INFORMATION THEORETIC CRITERIA FOR IMAGE QUALITY ASSESSMENT BASED ON NATURAL SCENE STATISTICS

by

Di Zhang

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2006

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.

Zhang, Di

Abstract

Measurement of visual quality is crucial for various image and video processing applications.

The goal of objective image quality assessment is to introduce a computational quality metric that can predict image or video quality. Many methods have been proposed in the past decades. Traditionally, measurements convert the spatial data into some other feature domains, such as the Fourier domain, and detect the similarity, such as mean square distance or Minkowsky distance, between the test data and the reference or perfect data. However only limited success has been achieved. None of the complicated metrics show any great advantage over other existing metrics.

The common idea shared among many proposed objective quality metrics is that human visual error sensitivities vary in different spatial and temporal frequency and directional channels. In this thesis, image quality assessment is approached by proposing a novel framework to compute the lost information in each channel not the similarities as used in previous methods. Based on natural scene statistics and several image models, an information theoretic framework is designed to compute the perceptual information contained in images and evaluate image quality in the form of entropy.

The thesis is organized as follows. Chapter I gives a general introduction about previous work in this research area and a brief description of the human visual system. In Chapter II statistical models for natural scenes are reviewed. Chapter III proposes the core ideas about the computation of the perceptual information contained in the images. In Chapter IV, information theoretic criteria for image quality assessment are defined. Chapter V presents the simulation results in detail. In the last chapter, future direction and improvements of this research are discussed.

Acknowledgments

I would like to thank all the many people that have made the writing of this thesis possible. I would first like to thank Dr. Ed Jernigan, for the encouragement and support he provided. I would also like to thank Dr. David A. Clausi and Dr. Daniel W. Stashuk for being the readers of my thesis.

To the many friends I have made, I give thanks for their part in making the MAsc. experience both pleasurable and memorable; special thanks go to the members of the VIP group. To Mom, Dad, I am really grateful for the birth you gave me and your endless support for me to pursue my interest.

Finally, I acknowledge with thanks the financial support I received over the course of my graduate program from NSERC and the Faculty of Engineering of the University of Waterloo.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

1.1 Motivation for Image Quality Assessment

A great effort has been made over many years to improve computational methods to evaluate the quality of images. Generally speaking, there are three areas in which image quality metrics play an important role.

First, metrics can be used to monitor an image acquisition system. In most storage systems designed for images, compression technologies are used to save memory. For instance, digital cameras use lossy compression technology (JPEG) for saving memory. In other words, a captured image will be compressed and then saved to a memory card. The compression process will cause some details in the original image to disappear. Consequently, if more compression is applied to an image, more space is saved but the compressed image has lower quality. For critical work that requires higher image quality, less compression should be used. One problem here is how to evaluate the compressed image quality.

Second, image quality assessment is needed to compare the results of different image processing systems. Suppose several algorithms may be considered for a specific task. Image quality assessment can select the algorithms with the best performance.

Third, a metric can be embedded into image processing systems to optimize the algorithm and configuration files, especially for real-time applications. For example, in digital television broadcasting networks, image quality assessment technologies are embedded to ensure a 'user-oriented' quality of service. In this case, a parametric model is needed for an image quality metric that relies on the evaluation of characteristic coding and transmission impairments with a set of features. This

allows assessment of the approach and future developments for quality of service monitoring in digital television.

The best way to evaluate the quality of images is to ask several viewers, because human beings are the ultimate viewers in most image applications. A mean opinion score (MOS) of the human judgments is computed as the subjective evaluation. This method has been used for several years, however it also has an obvious limitation. Subjective video quality measurement usually is costly and time-consuming, and requires many human viewers to obtain statistically meaningful results.

The goal is to approach the quality assessment(QA) problem by proposing computational algorithms to predict the quality of images or video. Several methods have been presented in the past years. The following section reviews the basic concepts of the previous methods and assesses their performances.

1.2 Subjective Image Quality Assessment

Presently, the most reliable judgment of image quality assessment is subjective rating by human observers. Typically, the observer group includes "nonexpert" observers and "expert" observers. A nonexpert viewer may pay more attention on the overview, but a trained viewer may focus on the details.

There are two common kinds of subjective evaluation: full reference judgment and blind judgment. In the former case, observers are provided standard reference images that act as the calibrations in the subjective judgment. In the later case, observers have to judge the image quality based on the test image without the comparison.

Subjective testing for visual quality assessment has been finalized in ITU-R Recommendation BT.500-10 (2000) [1].The three most commonly used procedures are the following:

◇ Double Stimulus Continuous Quality Scale (DSCQS) Method[1]: "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)

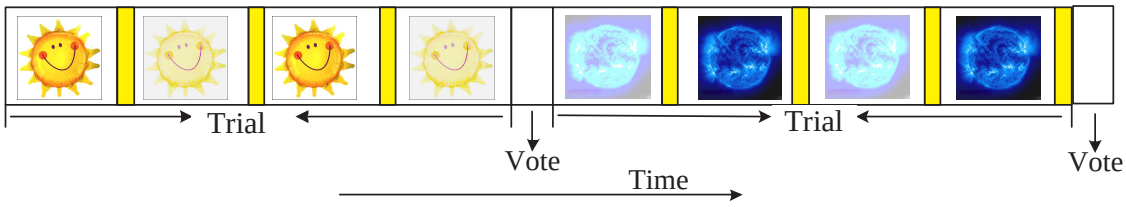


Figure 1.1: Presentation sequence for the DSCQS test method

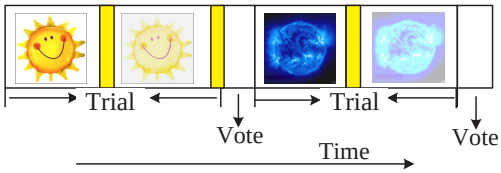


Figure 1.2: Rating scales for a trial with the DSCQS method

◇ Double Stimulus Impairment Scale (DSIS) Method[1]: The DSIS experiment procedure is illustrated in Figure 1.3. 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.

