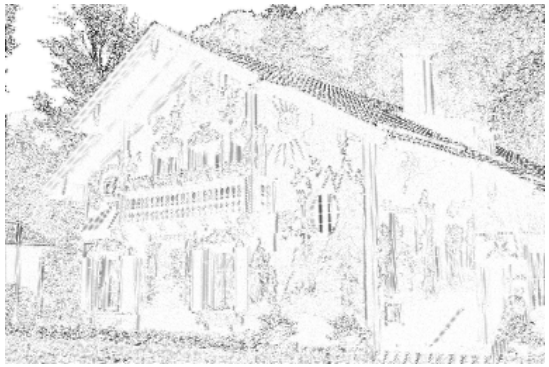
Figure 6.17: Fastfading distortion behaviors. Figure (a,b,c) show the absolute difference between the distorted images and their reference. Figure (a) shows the shift behavior. Compared to the reference in Figure (d), the roof edge is duplicated and shifted to the left. Figure (b) shows the elimination behavior. Compared to its reference, the difference on the top left is smaller, but larger on the rest part (information is eliminated). Figure (c) shows the quantization behavior, which is not easily recognized just from the image difference in the spatial domain.



(a) Shift behavior

(b) Elimination behavior

(c) Quantization behavior



(d) Reference image

Figure 6.18: Fastfading distortion behaviors in wavelet domain. The figures display the coefficients of the distorted images and their reference in the horizontal direction of the first scale in wavelet domain. Figure (a) shows the coefficients circular shift to the left. Figure (b) shows the coefficients other than the top left are eliminated. Figure (c) shows most structures are preserved but detail coefficients are discarded.

(a) Shift behavior

(b) Elimination behavior

(c) Quantization behavior

(d) Reference

Figure 6.19: Fastfading distortion behaviors in the wavelet domain (continued). The figures display the coefficients of the distorted images and their reference in the horizontal direction of the first two scales in wavelet domain. Compared with the second scale, it is easier to notice the distortion behaviors. In Figure (a), the coefficients in the first scale are circular shifted to the left. Figure (b) shows the coefficients in the first scale other than the top left are eliminated. Figure (c) shows most structures are preserved in both scales are kept but detail coefficients are discarded.

(a) Shift behavior (b) Elimination behavior

(c) Quantization behavior (d) Reference

Figure 6.20: 2-D histogram of local energy pair $\{\mathcal{E}_1(m_1, n_1), \mathcal{E}_2(m_2, n_2)\}$ in the fist two scales. In Figure (a), the local energy pair spreads widely, since the dependency of the first two scales is destroyed by the shift behavior. In Figure (b), the local energy pair spread to the left, since the coefficients (no matter the details or structures) in the first scale are eliminated. In Figure (c), a small portion of the local energy pairs line up, since the detail coefficients are quantized to zero.

(a) Chapter 3 Section 3 proposed a method to identify whether the wavelet secondary property is damaged (Page 55-57). If the linear correlation $\rho(m_i, n_i; m_{i+1}, n_{i+1})$ (Equation 3.25) between two energy set $\{\mathcal{E}_{i+1}(m_{i+1}, n_{i+1})\}$ and $\{\mathcal{E}_i(m_i, n_i)\}$ is less than $T_c = 0.4695$, then the neighborhood centered at (m_i, n_i) in the ith scale has the shift distortion.

(b) The elimination distortion is caused by the data lost during the communication. The result is that the wavelet coefficients within certain ares are replaced by zeros. In other words, the local energy of these areas are extremely small. Hence the neighborhood is classified to have elimination distortion if $\mathcal{E}_i(m_i, n_i) < T_e$. The threshold T_e is chosen by testing some possible values (very small values). We found that our quality metric has the best performance when T_e is set to be less than 10^{-14}.

(c) Let us denote $I(m_i, n_i)$ as the indicator function in Equation 6.41. $I(m_i, n_i)$ indicates whether the current neighborhood has the shift or elimination distortion.

$$I(m_i, n_i) = \begin{cases} 1 & \text{if } \rho(m_i, n_i; m_{i+1}, n_{i+1}) < T_c \text{ or } \mathcal{E}_i(m_i, n_i) < T_e \\ \\ 0 & else \end{cases} \tag{6.41}$$

- Evaluate the shift and elimination distortion

The shift distortion moves the random coefficients to the current neighborhood, hence the original information was totally "erased". The elimination distortion causes the same result because of the data loss during communication. Hence the perceptual information of these two distortion should be the original perceptual information of the neighborhood. In Chapter 6 Sections 1 and 2, we discussed the estimation of the original perceptual information in a neighborhood, as shown in Equation 6.42. Hence the perceptual information of the shift or elimination distortion of the current neighborhood can be computed with Equation 6.43, where $D_{m_i,n_i\ \mathrm{E/S}}$ denotes the shift or elimination distortion of the neighborhood centered at (m_i, n_i) in the ith scale.

$$\hat{\mathcal{P}}(X_{m_i,n_i}) \quad = \quad \frac{1}{2}\log\left(1 + \frac{\hat{\mathcal{E}}_i(m_i, n_i)}{\sigma^2_{Z_i}}\right) \tag{6.42}$$

$$\hat{\mathcal{P}}(D_{m_i,n_i\ \mathrm{E/S}}) \quad = \quad \frac{1}{2}\log\left(1+\frac{\hat{\mathcal{E}}_i(m_i,n_i)}{\sigma^2_{Z_i}}\right) \tag{6.43}$$

- Evaluate the perceptual information of the reference image

  To follow the strategy of the full reference quality assessment ITC in Equation 6.1, we need to estimate the perceptual information of the reference image. Based on the local energy consistency model in Chapter 3 Section 3, the local energy of the neighborhood can be estimated. Hence the original perceptual information of the current neighborhood can be computed with Equation 6.44. The total perceptual information of the reference image is estimated as the summation of the perceptual information in each neighborhood across all scales, as shown in Equation 6.45

$$\hat{\mathcal{P}}(X_{m_i,n_i}) = \frac{1}{2}\log\left(1+\frac{\hat{\mathcal{E}}_i(m_i,n_i)}{\sigma^2_{Z_i}}\right) \tag{6.44}$$

$$\hat{\mathcal{P}}(\underline{X}) = \sum_i \sum_{m_i,n_i} \hat{\mathcal{P}}(X_{m_i,n_i}) \tag{6.45}$$

- Compute the visual quality with Equation 6.46.

  Since the fastfading distortion is a mixture of the quantization and shift/elimination distortions, the total perceptual distortion is the summation of the perceptual information of the quantization distortion and the perceptual information of the shift/elimination distortion. Follow the ITC strategy, the visual quality is is normalized perceptual distortion across all scales by the perceptual information of the reference image.

$$BITC_{\text{fastfading}} = \frac{\sum_i \left(\sum_{m_i,n_i} I(m_i,n_i)\hat{\mathcal{P}}(D_{m_i,n_i\ \mathrm{E/S}}) + \hat{\mathcal{P}}(D_{i\ \mathrm{Q}})\right)}{\sum_i \sum_{m_i,n_i} \hat{\mathcal{P}}(m_i,n_i)} \tag{6.46}$$

6.6.2 Performance and Discussion

Figure 6.6.2 plots the performance of $BITC_{\text{fastfading}}$. The linear correlation between the subjective and objective judgments is 0.8665. Spearman correlation is 0.8819. We

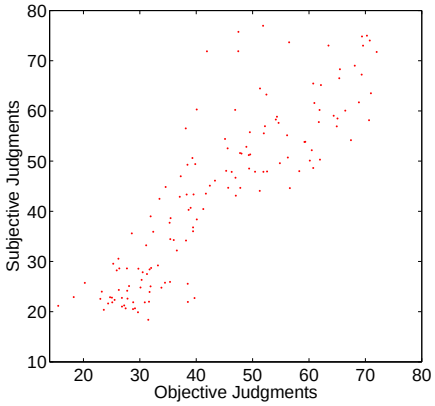


Figure 6.21: The performance of blind quality metric on fastfading distortion. The objective evaluations by $BITC_{\text{fastfading}}$ has the linear correlation of 0.8665 with human judgments. Spearman correlation is 0.8819.

notice that the blind quality metric performs well on small distortions. But the inaccurate estimation degrades the performance when the distortion increases. $BITC_{\text{fastfading}}$ is the first method that is proposed to evaluate the quality of an image with local distortion (shift/elimination) and global distortion (quantization) simultaneously.

6.7 Summary

In this chapter, several blind quality assessment metrics are proposed for different distortions. All of them illustrate an excellent ability to predict image quality without reference image. The basic idea of blind quality assessment is based on the success of full reference ITC in Chapter 5. It can be summarized as the estimation of the perceptual distortion D.

At the beginning of this chapter, two distortion estimation methods for white noise and Gaussian distortion are proposed based on the energy consistency and local energy consistency models. The test results in Chapter 6 and the appendix show that $BITC_{\text{wn}}$ and $BITC_{\text{gblur}}$ have excellent consistency to human judgments under most multi-resolution frameworks. $BITC_{\text{wn}}$ can also deal with other four distortions: noise in the color component [81], high frequency noise [81], spatially correlated noise [81] or impulse noise [81].

A blind quality metric is then presented for JPEG and JPEG2000 distortions. Al-

Distortion	Energy Consistency	Local Energy Consistency	Quantization Pattern
White Noise	×	×	
Gaussian Blur	×	×	
JPEG2000 Compression			×
JPEG Compression			×
Fastfading		×	×

Table 6.7: Blind quality assessment methods

though this method is different from the estimation of the distortion for white noise and Gaussian blur, the basic idea is same: evaluate the perceptual distortion. Compared with the existing metrics, our method has four advantages. First, *BITC* exhibits excellent compatibility to four different subjective databases. Second, it does not need a training procedure. Third, the metric can be easily embedded into JPEG or JPEG2000 codec for quality estimation or quality control. Fourth, the metric has very low computation complexity which makes real-time application possible. At the end of this chapter, we propose the first blind quality metric for images with the fastfading distortion. Three distortion behaviors are combined successfully to evaluate image quality.