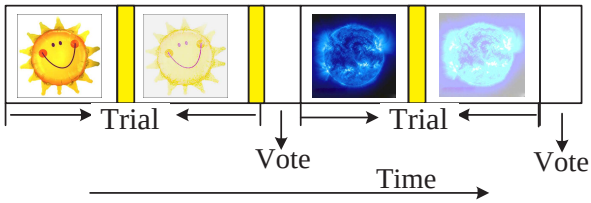


Figure 1.3: Presentation sequence for the DSIS test method

◇ Single Stimulus Continuous Quality Evaluation (SSCQE)[1]: 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".

In the subjective evaluation procedure, there are several points that should be emphasized. First of all, the number of the observers should be large enough to

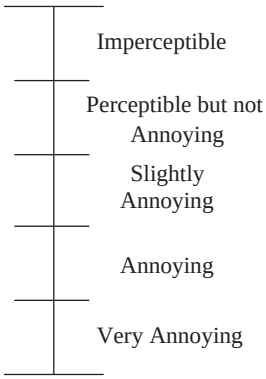


Figure 1.4: Rating scales for a trial with the DSIS method

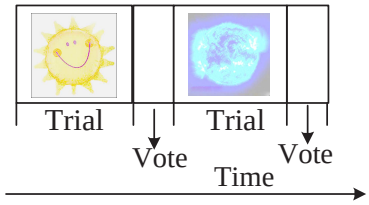


Figure 1.5: Presentation sequence for the SSCQE test method

ensure statistical confidence in image quality experiments, according to the Cramer-Rao bound. Second, the results of subjective testing are influenced by the experimental conditions, which means the experimental viewing conditions should be designed to match the viewing conditions in practice as closely as possible. In addition, there are a number of other procedural elements worth noting, such as: choice of observers, reference conditions, signal sources for the test scenes, timing of the presentation of the various test scenes, selection of a range of test scenes.

Generally speaking, subjective testing has the following weaknesses:

(1) a wide variety of possible methods and experiment parameters must be considered;

(2) necessary experiment controls are required;

(3) numerous observers must be selected;

(4) the experiment complexity makes the whole procedure very time-consuming.

The database set used in this thesis is the subjective judgment result from Laboratory for Image and Video Engineering, University of Texas, Austin [2]. Twenty-nine high-resolution 24-bits/pixel RGB color images (typically 768×512) were distorted using five distortion types: JPEG2000, JPEG, white noise, Gaussian blur,

and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model. A database was produced from the 29 reference images so that each reference image had test versions with each distortion type and different distortion level. Observers were asked to provide their perception of quality on a continuous linear scale that was divided into five equal regions marked with adjectives "Unsatisfactory", "Poor", "Fair", "Good" and "Excellent". 20-29 human observers rated each image. Each distortion type was evaluated by different subjects in different experiments using the same equipment and viewing conditions. In total, 982 images are evaluated by test viewers. Finally, a Mean Opinion Score (MOS) value for each distorted image was computed as a quantitative validation tool for objective image quality assessment.

1.3 Classical Objective Image Quality Assessment

Typically, image and video QA algorithms may be classified into three broad categories:

Full-Reference (FR) QA methods, in which the QA algorithm has access to a 'perfect version' of the image or video which acts as the reference during the quality evaluation procedure. The reference image generally comes from a high-quality acquisition device.

No-Reference (NR) QA methods, in which the QA algorithm has the access only to the distorted signal and must estimate the quality of the signal without any knowledge of the reference image. Since NR methods do not require any reference information, they can be used in any application where a quality measurement is required.

Reduced-Reference (RR) QA methods, in which partial information regarding the reference image is available. QA algorithms use this partial reference information to judge the quality of the distorted signal.

1.3.1 Distance Metrics

Generally speaking, in the area of full reference quality assessment, each method defines its own distance metric in a specific domain. Numerous distance metrics have been presented which can be divided into two major groups: those based on the statistical properties of the data (statistical distance metrics) and those without considering statistics (algebraic distance metrics).

Assume for each pixel or segmentation, a corresponding feature vector is F ($F = [F(1), F(2), F(3), ..., F(N)]^T$) and the distance between the feature vector X of a test image and the feature vector Y of the reference is defined as:

a. Algebraic Distance Metrics

1) Minkowsky metrics

This measurement is widely used:

$$D = \{\frac{1}{N}\sum_{i=1}^{N}|X(i) - Y(i)|^{\gamma}\}^{\frac{1}{\gamma}} \tag{1.1}$$

If $\gamma = 2$, the well-known mean square distance:

$$D_1 = \{\frac{1}{N}\sum_{i=1}^{N}|X(i) - Y(i)|^{2}\}^{\frac{1}{2}} \tag{1.2}$$

If $\gamma = 1$, the mean absolute distance:

$$D_2 = \frac{1}{N}\sum_{i=1}^{N}|X(i) - Y(i)| \tag{1.3}$$

If $\gamma = \infty$, the Minkowsky metric reduces to maximum distance measurement:

$$D_3 = \max_{i}|X(i) - Y(i)| \tag{1.4}$$

2) Weighted Minkowsky metrics

$$D = \{\frac{1}{N}\sum_{i=1}^{N}\omega_i|X(i) - Y(i)|^{\gamma}\}^{\frac{1}{\gamma}} \tag{1.5}$$

where ω_i is the weight for each element in feature vector.

If $\gamma = 2$, the well-known weighted mean square distance:

$$D_4 = \{\frac{1}{N}\sum_{i=1}^{N}\omega_i|X(i) - Y(i)|^{2}\}^{\frac{1}{2}} \tag{1.6}$$

3) Moments of the angles [3]

$$D_5 = \frac{2}{\pi} cos^{-1} \frac{X \odot Y}{||X|| \cdot ||Y||} \tag{1.7}$$

$$D_6 = [\frac{2}{\pi} cos^{-1} \frac{< X, Y >}{||X|| \cdot ||Y||} + 1][\frac{||X - Y||}{\max(||X||, ||Y||)} + 1] - 1 \tag{1.8}$$

b. Statistical Distance Metrics

If $p_X$is the probability density function of the feature X, $p_Y$is the probability density function of the feature Y, and $p_{XY}$is the joint distribution of the feature (X, Y)

1) Kullback-Leibler divergence (relative entropy) [4]:

$$D_7 = \int_{x \in \mathcal{X}} p_X(x) log[\frac{p_Y(x)}{p_X(x)}] dx \tag{1.9}$$

2) Information theoretic distortion measures [5]

$$D = g\{E_{p_X}[f(\frac{p_Y(y)}{p_X(x)})]\} \tag{1.10}$$

E_{p_X} is the expectation with respect to p_X

Hellinger distance:

$$\begin{aligned} f(x) &= (\sqrt{x} - 1)^2 \\ g(x) &= \frac{1}{2} x \end{aligned}$$

Substitute f(x) and g(x) into Formula (1.10), Hellinger distance will be derived.

$$\begin{aligned} D_8 &= \frac{1}{2} \int_{x \in \mathcal{X}} (\sqrt{\frac{p_Y(x)}{p_X(x)}} - 1)^2 dx \\ &= \frac{1}{2} \int_{x \in \mathcal{X}} (\frac{\sqrt{p_Y(x)} - \sqrt{p_X(x)}}{\sqrt{p_X(x)}})^2 dx \\ &= \frac{1}{2} \int_{x \in \mathcal{X}} \frac{(\sqrt{p_Y(x)} - \sqrt{p_X(x)})^2}{p_X(x)} dx \\ &= \frac{1}{2} \int_{x \in \mathcal{X}} (\sqrt{p_Y(x)} - \sqrt{p_X(x)})^2 dx \end{aligned}$$

Generalized Matusita distance: Substitute f(x) and g(x) into Formula (1.10), Generalized Matusita distance will be derived.

$$\begin{aligned} f(x) &= |1 - x^{1/r}|^r \\ g(x) &= x^{1/r} \\ D_9 &= \sqrt[r]{\int_{x\in\mathcal{X}} |p_Y(x)^{1/r} - p_X(x)^{1/r}|^r dx} \end{aligned} \tag{1.11}$$

3) Mutual Information [6]

$$D_{10} = I(X,Y) = \int_{x\in\mathcal{X},y\in\mathcal{Y}} -p_{XY}(x,y) log(\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}) dxdy \tag{1.12}$$

4) Normalized Mutual Information Measure [7]

$$\begin{aligned} D_{11} &= \frac{H(X)+H(Y)}{H(X,Y)} \\ H(X) &= \int_{x\in\mathcal{X}} -p_X(x) log(p_X(x)) dx \\ H(Y) &= \int_{x\in\mathcal{X}} -p_Y(x) log(p_Y(x)) dx \\ H(X,Y) &= \int_{x\in\mathcal{X},y\in\mathcal{Y}} -p_{XY}(x,y) log(p_{XY}(x,y)) dxdy \end{aligned} \tag{1.13}$$

5) Entropy correlation coefficient [8]

$$\begin{aligned} D_{12} &= \frac{2I(X,Y)}{H(X)+H(Y)} \\ &= \frac{\int_{x\in\mathcal{X},y\in\mathcal{Y}} -p_{F_1F_2}(x,y) log(\frac{p_{F_1F_2}(x,y)}{p_{F_1}(x)p_{F_2}(y)}) dxdy}{\int_{x\in\mathcal{X}} -p_X(x) log(p_X(x)) dx + \int_{x\in\mathcal{X}} -p_Y(x) log(p_Y(x)) dx} \end{aligned} \tag{1.14}$$

6) Weighted relative entropy [9]

$$D_{13} = A\int_{x\in\mathcal{X}} p_X(x) log(\frac{p_Y(x)}{p_X(x)}) dx + B\int_{x\in\mathcal{X}} p_Y(x) log(\frac{p_X(x)}{p_Y(x)}) dx \tag{1.15}$$

7) Kolmogorov-Smirnov distance [10]

$$D_{14} = \max_{x \in \mathcal{X}} \{ |\mathbf{p}_X^i(x) - \mathbf{p}_Y^i(x)| \} \tag{1.16}$$

where $\mathbf{p}_X^i(x) = \int_{t<x} p_X^i(t)dt$,$p_X^i(t)$ is the distribution of ith component of feature X