Unfortunately, we still have no idea how to derive a universal method that is capable to estimate the perceptual distortion D without extra caution. Table 6.7 gives us a guide as to which model is applied to approach blind quality assessment for certain distortion. White noise and Gaussian blur distortions only require one model (either the energy consistency or local energy consistency). JPEG and JPEG2000 distortions only require one model (quantization pattern). Fastfading distortion requires two models in order to catch the inter-scale consistency degradation and information loss due to quantization procedure. The performance of the blind quality assessment is summarized in Table 6.8.

	WN_1	WN_2	WN_3	WN_4
CC	0.8448	0.8833	0.9695	0.9719
SC	0.9511	0.9501	0.9617	0.9775
	$Blur_1$	$Blur_2$	JPEG2000	JPEG
CC	0.9204	0.9066	0.9474	0.9206
SC	0.9314	0.9066	0.9363	0.9065
	Fastfading			
CC	0.8665			
SC	0.8819			

Table 6.8: Blind quality assessment assessment result. "CC": linear correlation between subjective and objective judgments. "SC": Spearman correlation. WN_1,WN_2,WN_3 and WN_4 are the blind metrics based on the tradition method, color information, energy consistency and local energy consistency respectively. $Blur_1$ and $Blur_2$ are blind metrics based on the energy consistency and local energy consistency.

Chapter 7

Conclusions and Future works

7.1 Conclusions

This dissertation proposes successful image quality assessment methods. Full reference and blind quality assessment strategies are developed based on a the human visual system model and image statistical models. The following sections summarize the contributions in four aspects: image statistical model, human visual system model, full reference quality assessment and non-reference quality assessment.

- **Image Statistics**: Considering the strong dependency of natural images among the scales in the wavelet domain, Chapter 3 proposes the energy consistency model and the local energy model. The energy consistency model emphasizes the wavelet coefficient activity of the entire scale, while the local energy consistency model emphasizes the local activity within a small neighborhood. These two models enable us to estimate the original energy level in the reference image if the distorted (test) image is only accessible. Furthermore, the distortion on the wavelet secondary property by fastfading distortion can be detected with the local energy consistency model.

- **Human Contrast Sensitivity Model**: The human contrast sensitivity at a frequency is considered playing an important role in quality assessment. We modeled contrast sensitivity first with a quantization channel and then with a Gaussian communication channel. With the sphere decoding strategy by C. E. Shannon [94], we explained that the Gaussian communication channel simulates the contrast sensitivity behavior similarly as the quantization channel. A parrel

Gaussian channel is proposed to model contrast sensitivity at different frequencies with each individual channel deals with the image at a certain frequency while the noise level of the channel indicates the contrast sensitivity. The perceptual information of an image is defined as the transmitted information (mutual information) through the contrast sensitivity model. To compute the perceptual information, three statistical models are applied to describe the natural image statistics, since we are interested if the statistical model will make significant influence when evaluating visual quality with the perceptual information.

- **Full Reference Quality Assessment**: A reliable method is proposed to evaluate the visual quality of an image according to its reference image. The visual quality is defined as the perceptual difference between two image against the perceptual information of two images. Five subjective databases were used to validate our method's performance compared with other 18 quality metrics. Our method not only exhibits the excellent compatibility to difference distortions but also outperforms the other existing methods. Furthermore, our method on full reference quality assessment provides the foundation to address blind quality assessment.

- **Blind Quality Assessment**: Our basic idea of blind quality assessment is to blindly estimate the perceptual information of the distortion that occurs during image processing or transmission. The blind quality metric for compressed pictures surpasses all existing method in excellent consistency with human judgment and ease of use. The blind quality assessment on white noise, blur and fastfading distortions is first addressed in this thesis and the test results illustrate the excellent consistency with human opinion.

7.2 Future Works

Although this thesis addresses the quality assessment problem and achieves excellent results, the quality assessment problem still requires more effort on different aspects.

- **Subjective Database**:

  Due to the limitation of knowledge on how the human visual system perceives images, the quality assessment metric is still validated by subjective databases. In other words, subjective databases will determine the performance of the quality assessment metrics. Currently we have three subjective databases which contain over 2800 samples. Each subjective database tends to choose the test samples over a reasonable quality range. Figure 7.1 shows the histogram of subjective quality from 0 (best quality) to 100 (worst quality). We notice very few test samples fall into the quality range between [1, 20]. Test samples with the quality score of zero are identical to their reference images. In practice people usually assign the visual quality of a test sample to be perfect even the test sample contains very small distortion. A study on these "perfect visual quality" test samples will be helpful to understand how human visual system perceives the difference.

  During the establishment of each subjective quality database, 10 - 30 subjects were required to give their judgments on the test samples quality. Unfortunately we do not know whether the listed subjective quality in each database is truly accurate or not. Different subjective qualities will be assigned to the same test sample by different research groups. It has been reported in [47] that different subject groups will agree more on others' opinions about the low quality video samples. There were eight groups that evaluated the video quality on two separate databases. The average linear correlation of subjective quality by eight groups for the database containing high quality video is 0.8587, for low quality video database is 0.8958. This will cause the confusion when a quality metric is validated by a subjective database: a quality metric could be outstanding according to the subjective judgments by one research group but only above the average according to another research group. Hence a subjective database with only the quality score in a future study will not improve the quality assessment tremendously but introduce uncertainty. It will be better if the subjective database could provide extra information about how each subject evaluates the quality. From my point of view, different types of distortion will help people to study how people will perceive the difference. But this study is more like to

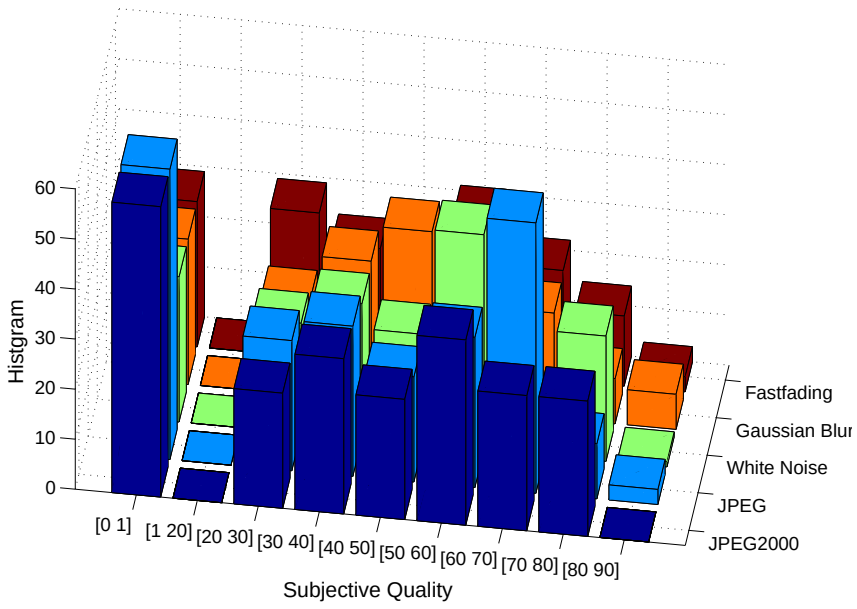


Figure 7.1: Subjective quality histogram

guess a crossword puzzle with only few hints. A more efficient database might contain the information at two different levels: where people perceive difference and how people evaluate difference. These information can be provided by local quality: the quality of every part in an image. More accurate information about where people perceive the difference can be achieved be eye tracking devices.

- **Image Quality Assessment**:

  In this thesis, we have proposed a method to estimate distortion by adopting an energy consistency model. We pointed out that the estimation is distortion dependent. This will force us to derive the corresponding estimation method when facing a new type of distortion. Is there a universal method to properly estimate the distortion?

  This thesis proposes quality assessment metrics for the images of medium size. The resolution is several hundred by several hundred. Great improvements have been achieved in digital imaging. Images of large size or extremely large size are becoming more and more popular. For instance, six years ago a digital camera with 5 million pixels was a standard model. Now a digital single-lens reflex

camera usually has the capacity to capture images with the resolution up to 12 million pixels. The resolution of test samples in the subjective databases is only around 0.4 million pixels, which is not even close to the standard resolution at the beginning of the century. Also high definition LCD monitors or TVs are becoming affordable and accepted by average families. All of these motivate appearance and daily use of high definition images. It is still unknown how people will evaluate the quality of images with the extremely high resolution. People may be more sensitive to the noise on the flat area. As the storage devices and communication technologies improve, images with poor or medium quality will become more and more rare. The quality assessment of images with high resolution and slight quality degradation might become a important direction in the quality assessment area.

Another direction for the future research is to evaluate the quality of the enhanced image. In chapter 1, we mentioned that quality assessment metrics are necessary when evaluating the performance of image processing systems, especially in image enhancement and restoration systems. Deriving a quality metric which can predict quality improvement consistently might be very difficult. We have run the subjective quality assessment experiment on enhanced images. Different enhancement operations are applied to improve the visual quality of noisy images. For a specific noisy image, different operations create 'new' distortions. Subjects often have difficulties comparing new distortions generated by the enhancement operations. We could imagine it will be more difficult to propose a quality metric for quality improvement.

Appendix A

Subjective Image Quality Assessment

◇ Double Stimulus Continuous Quality Scale (DSCQS) Method [2]: "Each trial consists of a pair of stimuli: one stimulus is the reference, and the other is the test. The test stimulus is usually the reference after undergoing some type of processing. The two stimuli are each presented twice in a trial, in alternating fashion, with the order of the two randomly chosen for each trial. To aid the test subjects in staying on track in their assessments, audio cues are used to indicate when a trial begins, when a new stimulus begins, when to vote, and what the current trial number is in the sequence of trials making up a test session." (*Subjective testing for visual quality assessment has been finalized in ITU-R Recommendation BT.500-10 (2000)*, P25)

◇ Double Stimulus Impairment Scale (DSIS) Method [2]: As in the DSCQS method, each trial consists of a pair of stimuli: the reference and the test. However, in the DSIS method, the two stimuli are always presented in the same order: the reference is always first, followed by the test. In the DSIS method, test subjects compare the two stimuli in a trial and rate the impairment of the test stimulus with respect to the reference, using a five-level degradation scale. Thus, only one vote is made for each DSIS trial.