8) Squared Euclidean distance between the CDFS

$$D_{15} = \int_{x \in \mathcal{X}} |\mathbf{p}_X^i(x) - \mathbf{p}_Y^i(x)|^2 dx \tag{1.17}$$

9) χ^2 -statistic [11]

$$D_{16} = \int_{x \in \mathcal{X}} \frac{p_X(x) - \hat{p}(x)}{\hat{p}(x)} dx \tag{1.18}$$

where $\hat{p}(x) = [p_X(x) + p_X(x)]/2$

10) Jeffrey-divergence [12]

$$D_{17} = \int_{x \in \mathcal{X}} p_X(x) log(\frac{p_Y(x)}{\hat{p}(x)}) dx + \int_{x \in \mathcal{X}} p_Y(x) log(\frac{p_X(x)}{\hat{p}(x)}) dx \tag{1.19}$$

where $\hat{p}(x) = [p_X(x) + p_X(x)]/2$

In Table 1.1, the performance of the different distance metrics is listed. During the simulation, the feature vector is simply defined as the value of each pixel in the test image and the reference image.

From Table 1.1, the performance of the objective judgments varies a lot. Although some distance metrics have poor correspondence to human judgment under the simulation condition listed above, according to references [3][4][5][6][7][8], these distance metrics will achieve much better results under more complicated assumptions and simulation conditions. The details of the implementation of the above metrics are listed in Appendix.

1.3.2 Image Quality Metrics

Hundreds of image quality metrics [3],[4],[5],[6],[7],[8],[9],[10],[11],[12] have been proposed. It is difficult to compare the performances of different criteria without the implementation detail mentioned in the literature. It is more useful to give a general

	Gaussian Blur	White Noise	J2k Distortion
Minkowsky Distance (D_1)	0.8248	0.8394	0.8739
Minkowsky Distance (D_2)	0.8754	0.8605	0.8901
Minkowsky Distance (D_3)	0.9301	0.9303	0.9351
Moments of Angels (D_5)	0.8570	0.8981	0.8684
Moments of Angels (D_6)	0.8640	0.8807	0.8592
Relative Entropy (D_7)	0.3091	0.7576	0.5172
Matusita (D_9)	0.6728	0.9020	0.5528
Mutual Information (D_{10})	0.8704	0.7073	0.9227
Entropy Correlation (D_{12})	0.8741	0.6558	0.9325
Normalized Mutual Information (D_{11})	0.8480	0.8831	0.9088
Kolmogorov Distance (D_{14})	0.5017	0.8763	0.6584
SquaredCDF (D_{15})	0.4204	0.6729	0.4303
Xstatistics (D_{16})	0.3402	0.4307	0.1469

Table 1.1: Performance of different distance metrics (Linear Correlation between Subjective Judgments and Objective Judgments)

description of the development process for the full reference image quality assessment algorithm than to just list certain criteria.

Figure 1.6 presents a graphical depiction of the development process. The library of test scenes provides a set of images which exhibits various amounts of spatial information content. In the following design, these images serve as the references. The impairment generator produces the degraded images with different types of distortions, such as blur impairment, noise injection and compression degradation. Before the algorithm design, the subjective test is an essential step to get the human judgment in order to validate the quality metric. Then the algorithm designer proposes the models to compute the distortion between the test image and the reference. The analyses of the correlation between the distortion and the corresponding objective feedback the information to the parameters setting (the dash line). This optimal set of parameters is then used to develop a quality assessment algorithm that gives objective results that agree closely with human results.

Among the present quality metrics, there are many measurements which share a human visual system (HVS) paradigm. These measurements aim to quantify the strength of the errors between the reference and the distorted image by mimicking the human eye's behaviors. Figure 1.7 shows a generic quality assessment framework that is based on HVS modeling. Most quality assessment algorithms that model the HVS can be explained with this framework [14], although they may differ in the specifics.

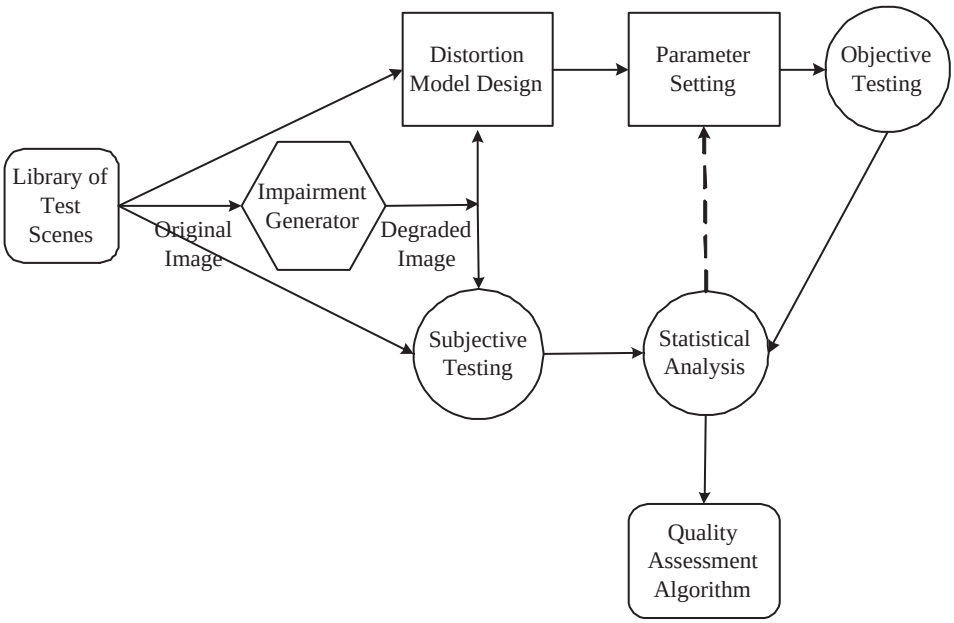


Figure 1.6: Development process for image quality assessment algorithm

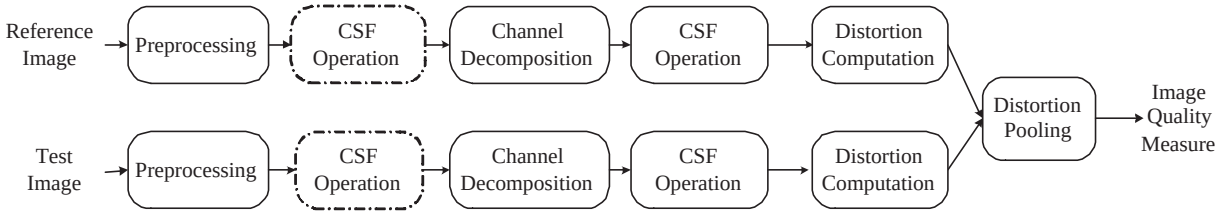


Figure 1.7: HVS Based Image Quality Assessment Algorithm

The first step of HVS based image quality assessment algorithm is a preprocessing stage, which may contain alignment and transformations of color spaces. The purpose of the alignment operation is to eliminate the misalignment due to various reasons such as unstable transmission or very low bit compression. In this case, it is necessary to establish the point-to-point correspondence between the reference and test image. The color space transformation is typically used in order to convert the signal into another color space which has better conformance to HVS, for instance the RGB space to the HSV space.

A Contrast Sensitivity Function (CSF) is broadly applied in numerous measures. The basic idea of CSF comes from the variation in the sensitivity of the HVS to the visual stimulus in the different spatial and temporal frequency domains. CSF may be implemented before the channel decomposition using linear filters to approximate the frequency responses. Meanwhile some metrics choose to implement CSF as a weighting for channels after the channel decomposition.

Quality metrics commonly model the frequency selective channels in the HVS with different transform tools. The channels serve to separate the visual stimulus into different spatial and temporal subbands. Some quality assessment measures use Gabor filters, some criteria employ different wavelet families, and some other simple methods even implement the discrete cosine transform.

Distortion computation is typically implemented within each channel. Numerous distance metrics (listed in section 1.3.1) can be defined in this step to compute the distortion between the reference signal and the test signal in each channel.

Distortion pooling is the process of combining the distortion from different channels into a single distortion or quality interpretation. For most quality assessment methods, the following formula is used:

$$C = (\sum_{j} |D_j|^{\beta})^{1/\beta} \tag{1.20}$$

where D_j is the distortion computed in the ith channel. β is a constant typically with a value between 1 and 4. This form of distortion pooling is commonly called Minkowski distortion pooling.

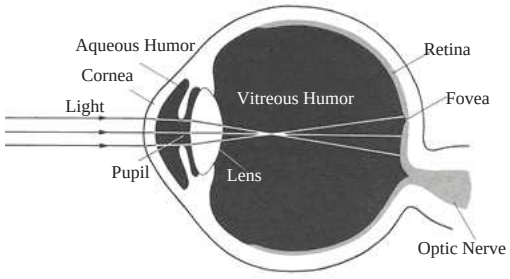
1.4 Human Visual System

1.4.1 Structure of the Human Visual System

"Vision is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information." (Marr and Nishihara) The human visual system (HVS) can be divided into two major components: the eyes, which capture light and convert it into bio-signals, and the visual pathways in the brain, a set of neural systems in which these signals are transmitted and processed. Because of the complexity of the HVS, the following section briefly describe the components and corresponding functions [15].

a. Structure of human eye

For most animals on the planet, the eye is the organ to perceive visual signals. Roughly speaking, the human eye has the same structure as a camera. As illustrated in Figure 1.8 [16], the optical system of the human eye is composed of the cornea, the aqueous humor, the lens, and the vitreous humor. The light first passes through the refractive system that contains the cornea and the lens. The light is focused on the retina, the organ that converts the light signals into bio-signals.



The volume of the eye is around 6.5 cm^3.
The diameter is about 24 mm and the weight is approximately 7 gram.

Figure 1.8: Human eye structure[16]

Generally, our human optical imaging system follows the *Gaussian lens formula*:

$$\frac{1}{d_s} + \frac{1}{d_i} = \frac{1}{f} \tag{1.21}$$

where d_s is the distance between the source and the lens, d_i is the distance between the image and the lens, and f is the focal length of the lens.

The retina connects the human eye and visual pathways. There are five main kinds of cells in the retina in Figure 1.9 [17]:

(1)Photoreceptors (rods & cones): rods are responsible for low intensity vision; cones are responsible for high intensity vision;

(2)Horizontal cells: allow for lateral interactions

(3)Amacrine cells: influence temporal characteristics of the ganglion cells' responses;

(4)Bipolar cells: connect photoreceptors to retinal ganglion cells;

(5)Retinal ganglion cells: form the optic nerve.