◇ Single Stimulus Continuous Quality Evaluation (SSCQE) [2]: Instead of seeing separate short sequence pairs, viewers watch a program of typically 20-30 minutes duration which has been processed by the system under test; the reference is not shown. Using a slider, the subjects continuously rate the instantaneously perceived quality on the DSCQS scale from "bad" to "excellent".

Appendix B

Entropy Estimation

Suppose we divide the range of x into bins of length $\triangle$ (the quantization step is $\triangle$). Recall the definition of the differential entropy and the basic idea of calculus, by the mean value theorem, there exists a value $\tilde{x}_i$ within each bin such that, as shown in Figure B:

$$f(\tilde{x}_i)\triangle = \int_{i\triangle}^{(i+1)\triangle} f(x)dx \tag{B.1}$$

Consider the quantized random variable:

$$x_\triangle = \tilde{x}_i \qquad if \quad i\triangle \leq x \leq (i+1)\triangle \tag{B.2}$$

Then the probability that $x_\triangle = \tilde{x}_i$

$$p_i = \int_{i\triangle}^{(i+1)\triangle} f(x)dx = f(\tilde{x}_i)\triangle \tag{B.3}$$

Then the entropy of the quantized version $x_\triangle$

$$\begin{aligned} H(x_\triangle) &= -\sum_{i=-\infty}^{+\infty} p_i log_2(p_i) \\ &= -\sum_{i=-\infty}^{+\infty} f(\tilde{x}_i)\triangle \log_2(f(\tilde{x}_i)\triangle) \end{aligned}$$

$$
\begin{aligned}
&= -\sum_{i=-\infty}^{+\infty} f(\tilde{x}_i) \triangle \log_2(f(\tilde{x}_i)) - \sum_{i=-\infty}^{+\infty} f(\tilde{x}_i) \triangle log_2(\triangle) \\
&= -\sum_{i=-\infty}^{+\infty} f(\tilde{x}_i) \triangle \log_2(f(\tilde{x}_i)) - \log_2(\triangle)
\end{aligned}
$$

Since $\sum_{i=-\infty}^{+\infty} f(\tilde{x}_i)\triangle = \int f(x) = 1$, the first term will approach the differential entropy $H(x) = \int f(x) \log_2(f(x))dx$

Hence, we derive the following theorem about he relationship of the differential entropy and discrete entropy. Consider a continuous variable x and its quantized random variable $x_\triangle$

$$H(x_\triangle) \to H(x) - \log_2(\triangle) \qquad \triangle \to 0 \tag{B.4}$$

where $\triangle$ is the quantization step.

Equation B.4 also gives a method to estimate the differential entropy of any source:

$$H(x) \to H(x_\triangle) + \log_2(\triangle) \qquad \triangle \to 0 \tag{B.5}$$

where $\triangle$ is the quantization step. Now let us take a look at the performance of the

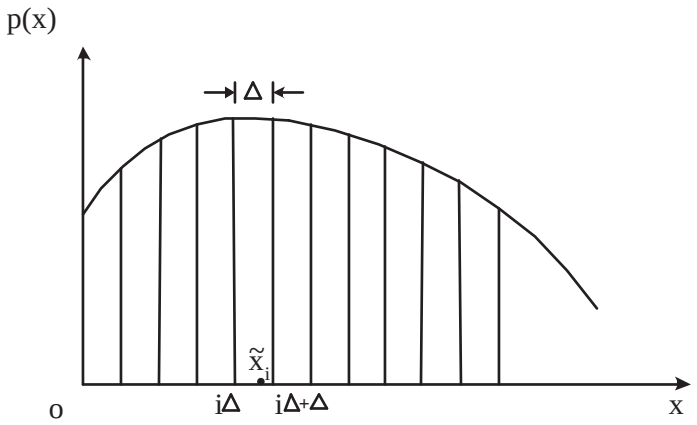


Figure B.1: Quantization of a continuous random variable

above entropy estimation method. If x is Gaussian, by information theory:

$$H(x) = -\int f(x)\log_2(f(x))dx = \frac{1}{2}\log_2(2\pi e\sigma_x^2) \tag{B.6}$$

The quantization step in the entropy estimation is $\frac{\sigma_x}{20}$. Figure B displayed the esti-

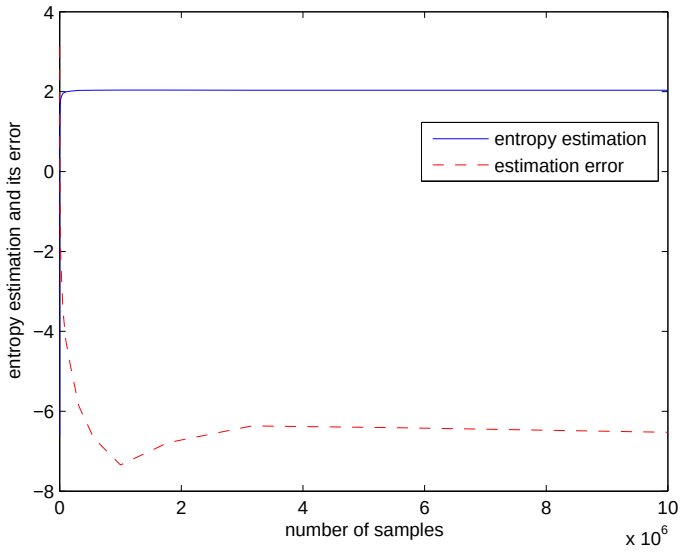


Figure B.2: Performance of entropy estimation

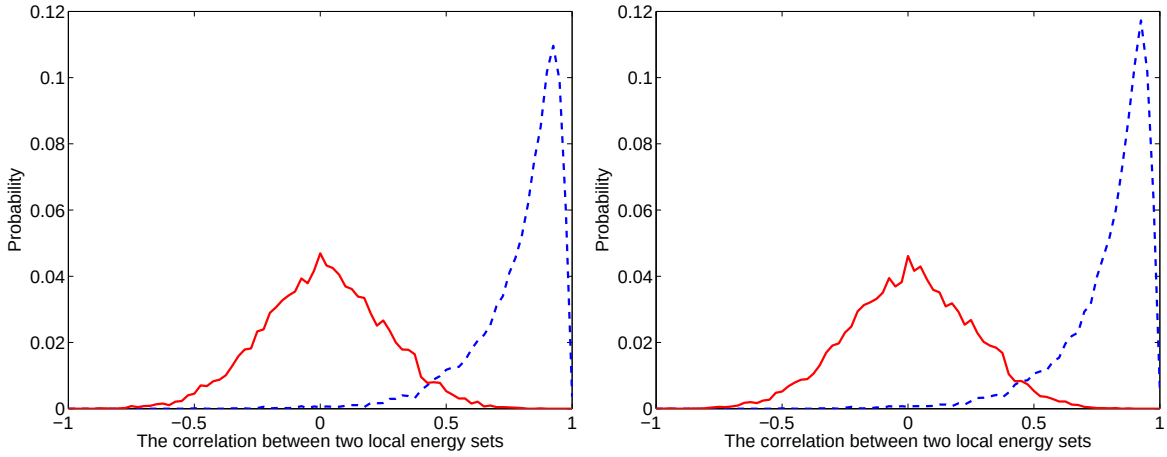
mated entropy of x (with solid line) and the estimation error (with dash line). For display purposes, estimated entropy and error are displayed logarithmically. It is clear that as the number of prior samples gets larger, the estimation error approach to a constant(2^{-6} bit). Compared with theoretic result(2.0471 bit), the error percentage is 0.76%. The constant estimation error corresponds to the tolerance when computing the histogram of the quantized signal/image. With the smaller tolerance, entropy estimation can improve its accuracy.

Appendix C

The local energy set correlation under the db97 wavelet

Wavelet Framework	Local Energy Set Size	A local energy set & its corresponding set		A local energy set & a random set	
		average	standard deviation	average	standard deviation
The 1st and 2nd scales	3×3	0.7185	0.2723	4.438e-3	0.3968
	4×4	0.7552	0.2170	1.453e-2	0.3162
	5×5	0.7781	0.1906	7.825e-3	0.2712
	6×6	0.7960	0.1726	1.181e-2	0.2410
The 2nd and 3rd scales	3×3	0.7002	0.2999	-9.038e-4	0.4043
	4×4	0.7378	0.2477	-6.602e-3	0.3237
	5×5	0.7709	0.2105	-5.817e-3	0.2909
	6×6	0.7905	0.1918	-7.734e-3	0.2633

Table C.1: The correlation between local energy sets. The table lists the average and standard deviation of the correlation between two local energy sets. The test results show that there is strong correlation (the large average and small deviation) between the local energy set and its corresponding set. This verifies the wavelet secondary property. The table also shows the very weak correlation between the local energy set in the ith scale and a random local energy set in the $i+1th$ scale. The test uses 200 natural images and the 'db1' wavelet. The neighborhood of the local energy is 5×5 and 10×10 in the $i+1th$ and ith scale respectively.



(a) (b)

Figure C.1: The probability of the local energy set correlation under the db97 wavelet. Figure (a) plots the probability of the local energy set correlation between the first and second scales. Figure (b) plots the probability of the local energy set correlation between the second and third scales. The solid lines (red) indicate the correlation between two random local energy sets, while the dash lines (blue) indicate the correlation between the local energy set and its corresponding set in the adjacent scale.