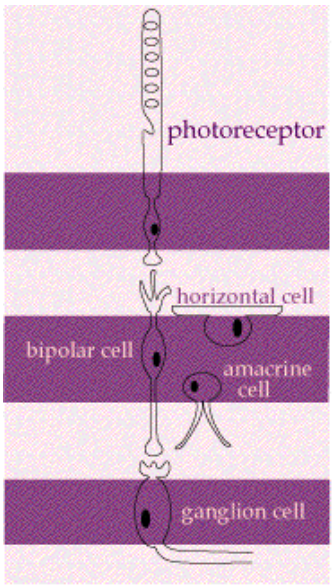


Figure 1.9: The structure of retina [17]

b. Visual pathways

Figure 1.10 [18] shows the optical path in the visual system. The optic nerves from the retina meet at the optic chiasm, in which the fibers are rearranged. All the fibers from the nasal halves of each retina cross to the opposite side, where they join the fibers from the temporal halves of the opposite retinas to form the optic tracts. Since the retinal images are reversed by the optics, the left visual field is thus processed in the right hemisphere, and the right visual field is processed in the left hemisphere. Most of the fibers from each optic tract synapse in the lateral geniculate nucleus (LGN). From there fibers pass by way of the optic radiation to the visual cortex. Throughout these visual pathways, the neighborhood relations of the retina are preserved, i.e. the input from a certain small part of the retina is processed in a particular area of the LGN and of the primary visual cortex.

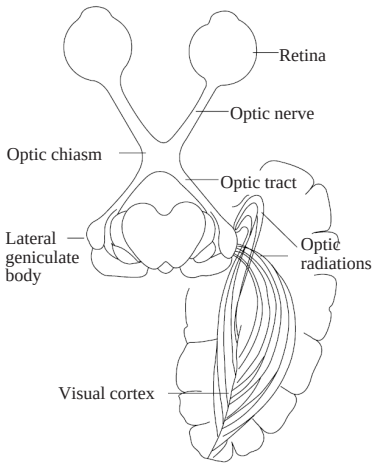


Figure 1.10: The optical path in the visual system[18]

1.4.2 Psychophysical HVS Features

An intuitive method to design an efficient evaluation algorithm for image quality is to mimic the human visual system. Understanding the psychophysical properties of the HVS and its corresponding mathematical models will be helpful. The following section will briefly discuss the HVS features. And some of which will play important roles in the algorithm presented later.

Foveal and Peripheral Vision

The densities of the cone cells and the ganglion cells are not uniform across the retina. We have the peak resolution at the fovea and rapidly decreasing low resolution away from the fovea. The high spatial resolution vision due to fixation on a region is called foveal vision, while the lower resolution vision is called peripheral vision. Most image quality assessment algorithms employ foveal vision.

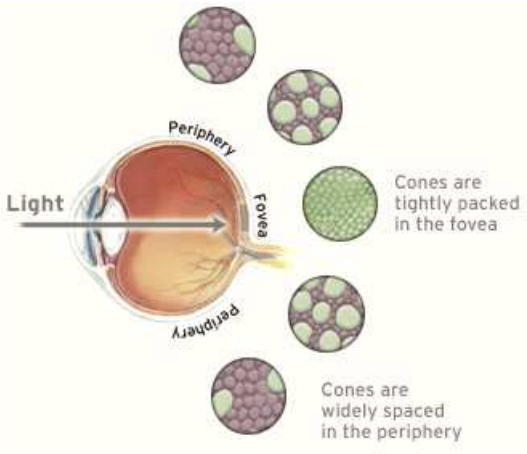


Figure 1.11: Foveal and Peripheral Vision [48]

Light Adaptation and Dark Adaptation

One of the important properties of the human visual system is the adaptation to a large range of light levels.

Light Adaptation: This phenomenon occurs when we move from the dark into bright light. Because the sensitivity of the receptors is set to dim light, when human eyes are exposed to a bright light environment, rods and cones are both stimulated. In the following several seconds, adaptation occurs in two ways: the

sensitivity of the retina decreases dramatically; retinal neurons rapidly inhibit rod function and arouse the cone system.

Dark Adaptation: Dark adaptation is the reverse of light adaptation. It occurs when we go from a bright light area to a dark area. Initially human eyes perceive nothing but blackness because cones cease to function in low intensity light. Once in the dark, rhodopsin (the pigment sensitive to red light in the retinal rods)is generated, and the sensitivity of the retina increases. The pupil size will also increase to allow more light to reach the retina.

Contrast Sensitivity Function

The contrast sensitivity function (CSF) tells us how sensitive we are to the various spatial frequencies of visual stimuli. If the spatial frequency of visual stimuli is too high we will not be able to recognize the stimulus pattern any more. Imagine an image consisting of vertical black and white stripes, as shown in 1.12[19].