Appendix D

Linear Energy Correlation Between Two Adjacent Scales

Multi-resolution Framework	1H-2H	1V-2V	1D-2D
db1	0.97531	0.97255	0.96897
db3	0.94575	0.94651	0.95212
db5	0.92088	0.92563	0.93627
db7	0.95657	0.95072	0.94069
db9	0.95127	0.95953	0.94094
db11	0.90356	0.91000	0.93174
db13	0.94068	0.93450	0.93901
db15	0.94543	0.94468	0.93767
db17	0.94028	0.93252	0.93439
db19	0.91951	0.90186	0.93462
db21	0.94493	0.93281	0.93467
db23	0.94791	0.94024	0.93411
db25	0.94122	0.92802	0.93516
db27	0.93012	0.90381	0.93156
db29	0.94722	0.92731	0.93318
db31	0.94388	0.92755	0.93425
db33	0.93948	0.91160	0.93156
db35	0.93386	0.89809	0.93143
db37	0.94394	0.91878	0.93143
db39	0.94575	0.92156	0.92901
db41	0.93796	0.90245	0.93011
db43	0.93586	0.89272	0.92704
db45	0.94517	0.91324	0.928
coif2	0.95253	0.94558	0.94377
coif4	0.95028	0.94398	0.93784
sym2	0.96074	0.95299	0.96118
sym4	0.95516	0.9519	0.94565
sym6	0.94939	0.94915	0.93495
sym8	0.94570	0.94748	0.93638
sym10	0.94551	0.94721	0.93206
sym12	0.94154	0.948	0.93313
sym14	0.94543	0.94768	0.93172
sym16	0.94145	0.94692	0.9317
sym18	0.94357	0.94492	0.92901
sym20	0.94133	0.946	0.9305
bior1.1	0.97531	0.97255	0.96897
bior1.5	0.96039	0.95056	0.9577
bior2.4	0.95169	0.95226	0.94414
bior2.8	0.95093	0.95086	0.94106
bior3.3	0.87063	0.91848	0.92022
bior3.7	0.8677	0.90987	0.91319
bior4.4	0.93752	0.93641	0.93949
bior6.8	0.93572	0.93815	0.93358
rbio1.3	0.90428	0.93958	0.94501
rbio2.2	0.95516	0.9531	0.94761
rbio2.6	0.93809	0.93705	0.93685
rbio3.1	0.94492	0.95749	0.94854
rbio3.5	0.83625	0.86631	0.93704
rbio3.9	0.8442	0.85839	0.93392
rbio5.5	0.95129	0.95305	0.94114

Appendix E

The Performance of FRQA

QA methods	JPEG2000	JPEG	White Noise
Absolute Error	0.8568	0.8996	0.9864
Mean Squared Error	0.8887	0.9011	0.9856
Maximum Error	0.9408	0.9281	0.9736
Moments of the Angles (D_4)	0.6491	0.8065	0.9550
Moments of the Angles (D_5)	0.7232	0.8171	0.9544
Normalized Cross-correlation	0.8593	0.8820	0.9685
Mutual Information	0.7999	0.8278	0.9151
Entropy Correlation Coefficient	0.8713	0.7590	0.9784
Peak Signal Noise Ratio	0.8887	0.9011	0.9856
Visual Fidelity Criterion	0.8720	0.8619	0.9174
DCTune	0.7901	0.8840	0.9327
Picture Quality Scale	0.8970	0.9327	0.9416
Visible Difference Predictor	0.8932	0.9109	0.9773
Visual Signal Noise Ratio	0.5704	0.7214	0.8985
Structural Similarity	0.9613	0.9764	0.9693
Multi-structural Similarity	0.9651	0.9795	0.9729
Visual Information Fidelity	**0.9695**	**0.9845**	**0.9857**
Information Theoretic Criterion	0.9674	0.9729	0.9722

QA methods	Blur	Fastfading	Overall
Absolute Error	0.7333	0.8579	0.9220
Mean Squared Error	0.7824	0.8873	0.9350
Maximum Error	0.8672	0.8409	0.9351
Moments of the Angles (D_4)	0.4896	0.5418	0.7935
Moments of the Angles (D_5)	0.6109	0.7136	0.8558
Normalized Cross-correlation	0.7512	0.8922	0.9283
Mutual Information	0.6607	0.8385	0.9022
Entropy Correlation Coefficient	0.7100	0.8656	0.9111
Peak Signal Noise Ratio	0.7824	0.8873	0.9350
Visual Fidelity Criterion	0.7145	0.8128	0.9277
DCTune	0.6722	0.7645	0.8997
Picture Quality Scale	0.9303	0.9074	0.9510
Visible Difference Predictor	0.8798	0.8696	0.9426
Visual Signal Noise Ratio	0.6642	0.6966	0.7042
Structural Similarity	0.9516	0.9556	0.9728
Multi-structural Similarity	0.9586	0.9318	0.9707
Visual Information Fidelity	0.9728	0.9650	0.9765
Information Theoretic Criterion	0.9700	**0.9670**	**0.9780**

Table E.1: The performance of quality metrics for LIVE database. Bold numbers indicate the best performances.

QA methods	JPEG2000	JPEG	LAR coding
Absolute Error	0.8099	0.6764	0.6105
Mean Squared Error	0.8133	0.6641	0.6193
Maximum Error	0.7755	0.4664	0.3545
Moments of the Angles (D_4)	0.4839	0.3315	0.0194
Moments of the Angles (D_5)	0.5575	0.3575	0.1627
Normalized Cross-correlation	0.7521	0.5477	0.5043
Mutual Information	0.6900	0.4968	0.4273
Entropy Correlation Coefficient	0.7815	0.5389	0.4956
Peak Signal Noise Ratio	0.8133	0.6641	0.6193
Visual Fidelity Criterion	0.8018	0.6379	0.5663
DCTune	0.8517	0.6654	0.4035
Picture Quality Scale	0.8268	0.5490	0.5306
Visible Difference Predictor	0.8414	0.5707	0.7022
Visual Signal Noise Ratio	0.8224	0.6237	0.6727
Structural Similarity	0.9297	0.9225	0.8871
Multi-structural Similarity	0.9298	0.9179	0.8745
Visual Information Fidelity	0.9355	0.9239	0.8877
Information Theoretic Criterion	**0.9387**	**0.9507**	**0.9022**

QA methods	Blur	JPEG gray compression	Overall
Absolute Error	0.7953	0.5350	0.6751
Mean Squared Error	0.7893	0.5312	0.6825
Maximum Error	0.7699	0.6539	0.5654
Moments of the Angles (D_4)	0.6102	0.3066	0.2937
Moments of the Angles (D_5)	0.5861	0.3513	0.3832
Normalized Cross-correlation	0.7343	0.5285	0.6128
Mutual Information	0.6824	0.4179	0.5122
Entropy Correlation Coefficient	0.7644	0.4341	0.5672
Peak Signal Noise Ratio	0.7893	0.5312	0.6825
Visual Fidelity Criterion	0.7248	0.4437	0.5938
DCTune	0.7155	0.4784	0.6117
Picture Quality Scale	0.8397	0.7534	0.6650
Visible Difference Predictor	0.8389	0.6275	0.6916
Visual Signal Noise Ratio	0.9413	0.7800	0.7635
Structural Similarity	0.9315	0.8921	0.9018
Multi-structural Similarity	0.9443	0.8463	0.8844
Visual Information Fidelity	**0.9729**	0.8775	0.8963
Information Theoretic Criterion	0.9653	**0.9179**	**0.9184**

Table E.2: The performance of quality metrics for IVC database. Bold numbers indicate the best performances.

QA methods	JPEG2000	JPEG	Overall
Absolute Error	0.8237	0.3284	0.6040
Mean Squared Error	0.8496	0.3744	0.6386
Maximum Error	0.8950	0.5011	0.6821
Moments of the Angles (D_4)	0.6758	0.4097	0.5357
Moments of the Angles (D_5)	0.7429	0.3166	0.5519
Normalized Cross-correlation	0.8108	0.3264	0.5944
Mutual Information	0.7866	0.3183	0.5797
Entropy Correlation Coefficient	0.8236	0.1916	0.5460
Peak Signal Noise Ratio	0.8496	0.3744	0.6386
Visual Fidelity Criterion	0.8313	0.2868	0.5842
DCTune	0.8830	0.7466	0.7639
Picture Quality Scale	0.9030	0.8052	0.8552
Visible Difference Predictor	0.8545	0.4712	0.6791
Visual Signal Noise Ratio	0.9009	0.7974	0.8347
Structural Similarity	0.9399	0.8590	0.8639
Multi-structural Similarity	0.9423	0.8307	0.8663
Visual Information Fidelity	0.9569	0.9059	0.8865
Information Theoretic Criterion	**0.9581**	**0.9178**	**0.9265**

Table E.3: The performance of different quality metrics for TOYAMA database. Bold numbers indicate the best performances.

QA methods	Overall
Absolute Error	0.3824
Mean Squared Error	0.4844
Maximum Error	0.3824
Moments of the Angles (D_4)	-*
Moments of the Angles (D_5)	-*
Normalized Cross-correlation	0.4835
Mutual Information	0.2434
Entropy Correlation Coefficient	0.1473
Peak Signal Noise Ratio	0.4844
Visual Fidelity Criterion	0.4888
DCTune	0.2616
Picture Quality Scale	0.0671
Visible Difference Predictor	0.2110
Visual Signal Noise Ratio	**0.9339**
Structural Similarity	0.8077
Multi-scale Structural Similarity	0.8394
Visual Information Fidelity	0.6224
Information Theoretic Criterion	0.8454

Table E.4: The performance of different quality metrics for the Cornell A57 database [25]. *Test samples in Cornell A57 database are all gray level images. Moments of the angles metrics (D_4, D_5) can only work on colorful images. Bold numbers indicate the best performances.

QA methods	Gaussian noise	Noise in color components	Spatially correlated noise
Absolute Error	0.9078	0.9056	0.9178
Mean Squared Error	0.9114	0.9067	0.9228
Maximum Error	0.9179	0.9078	**0.9309**
Moments of the Angles (D_4)	0.7694	0.7550	0.7481
Moments of the Angles (D_5)	0.7642	0.7480	0.7438
Normalized Cross-correlation	0.8404	0.8469	0.8389
Mutual Information	0.7934	0.7945	0.8112
Entropy Correlation Coefficient	0.8283	0.8401	0.8457
Peak Signal Noise Ratio	0.9114	0.9067	0.9228
Visual Fidelity Criterion	0.9069	0.9033	0.9159
DCTune	0.8416	0.7841	0.8605
Picture Quality Scale	0.7604	0.7432	0.6987
Visible Difference Predictor	**0.9235**	0.8967	0.9306
Visual Signal Noise Ratio	0.7882	0.7849	0.7858
Structural Similarity	0.8106	0.8029	0.8145
Multi-structural Similarity	0.8086	0.8045	0.8198
Visual Information Fidelity	0.8797	0.8757	0.8698
Information Theoretic Criterion	0.9184	**0.9092**	0.9080

QA methods	Masked noise	High frequency noise	Impulse noise
Absolute Error	0.8350	0.9281	**0.9179**
Mean Squared Error	0.8486	0.9323	0.9176
Maximum Error	0.6669	**0.9565**	0.3851
Moments of the Angles (D_4)	0.5771	0.8708	0.8705
Moments of the Angles (D_5)	0.5733	0.8679	0.8498
Normalized Cross-correlation	0.7181	0.9080	0.8831
Mutual Information	0.6767	0.8638	0.6583
Entropy Correlation Coefficient	0.6964	0.8782	0.7902
Peak Signal Noise Ratio	0.8486	0.9323	0.9176
Visual Fidelity Criterion	0.8663	0.8712	0.8627
DCTune	0.6774	0.8721	0.7440
Picture Quality Scale	0.6963	0.8713	0.8365
Visible Difference Predictor	**0.8750**	0.9056	0.8180
Visual Signal Noise Ratio	0.7921	0.8961	0.6655
Structural Similarity	0.7794	0.8728	0.6732
Multi-structural Similarity	0.8150	0.8676	0.6867
Visual Information Fidelity	0.8683	0.9074	0.8327
Information Theoretic Criterion	0.8479	0.9362	0.8841

QA methods	Quantization noise	Gaussian blur	Image denoising
Absolute Error	0.8647	0.8505	0.9189
Mean Squared Error	0.8699	0.8681	0.9380
Maximum Error	**0.9060**	0.8221	0.8666
Moments of the Angles (D_4)	0.6771	0.6481	0.6759
Moments of the Angles (D_5)	0.7050	0.7773	0.7465
Normalized Cross-correlation	0.7998	0.8170	0.8985
Mutual Information	0.7293	0.7807	0.8490
Entropy Correlation Coefficient	0.6706	0.7916	0.8555
Peak Signal Noise Ratio	0.8699	0.8681	0.9380
Visual Fidelity Criterion	0.8587	0.9161	0.9270
DCTune	0.8115	0.8752	0.8696
Picture Quality Scale	0.7063	0.7358	0.8571
Visible Difference Predictor	0.8897	0.9343	0.9358
Visual Signal Noise Ratio	0.8147	0.9236	0.9134
Structural Similarity	0.8530	0.9544	0.9529
Multi-structural Similarity	0.8530	**0.9605**	**0.9566**
Visual Information Fidelity	0.7970	0.9540	0.9160
Information Theoretic Criterion	0.8938	0.9192	0.9330

Table E.5: The performance of quality metrics for TID2008 database. Bold numbers indicate the best performances.

QA methods	JPEG	JPEG2000	JPEG errors
Absolute Error	0.9160	0.8144	0.8355
Mean Squared Error	0.9011	0.8301	0.7664
Maximum Error	0.8748	0.7841	0.4394
Moments of the Angles (D_4)	0.7882	0.5732	0.5101
Moments of the Angles (D_5)	0.7910	0.5998	0.6891
Normalized Cross-correlation	0.8319	0.7800	0.7177
Mutual Information	0.7498	0.8038	0.7893
Entropy Correlation Coefficient	0.6036	0.7872	0.7988
Peak Signal Noise Ratio	0.9011	0.8301	0.7664
Visual Fidelity Criterion	0.8472	0.7771	0.7681
DCTune	0.9079	0.9401	0.7427
Picture Quality Scale	0.8002	0.8914	0.7940
Visible Difference Predictor	0.9229	0.8466	0.8599
Visual Signal Noise Ratio	0.9278	0.9489	0.7842
Structural Similarity	0.9251	0.9624	0.8677
Multi-structural Similarity	**0.9347**	**0.9734**	**0.8737**
Visual Information Fidelity	0.9167	0.9709	0.8585
Information Theoretic Criterion	0.9318	0.9639	0.6466

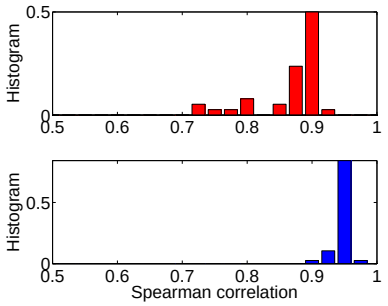
QA methods	JPEG2000 errors	Non-eccentricity pattern noise	Local block-wise distortions
Absolute Error	0.8664	0.6476	0.8680
Mean Squared Error	0.7765	0.5930	0.5851
Maximum Error	0.5278	0.2815	0.6713
Moments of the Angles (D_4)	0.4347	0.7167	0.6222
Moments of the Angles (D_5)	0.6256	0.7230	0.7648
Normalized Cross-correlation	0.7697	0.5046	0.2718
Mutual Information	0.8042	0.6845	0.4263
Entropy Correlation Coefficient	0.8180	0.7690	**0.8688**
Peak Signal Noise Ratio	0.7765	0.5930	0.5851
Visual Fidelity Criterion	0.8210	0.5467	0.6132
DCTune	**0.8802**	**0.8073**	0.0686
Picture Quality Scale	0.7424	0.5973	0.4994
Visible Difference Predictor	0.7272	0.6175	0.8554
Visual Signal Noise Ratio	0.7926	0.5144	0.1479
Structural Similarity	0.8576	0.7107	0.8462
Multi-structural Similarity	0.8521	0.7334	0.7607
Visual Information Fidelity	0.8500	0.7619	0.8323
Information Theoretic Criterion	0.5628	0.7946	0.7542

QA methods	Mean shift (intensity shift)	Contrast change	Overall
Absolute Error	0.7064	0.5927	0.2640
Mean Squared Error	0.6973	0.6125	0.5245
Maximum Error	0.7218	0.6182	0.3843
Moments of the Angles (D_4)	0.6005	0.0058	0.0602
Moments of the Angles (D_5)	0.6977	0.1515	0.1255
Normalized Cross-correlation	**0.7577**	0.5666	0.4787
Mutual Information	0.2383	0.4348	0.4707
Entropy Correlation Coefficient	0.3870	0.2503	0.4247
Peak Signal Noise Ratio	0.6973	0.6125	0.5245
Visual Fidelity Criterion	0.4094	0.4093	0.5861
DCTune	0.7145	0.4769	0.4758
Picture Quality Scale	0.5596	0.2346	0.4252
Visible Difference Predictor	0.6082	0.4980	0.7459
Visual Signal Noise Ratio	0.3643	0.4397	0.7118
Structural Similarity	0.7230	0.5245	0.7749
Multi-structural Similarity	0.7347	0.6393	**0.8526**
Visual Information Fidelity	0.5095	**0.8188**	0.7490
Information Theoretic Criterion	0.7079	0.6367	0.5128

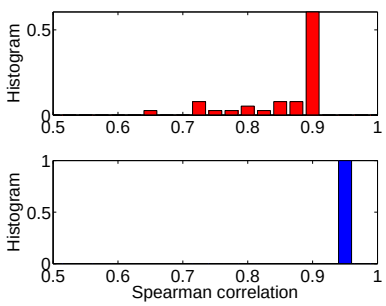
Table E.6: The performance of quality metrics for TID2008 database(continued). Bold numbers indicate the best performances.

Appendix F

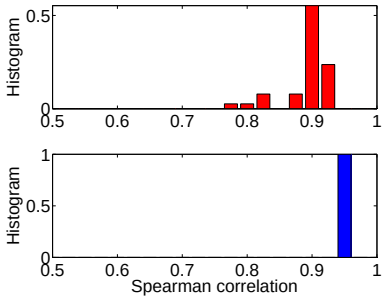
Performance of Blind ITC on Noise



(a) Horizontal



(b) Vertical

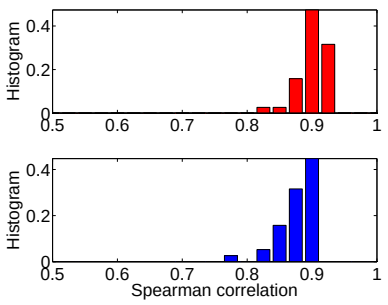


(c) Diagonal

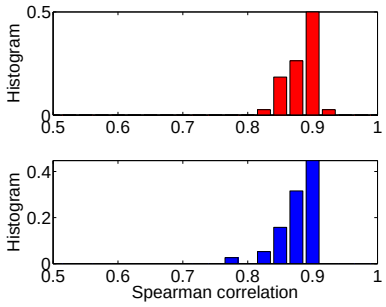
Figure F.1: The performance of $BITC_{\text{wn}}$ (Spearman correlation).Figure (a) - (c) plot the density of correlation coefficients. The density of Spearman correlation coefficients using energy consistency model in $BITC_{\text{wn}}$ is plotted in the top half of each figure, and the density of Spearman correlation coefficients using local energy consistency model in $BITC_{\text{wn}}$ is plotted in the bottom half.

Appendix G

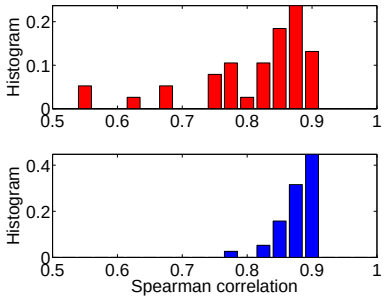
Performance of Blind ITC on Blur



(a) Horizontal



(b) Vertical



(c) Diagonal

Figure G.1: The performance of $BITC_{\text{gblur}}$ (Spearman correlation). Figure (a) - (c) plot the density of correlation coefficients. The density of Spearman correlation coefficients using energy consistency model in $BITC_{\text{gblur}}$ is plotted in the top half of each figure, and the density of Spearman correlation coefficients using local energy consistency model in $BITC_{\text{gblur}}$ is plotted in the bottom half.