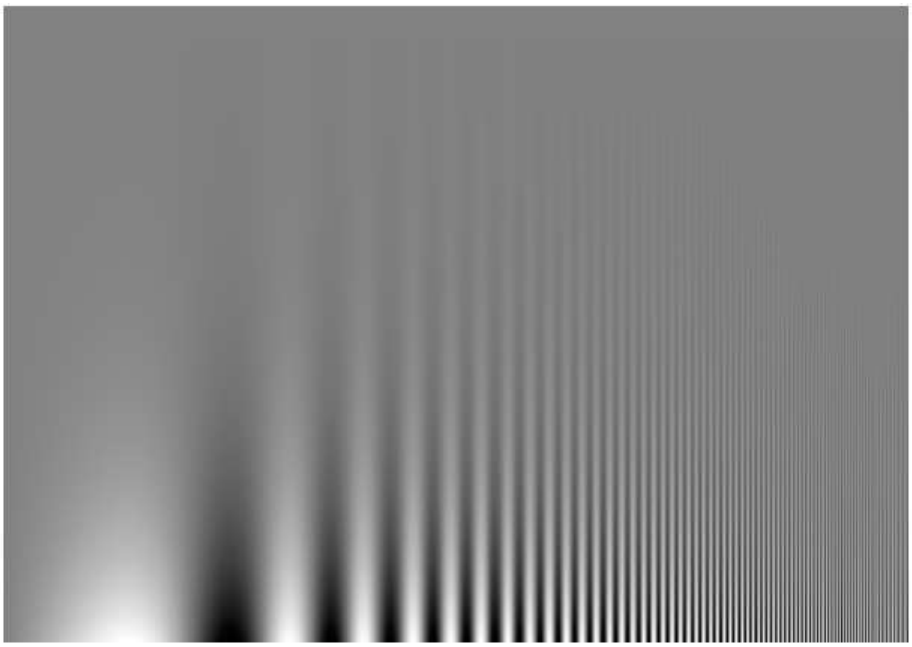


Figure 1.12: Contrast Sensitivity Chart

If the stripes are very thin, we will be unable to see individual stripes. All visible is a gray image. If the stripes then become wider and wider, there is a threshold width beyond which we are able to distinguish the stripes. That threshold width corresponds the cut-off frequency of the human eyes, which means one basic function

of the human visual system is a low pass filter to get rid of the high frequency details of the natural scene.

On the other hand, for each stripe, as the luminance decreases, the lowest detectable luminance varies from low frequency to high frequency. As shown in 1.13, the contrast sensitivity function appears as the envelope of visibility of the modulated pattern. The contrast sensitivity function proposed by Manos and Sakrison [20] is

$$A(f) = 2.6(0.0192 + 0.114f)e^{(-0.114f)^{1.1}} \tag{1.22}$$

f is the spatial frequency of the visual stimuli given in cycles/degree. The function has a peak of value 1 approximately at $f = 8.0$ cycles/degree, and is insignificant for frequencies above 60 cycles/degree. 1.13 shows the contrast sensitivity function A(f).

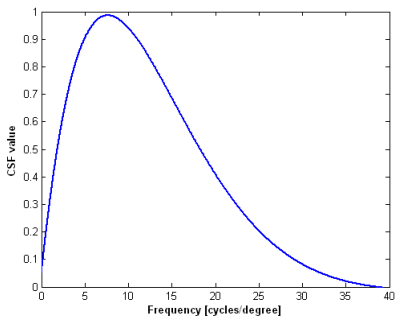


Figure 1.13: Contrast Sensitivity Function

In the algorithm proposed in this thesis, the contrast sensitivity phenomenon is explained as a quantization with the step size defined as the luminance threshold at different frequencies. This quantization process acts as a nonlinear operation in the human visual system model proposed here. The details about the implementation and its benefits will be discussed in Chapter III.

Masking

Masking is a very important phenomenon in vision and image processing, which describes interactions between stimuli. Spatial masking occurs when a stimulus that is visible by itself cannot be detected in the presence of another. Several experiments show that masking is strongest when the interacting stimuli have similar

characteristics, i.e. similar frequencies, orientations, colors, etc. Masking also occurs between stimuli of different orientations or spatial frequency, but such masking is generally weaker. Figure 1.14 gives us a simple explanation of the relationship between the masking and image quality assessment. In the right figure the Gaussian noise is added to the upper part and the same amount noise is implanted into the lower part in the left figure. The noise is masked by the original image acting as the background stimuli. In the first case, the background stimuli is the plain sky, the noise is disturbing in this area. In the second case, the background stimuli are much more complex and unfamiliar, the noise stimuli is less noticeable. Hence the masking results in different subjective quality evaluations even though the noise is same.

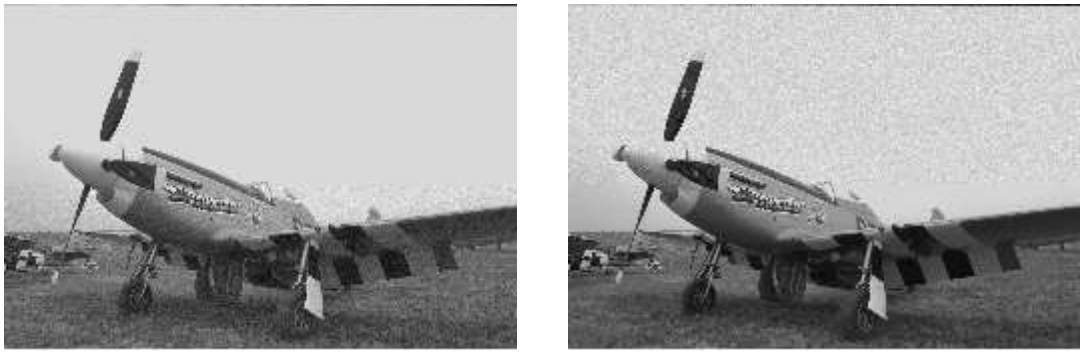


Figure 1.14: Masking and Image Quality Assessment

Temporal Masking: temporal masking is an elevation of visibility thresholds due to temporal discontinuities in intensity, for example scene cuts. This phenomenon plays an important role in video quality assessment.