Appendix H

Quantization Step Determination

```
function quantizationstep = findQuantizationStep(coefficients)

% input: coefficients
% The coefficients within a scale (JPEG2000) or at a frequency (JPEG)
% output: quantizationstep
% The quantization step within the scale or at that frequency.

% derive the histogram of the coefficients

mincoefficient = min(coefficients); maxcoefficient = max(coefficients);

histogram = hist(coefficients, mincoefficient: (maxcoefficient - mincoefficient)/200: maxcoefficient);

% derive the position where the histogram reaches the maximum

[sortedhistogram,index] = sort(histogram);
maxposition = index(length(index));

% determine the quantization step with the the right half of the histogram

histgramrighthalf = histogram(maxposition:length(histogram));

% scan the right half of the histogram, if the histogram drop less than 0.02 of the maximum histogram

position = 1;
while(histgramrighthalf(position) > 0.02*max(histogram))
position = position + 1;
if(position > length(histgramrighthalf))        break;        end
end

% extremely large quantization step
if(position > length(histgramrighthalf))
quantizationstep = position*(maxcoefficient - mincoefficient)/200
return;
end

% search next histogram peak
[temp, index] = sort(histgramrighthalf(position: length(histgramrighthalf)));
%next peak position is index(length(index));

quantizationstep = position + index(length(index));
return;
```

Appendix I

The Perceptual Information Approximation

The perceptual information of X is defined as the mutual information between X and Y, where $Y = X + Z$ and Z is Gaussian. If X is uniformly distributed (Equation I.1), there is no closed form to compute $\mathcal{P}(X) = I(X;Y)$. There are at least two methods to approximate $I(X;Y)$. Equation I.3 comes from the definition of mutual information [28]. Equation I.4 is because that Z is independent of X. In Equation I.4, $h(Z) = \frac{1}{2}\log(2e\pi\sigma_Z^2)$ [28].

$$f(X) = \begin{cases} \frac{1}{\triangle} & \text{for } |X| < \frac{1}{2\triangle} \\ 0 & \text{else} \end{cases} \tag{I.1}$$

$$\mathcal{P}(X) = I(X;Y) \tag{I.2}$$

$$= h(Y) - h(Y|X) \tag{I.3}$$

$$= h(Y) - h(Z) \tag{I.4}$$

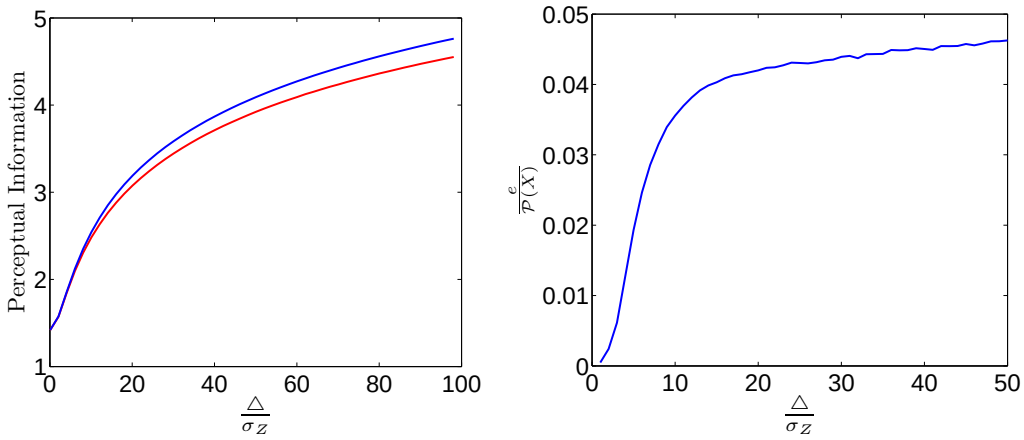
To compute $h(Y)$, we suggest the following approximation. Suppose that $\hat{X}$ is Gaussian and has the same variance as X. The differential entropy of $\hat{Y} = \hat{X} + Z$ is $\frac{1}{2}\log 2\pi e(\sigma_{\hat{X}}^2 + \sigma_Z^2)$ [28], since $\hat{Y}$ obeys $g(0, \sigma_{\hat{X}}^2 + \sigma_Z^2)$. We use $h(\hat{Y})$ substitute $h(Y)$ in Equation I.3 to approximate $\mathcal{P}(X)$, as shown in Equation I.5.

$$\mathcal{P}(X) \simeq \log(1 + \frac{\sigma_{\hat{X}}^2}{\sigma_Z^2}) \tag{I.5}$$

$$= \log(1 + \frac{\sigma_X^2}{\sigma_Z^2}) \tag{I.6}$$

$$= \log(1 + \frac{\triangle^2}{12\sigma_Z^2}) \tag{I.7}$$

The test results are shown in Figure I.1. Figure I.1(b) illustrates that the approximation error is less than 5% of the actual perceptual information. Hence the perceptual information of an uniform distribution can be computed by Equation I.7.



(a) (b)

Figure I.1: The perceptual information approximation. Figure (a) displays the actual perceptual information of an uniform distribution in the dash (red) line against the approximated perceptual information by Equation I.7 in the solid (blue) line. The ratio of $\triangle$ over σ_Z varies from 0.01 to 100. Figure (b) displays the estimation error e over the actual perceptual information as $\triangle/\sigma_Z$ changes.

Appendix J

The Wavelets

The performance of the information theoretic criterion (ITC) for image quality assessment is tested under 39 wavelet frameworks, which is listed in the following table. Table J.1 listed two types of wavelets: orthogonal wavelets [70] and non-orthogonal wavelets [70]. In the first category, eight Daubechies compactly supported wavelets are listed in the form of "db$N$". The number N (1, 2 or 3) after "db" indicates the vanishing moments [70] of the wavelet. In the second category, biorthogonal wavelets [70] (bior$M.N$) and reverse biorthogonal wavelets [70] (rbio$M.N$) are listed. M and N indicates the vanishing moments of wavelet bases for analysis and synthesis. We also listed the steerable pyramid [104] in the second category which has better orientation selectivity (6 orientations).

Orthogonal wavelets [70]	'db1', 'db2', 'db3', 'db4', 'db5', 'db6', 'db7', 'db8'
Non-orthogonal wavelets [70]	'bior1.1', 'bior1.3', 'bior1.5', 'bior2.2', 'bior2.4', 'bior2.6', 'bior2.8', 'bior3.1', 'bior3.3', 'bior3.5', 'bior3.7', 'bior3.9', 'bior4.4', 'bior5.5', 'bior6.8', 'rbio1.1', 'rbio1.3', 'rbio1.5', 'rbio2.2', 'rbio2.4', 'rbio2.6', 'rbio2.8', 'rbio3.1', 'rbio3.3', 'rbio3.5', 'rbio3.7', 'rbio3.9', 'rbio4.4', 'rbio5.5', 'rbio6.8','Steerable Pyramid' [104]

Table J.1: The wavelets used in the quality assessment metric design.

Bibliography

[1] Digital compression and coding of continuous-tone still images. *ITU Telecom. Standardization Sector of ITU*, 1992. 61

[2] *Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-10*, August 2000. 1, 2, 3, 26, 90, 143

[3] JPEG 2000 image coding system: Core coding system. *ITU Telecom. Standardization Sector of ITU*, 2002. 55, 61

[4] A. Ahumada and H. Peterson. Luminance-model-based DCT quantization for color image compression. *Human Vision, Visual Processing, and Digital Display III*, pages 365–374, 1992. 13

[5] I. Alexander and A. Cowey. The cortical basis of global motion detection in blindsight. *Experimental Brain Research*, 192(3):407–411, 2009.

[6] I. Alexander and A. Cowey. Differential and co-involvement of areas of the temporal and parietal streams in visual tasks. *Neuropsychologia*, 47:1609–1614, 2009.

[7] D. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos. Distance measures for color image retrieval. *International Conference on Image Processing*, 2:770–774, 1998.

[8] A.N. Avanaki, S. Sodagari, and A. Diyanat. Reduced reference image quality assessment metric using optimized parameterized wavelet watermarking. *Signal Processing, 9th International Conference on*, pages 868–871, Oct. 2008. 7

[9] I. Avcibas, B. Sankur, and K. Sayood. Statistical evaluation of image quality measures. *Journal of Electronic Imaging*, 11:206–223, 2002. 7, 8

[10] Z. Azimifar. *Image models for wavelet domain statistics*. PhD Thesis, Department of Systems Design Engineering, University of Waterloo, 2005. 33

[11] Z. Azimifar, P. W. Fieguth, and Ed Jernigan. Correlated wavelet shrinkage: models of local random fields across multiple resolutions. *International Conference on Image Processing*, 3:157–160, 2005.

[12] H. B. BARLOW. Increment thresholds at low intensities considered as signal/noise discriminations. *J. Physiol.*, 136, 1957. 4

[13] H. B. BARLOW. Temporal and spatial summation in human vision at different background intensities. *J. Physiol.*, 141, 1958. 4

[14] H. B. BARLOW and WIESEL T. N. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195, 1968. 5

[15] A. Bouzerdoum, A. Havstad, and A. Beghdadi. Image quality assessment using a neural network approach. *Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on*, pages 330–333, Dec. 2004. 7

[16] A. C. Bovik and S. Z. Liu. DCT-domain blind measurement of blocking artifacts in DCT-coded images. *ICASSP '01: Proceedings of the Acoustics, Speech, and Signal Processing, 2001. on IEEE International Conference*, pages 1725–1728, 2001. 20, 21

[17] A. P. Bradley. A wavelet visible difference predictor. *Image Processing, IEEE Transactions on*, 8(5):717–730, May 1999. 6, 7, 9, 16, 17, 24, 25, 27, 28, 61, 92

[18] R. W. Buccigrossi and E. P. Simoncelli. Image compression via joint statistical characterization in the wavelet domain. *IEEE Transactions on Image Processing*, 8:1688–1701, 1998. 21

[19] A. R. Calderbank and B.-L. Yeo I. Daubechies, W. Sweldens. Wavelet transforms that map integers to integers. *Applied and Computational Harmonic Analysis*, 5(3):332–369, Jul. 1998. 32

[20] J. W. Campbell and J. G. Robson. Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197:551 – 566, 1968. 4, 65

[21] P. Campisi, M. Carli, G. Giunta, and A. Neri. Tracing watermarking for multimedia communication quality assessment. *Communications, 2002. ICC 2002. IEEE International Conference on*, 2:1154–1158, 2002. 7, 23

[22] P. Campisi, M. Carli, G. Giunta, and A. Neri. Blind quality assessment system for multimedia communications using tracing watermarking. *Signal Processing, IEEE Transactions on*, 51(4):996–1002, Apr 2003. 7, 23

[23] M. Carnec, P. Le Callet, and D. Barba. Full reference and reduced reference metrics for image quality assessment. *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, 1:477–480, July 2003. 7

[24] M. Carnec, P. Le Callet, and D. Barba. Visual features for image quality assessment with reduced reference. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 1:I–421–4, September 2005. 7

[25] D. M. Chandler and S. S. Hemami. Online supplement to "vsnr: A visual signal-to-noise ratio for natural images based on near-threshold and suprathreshold vision". http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html. 24, 26, 27, 90, 94, 152

[26] H. Choi and R. G. Baranuik. Multiscale image segmentation using wavelet-domain hidden markov models. *IEEE Trans. Image Processing*, 10:1309–1321, 1999. 33

[27] K. Chono, Y. C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod. Reduced-reference image quality assessment using distributed source coding. *Multimedia and Expo, 2008 IEEE International Conference on*, pages 609–612, Apr. 2008. 7

[28] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991. 10, 59, 60, 61, 68, 158

[29] M. Crouse, R. Nowak, and R. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *Proceedings ICASSP-97 (IEEE International Conference on Acoustics, Speech and Signal Processing)*, pages 886–902, 1997. 21, 33, 35, 65, 66, 68, 70, 81, 82, 115

[30] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. *Digital images and human vision*, pages 179–206, 1993. 17

[31] 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 Transactions on Image Processing*, 9(4):636–650, 2000.

[32] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992. 31

[33] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles, 1980.

[34] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, 1985.

[35] W. R. Ding, Y. B. Tong, Q. S. Zhang, and D. K. Yang. Image and video quality assessment using neural network and svm. *Tsinghua University Press Published by Elsevier B.V.*, 13:112–116, 2008. 7

[36] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. 43, 44, 96, 97, 98, 99

[37] M. P. Eckert and A. P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Process.*, 70(3):177–200, 1998. 7

[38] A. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Transactions on Communications*, 43:2959 – 2965, December 1995. 7, 8, 9

[39] G. Fan and X. Xia. Image denoising using a local contextual hidden Markov model in thewavelet domain. *IEEE Signal Processing Letters*, 8(5):125–128, 2001. 33, 35, 81

[40] G. Fan and X. Xia. Improved hidden Markov models in the wavelet-domain. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 49(1):115–120, 2001. 33, 35

[41] G. Fan and X. Xia. Wavelet-based texture analysis and synthesis using hidden Markov models. *IEEE Transactions on Circuits and Systems*, 50(1):106 – 120, 2003. 33, 35

[42] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–2394, 1987. 36

[43] K. Fliegel. A model utilizing artificial neural network for perceptual image quality assessment in image compression algorithms. *Proceedings of the SPIE*, 6315, Dec. 2006. 7

[44] P. Gastaldo, S. Rovetta, and R. Zunino. Objective quality assessment of MPEG-2 video streams by using CBP neural networks. *IEEE Transactions on Neural Networks*, 13(4):939–947, July 2002. 7

[45] V. Di Gesú and V. Starovoitov. Distance-based functions for image comparison. *Pattern Recognition Letter*, 20(2):207–214, 1999.

[46] Video Quality Experts Group. Video quality experts group test sequences. 2000. http://www.vqeg.org/. 3

[47] Video Quality Experts Group. Final report from the video quality experts group on the validation of objective models of video quality assessment. 2000 March. 1, 3, 11, 140

[48] J. A. Guerrero-Colon, L. Mancera, and J. Portilla. Image restoration using space-variant Gaussian scale mixtures in overcomplete pyramids. *IEEE Transactions on Image Processing*, 17(1):27–41, 2008. 36

[49] I. Gunawan and M. Ghanbari. Reduced-reference picture quality estimation by using local harmonic amplitude information. *Proceedings of London Communications Symposium*, pages 137–140, 2003. 7

[50] A. Haar. *Zur theorie der orthogonalen funktionensysteme*. PhD Dissertation, University of Gttingen, 1909. 31

[51] D. K. Hammond and E. P. Simoncelli. Image denoising with an orientation-adaptive Gaussian scale mixture model. *IEEE International Conference on Image Processing*, pages 1433–1436, 2006. 36, 81

[52] M. Holliman and M. M. Yeung. Watermarking for automatic quality monitoring. *Proc. SPIE Security and Watermarking of Multimedia Contents*, 4675:458–469, 2002. 7, 23

[53] Y. Horita, S. Arata, and T. Murai. No-reference image quality assessment for JPEG/JPEG2000 coding. *EUSIPCO*, 2004. 21

[54] Y. Horita, M. Sato, Y. Kawayoke, Z. M. P. Sazzad, and K. Shibata. Image quality evaluation model based on local features and segmentation. *Image Processing, 2006 IEEE International Conference on*, pages 405–408, October 2006. 28

[55] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad. MICT image quality evaluation database. 2005. http://mict.eng.u-toyama.ac.jp/mict/index2.html/. 24, 26, 27, 28, 90, 94, 95, 107, 113, 116, 122, 124

[56] Yuukou Horita, Masaharu Sato, Yoshikazu Kawayoke, Z.M. Parvez Sazzad, and Keiji Shibata. Image quality evaluation model based on local features and segmentation. *Image Processing, 2006 IEEE International Conference on*, pages 405–408, Oct. 2006.

[57] M. Huber, T. Bailey, H. Durrant-Whyte, and U. Hanebeck. On entropy approximation for gaussian mixture random vectors. *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008. 68, 69

[58] A. K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989. 19

[59] M. H. Kayvanrad, S. Sodagari, A. N. Avanaki, and H. Ahmadi-Noubari. Reduced reference watermark-based image transmission quality metric. *Communications, Control and Signal Processing, 3rd International Symposium on*, pages 526–531, March 2008. 7

[60] T. M. Kusuma and H. J. Zepernick. A reduced-reference perceptual quality metric for in-service image quality assessment. *Mobile Future and Symposium on Trends in Communications*, pages 71 – 74, October 2003. 7

[61] T. M. Kusuma, H. J. Zepernick, and M. Caldera. On the development of a reduced-reference perceptual image quality metric. *Systems Communications, 2005. Proceedings*, pages 178–184, Aug. 2005. 7

[62] Y. K. Lai, J. Li, and J. Kuo. A haar wavelet approach to compressed image quality measurement. *Journal of Visual Communication and Image Representation*, 11:17–40, 2000. 6, 17

[63] C. Lawson and R. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Inc., Englewood Cliffs, NJ, USA, 1974. 40

[64] P. Le Callet and F. Autrusseau. Subjective quality assessment IRCCyN/IVC database. 2005. http://www.irccyn.ec-nantes.fr/ivcdb/. 3, 24, 26, 27, 28, 90, 94, 95, 113, 116, 122, 124

[65] G. E. Legge and J. M. Foley. Contrast masking in human vision. *J. Physiol.*, 70(12):1458 – 1471, 1980. 6

[66] J. L. Li, G. Chen, and Z. R. Chi. A fuzzy metric for image quality assessment. *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, 2:562–565, Dec. 2001. 7

[67] X. Li. Blind image quality assessment. *Image Processing. 2002. Proceedings. 2002 International Conference on*, 1:I–449–I–452 vol.1, 2002. 21

[68] S. Z. Liu and A. C. Bovik. Efficient DCT-domain blind measurement and reduction of blocking artifacts. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1139–1149, Dec 2002. 7, 15, 20, 21

[69] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2), 1997. 7, 8, 11

[70] S. Mallat. *A wavelet tour of signal processing*. AP Professional, London, 1997. 31, 32, 40, 45, 48, 76, 79, 87, 96, 160

[71] S. G. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, 38(2):617–643, 1992.

[72] J. L. Mannos and D. J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory*, 20(4):525–535, 1974. 5, 7, 12, 24, 25

[73] P. Marziliano, F. DuFaux, S. Winkler, and T. Ebrahimi. A no-reference perceptual blur metric. *International Conference on Image Processing*, 3:57–60, 2002. 7

[74] F. Meng, X. H. Jiang, H. Sun, and S. Yang. Objective perceptual video quality measurement using a foveation-based reduced reference algorithm. *Multimedia and Expo, 2007 IEEE International Conference on*, pages 308–311, July 2007. 7

[75] Y. Meyer. Orthonormal wavelets. *Wavelets. Time-Frequency Methods and Phase Space, Proceedings of the International Conference, Marseille, France*, 1989. 31

[76] T. Mitsa and K. Varkur. Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms. *International Conference on Acoustics, Speech, and Signal Processing*, 5:301–304, 1993. 12

[77] M. Miyahara, K. Kotani, and V. R. Algazi. Objective picture quality scale (PQS) for image coding. *Communications, IEEE Transactions on*, 46(9):1215–1226, Sep 1998. 4, 6, 7, 13, 16, 24, 25, 28

[78] A. Ninassi, P. L. Callet, and F. Autrusseau. Pseudo no reference image quality metric using perceptual data hiding. *SPIE Human Vision and Electronic Imaging*, 6057, Febrary 2006.

[79] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901. 24

[80] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin. Metrics performance comparison for color image database. *Fourth international workshop on video processing and quality metrics for consumer electronics*, Jan. 2009. 27

[81] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. Color image database for evaluation of image quality metrics. *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 403–408, Oct. 2008. 24, 26, 27, 28, 90, 94, 95, 101, 113, 116, 122, 124, 135

[82] J. Portilla and E. P. Simoncelli. Image denoising via adjustment of wavelet coefficient magnitude correlation. *Proceedings of the 7th ICIP*, pages 10–13, 2000. 33, 81

[83] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40(1):49–70, 2000. 33

[84] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli. Adaptive wiener denoising using a Gaussian scale mixture model in the wavelet domain. *International Conference on Image Processing*, 2:37–40, 2001. 33, 35, 65, 66, 69, 70, 72, 81

[85] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli. Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain. pages 37–40, 2001. 33, 81

[86] W. K. Pratt. *Digital image processing.* A Wiley-Interscience Publication, 2001.

[87] D. L. Ruderman. The statistics of narual images. *Network:computation in Neural Systems*, 5:517 –548, 1994. 65, 72

[88] S. Saha and R. Vemuri. An analysis on the effect of image activity on lossy coding performance. *IEEE Circuits and Systems, 2000. Proceedings*, 3:295 – 298, May 2000.

[89] Z. M. P. Sazzad and Y. Horita. Image quality assessment models for JPEG and JPEG2000 compressed color images. *3rd European Conference on Colour in Graphics, Imaging, and Vision*, pages 478–483, 2006. 15, 114, 115, 123

[90] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita. Spatial features based no reference image quality assessment for JPEG2000. *Image Processing, IEEE International Conference on*, 3:517–520, Oct. 2007. 22, 27, 28, 95, 107

[91] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita. No reference image quality assessment for JPEG2000 based on spatial features. *Signal Processing: Image Communication*, 23(4):257–268, 2008. 22, 95

[92] Z. M. P. Sazzad, M. Sato, Y. Kawayoke, and Y. Horita. No-reference image quality evaluation based on local features and segmentation. *The Journal of the Institute of Image Electronics Engineers of Japan*, 37(3):335–345, 2008. 21

[93] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948. 59

[94] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IEEE*, 72:1192–1201, 1984. 63, 138

[95] C. E. Shannon and W. Weaver. The mathematical theory of communication. 1949.

[96] H. R. Sheikh, A. C. Bovick, and L. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11):1918–1927, 2005. 7, 21, 22, 27, 28, 107

[97] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006. 7, 19, 25, 27, 28, 44, 65, 72, 88, 92, 102

[98] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Blind quality assessment for JPEG2000 compressed images. *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, 2:1735–1739, November 2002. 7, 27, 95, 114, 115

[99] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik. Live image quality assessment database release 2. 2005. 3, 24, 26, 27, 28, 41, 42, 43, 44, 47, 55, 73, 75, 76, 77, 90, 94, 95, 107, 111, 113, 114, 116, 122, 124

[100] H.R. Sheikh, A.C. Bovik, and L. Cormack. Blind quality assessment of jpeg2000 compressed images using natural scene statistics. *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2:1403–1407 Vol.2, Nov. 2003. 88, 95

[101] H.R. Sheikh, A.C. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *Image Processing, IEEE Transactions on*, 14(12):2117–2128, Dec. 2005. 32

[102] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *Image Processing, IEEE Transactions on*, 15(11):3440–3451, Nov. 2006. 27

[103] H.R. Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. Blind quality assessment for jpeg2000 compressed images. *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, 2:1735–1739 vol.2, Nov. 2002. 88, 95, 107

[104] E. P. Simoncelli and W. T. Freeman. The steerable pyramid: a flexible architecture for multi-scalederivative computation. *Image Processing, 1995. Proceedings., International Conference on*, 3:444–447, 1995. 32, 79, 87, 90, 160

[105] Eero P. Simoncelli and William T. Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. pages 444–447, 1995.

[106] C. Staelin and H. Nachlieli. Image noise estimation using color information. *HPL-2005-2R1*, 2007. 43, 44, 96, 98, 99

[107] G. Strang. Wavelet transforms versus Fourier transforms. *Bull. Amer. Math. Soc.*, 23, 1993. 31

[108] C. Studholme, D. Hawkes, and D. Hill. A normalized entropy measure for multimodality image alignemet. *SPIE Conference on Image Processing*, 3338:132 – 143, February 1998.

[109] D. Tao, X. Li, W. Lu, and X. Gao. Reduced-reference IQA in contourlet domain. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on : Accepted for future publication*, May 2009. 7

[110] P. C. Teo and D. J. Heeger. Perceptual image distortion. *ICIP (2)*, pages 982–986, 1994. 7, 17

[111] H. H. Tong, M. J. Li, H. J. Zhang, and C. S. Zhang. No-reference quality assessment for JPEG2000 compressed images. *Image Processing, International Conference on*, 5:3539–3542, Oct. 2004. 7, 15, 23, 24

[112] H.H. Tong, M.J. Li, H.J. Zhang, C.S. Zhang, J.R. He, and W.Y. Ma. Learning no-reference quality metric by examples. *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 247–254, Janurary 2005.

[113] S. Tourancheau, F. Autrusseau, Z. M. P. Sazzad, and Y. Horita. Impact of subjective dataset on the performance of image quality metrics. *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 365–368, October 2008. 23

[114] P. E. Trahanias, D. Karakos, and A. N. Venetsanopoulos. Directional processing of color images: Theory and experimental results. *IEEE Transtions on Image Process.*, 5(6):868–880, June 1996. 8, 9

[115] D. Tsai, Y. Lee, and E. Matsuyama. Information entropy measure for evaluation of image quality. *Journal of Digital Imaging*, 21(3), 2008. 7, 10

[116] S. Uchida and T. DU Yih. Evaluation of radiographic images by entropy : Application to development process. *Japanese journal of applied physics*, 17(11):2029–2034, 19781105.

[117] Weken D. V., Q. Yu, C.N. Zhang, and H. Li. Image quality assessment using rough fuzzy integrals. *Distributed Computing Systems Workshops - Supplements, 2007. ICDCSW 2007. 27th International Conference on*, pages 1–5, June 2007. 7

[118] D. Van der Weken, M. Nachtegael, and E. Kerre. Improved image quality measures using ordered histograms. *Multimedia Signal Processing, 2004 IEEE 6th Workshop on*, pages 67–70, Sept.-1 Oct. 2004. 7

[119] R. Venkatesh Babu, S. Suresh, and A. Perkis. No-reference JPEG-image quality assessment using GAP-RBF. *Signal Process.*, 87(6):1493–1503, 2007. 7, 15, 23

[120] B. A. Wandell. *Foundations of vision.*, 1995.

[121] Z. Wang and A. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81 – 84, March 2002. 4, 6, 7, 18, 25

[122] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error measurement to structural similarity. *IEEE Transactions on Image Processing*, 13, 2004. 4, 6, 7, 18, 24, 25, 27, 28, 92

[123] Z. Wang and A. C. Bovik. Embedded foveation image coding. *Image Processing, IEEE Transactions on*, 10(10):1397–1410, Oct 2001. 7

[124] Z. Wang, A. C. Bovik, and B. Evans. Blind measurement of blocking artifacts in images. *International Conference on Image Processing*, 3:981–984, 2000. 7, 15, 20, 27, 28, 88, 95, 123

[125] Z. Wang, A. C. Bovik, and L. G. Lu. Why is image quality assessment so difficult. *Acoustics, Speech, and Signal Processing, 2002. Proceedings. (ICASSP '02). IEEE International Conference on*, 4:IV–3313–IV–3316 vol.4, 2002. 7

[126] Z. Wang, A. C. Bovik, and E. P. Simoncelli. Structural approaches to image quality assessment. *Handbook of Image and Video Processing*, 2005. 7, 18

[127] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet-domain natural image statistic model. *Human Vision and Electronic Imaging X*, 5666:149 – 159, January 2005. 7, 95

[128] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. *IEEE Asilomar Conference on Signals, Systems and Computers*, Nov 2003. 27, 92

[129] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. *Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on*, 2:1398–1402 Vol.2, Nov. 2003. 19

[130] A. B. Watson. Dctune: A technique for visual optimization of DCT quantization matrices for individual images. *Society for Information Display Digest of Technical Papers*, XXIV:946–949, 1993. 4, 6, 7, 9, 12, 14, 17, 24, 25, 28, 61

[131] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor. Visibility of wavelet quantization noise. *Image Processing, IEEE Transactions on*, 6(8):1164–1175, Aug. 1997. 16

[132] D. V. Weken, M. Nachtegael, and E. E. Kerre. Using similarity measures and homogeneity for the comparison of images. *Image and Vision Computing*, 22:695–702, Aug. 2004. 7

[133] B. Widrow and Kollar I. *Quantization noise*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2008. 109

[134] G.W. Wornell. Wavelet-based representations for the 1/f family of fractal processes. *Proceedings of the IEEE*, 81(10):1428–1450, Oct 1993. 37

[135] G.W. Wornell and A.V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *Signal Processing, IEEE Transactions on*, 40(3):611–623, Mar 1992. 37

[136] G.W. Wornell and A.V. Oppenheim. Wavelet-based representations for a class of self-similar signals with application to fractal modulation. *Information Theory, IEEE Transactions on*, 38(2):785–800, Mar 1992. 37

[137] H. R. Wu and M. Yuen. A generalized block-edge impairment metric for video coding. *IEEE Signal Processing Letters*, 4(11):317 – 320, November 1997. 7

[138] H. R. Wu and M. Yuen. A generalized block-edge impairment metric for video coding. *Signal Processing Letters, IEEE*, 4(11):317–320, Nov 1997. 21

[139] Q. J. Yu and S. L. Xie. An image quality assessment method based on fuzzy inference rules. *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, 1:702–705, Dec. 2004. 7

[140] D. Zhang. *Information theoretic criteria for image quality assessment based on natural scene statistics*. MASc Thesis, University of Waterloo, 2006.

[141] D. Zhang. *INFORMATION THEORETIC CRITERIA FOR IMAGE QUALITY ASSESSMENT BASED ON NATURAL SCENE STATISTICS*. PhD Thesis, Department of Systems Design Engineering, University of Waterloo, 2009.

[142] D. Zhang and Ed Jernigan. An information theoretic criterion for image quality assessment based on natural scene statistics. *International Conference on Image Processing*, pages 2953–2956, September 2006. 64