1.5 Chapter Summary

This chapter starts with the motivation of image quality assessment and its applications in image processing areas. It continues with a literature review about subjective image quality assessment and image quality evaluation algorithms. Finally, this chapter briefly introduced the structure of the human eye, and psychophysical HVS features. The final goal of this thesis is to present an efficient algorithm for the image quality (IQ) task. The intuitive method for us to access IQ is to propose a HVS model to simulate the bio-processing in human brains. Then the first problem we face is how to mathematically represent the natural scenes. In the next chapter, we will discuss how to describe the images statistically with proper models.

Chapter 2

Statistical Modeling of Natural Scenes

Given a photograph, people can distinguish certain textures from others, analyze motions or measure objectives. In general, people abstract visual information from images or videos. How much visual information is contained in one image? Here is a simple game to serve as an inspiration about how to compute perceptual or visual information. Suppose that you are the only person allowed to see a picture, while other people want to know its content. The only way for others to gather information is to ask questions about the image. The game rule says that when you answer the question, the only word that you can say is *YES* or *NO*. For example, "Are there people in the image?" Suppose that finally people use N questions to learn about the image. Intuitively, the perceptual/visual information within the image can be defined as the number of the questions asked. Following that simple idea, how many different images can be described by the N questions and the corresponding answers? 2^N images.

How many question are necessary for people to obtain the content within one image? It greatly depends on the random fluctuations in intensity, color, texture object boundary/shape. The causes for these fluctuations are diverse and complex. The factors can be non-uniform lighting, random fluctuations in object surface orientation and texture, complex scene geometry, or even noise. In Chapter 4, the connection between the statistical properties and perceptual information will be defined and explained in detail.

2.1 Random Field

A random field X is the collection of random variables on a lattice L:

$$X = \{X_i, i \in L\} \tag{2.1}$$

In general, the lattice L can be any collection of discrete points. In this thesis, the shape of lattice L is chosen as a rectangle, the simplest one shown in Figure 2.1:

$$L = \{(i, j) | 1 \leq i \leq N, 1 \leq j \leq M\} \tag{2.2}$$

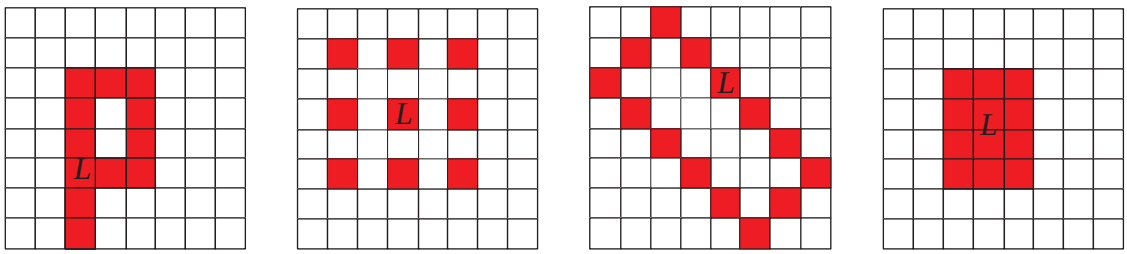


Figure 2.1: The Lattice

The random field can be represented as:

$$X = \{X_{i,j}, (i, j) \in L\} \tag{2.3}$$

The direct way to achieve the distribution of the random field is to compute the probability of each element in X. Suppose we have a set of images with the size of $M \times N$. To describe the random field of images, we need to characterize the joint statistics of $M \times N$. Just consider $M = 256, N = 256$, to describe a random field of 65536 ($M \times N = 65536$) elements will be a cumbersome and time-consuming mission. One important property is introduced by Geman [21] as Markovianity to derive a good approximation of a random field.

2.2 Markov Random Field

The basic idea of Markovianity is conditional independence. To illuminate this idea, let us consider a one dimensional random field, which is characterized by the joint probability function $p\{(X_1, X_2, X_3, ..., X_N) = (x_1, x_2, x_3, ..., x_N)\}, (x_1, x_2, x_3, ..., x_N) \in \mathscr{X}^n$.

Definition 2.1: **A one-dimensional random field is said to be *stationary* if the joint distribution of any subset of the sequences of random variables is invariant with respect to shifts in the time index, i.e.,**

$$\begin{aligned} p\{(X_1, X_2, X_3, ..., X_N) = (x_1, x_2, x_3, ..., x_N)\} = \\ p\{(X_{l+l}, X_{2+l}, X_{3+l}, ..., X_{N+l}) = (x_1, x_2, x_3, ..., x_N)\} \end{aligned} \tag{2.4}$$

for every shift l and for all $x_1, x_2, x_3, ...x_n \in \mathscr{X}$.

Definition 2.2: **A discrete random field** $(X_{l+l}, X_{2+l}, X_{3+l}, ..., X_{N+l})$ **is said to be a *Markov process*, if for** $l, m, n, p = 1, 2, 3, ...$ **and** $p < n$**,**

$$\begin{aligned} p(X_{l+1} = x_{l+1}, ..., X_{l+m} = x_{l+m} | X_{l+m+1} = x_{l+m+1}, ..., X_{n+l}) = \\ p(X_{l+1} = x_{l+1}, ..., X_{l+m} = x_{l+m} | X_{l+m+1} = x_{l+m+1}, ..., X_{p+l}) \end{aligned} \tag{2.5}$$

Definition 2.2 describes a Markov random field with the property that 'the past' is decoupled with 'the future' by knowing 'the now': $p(X_{future}|X_{now}, X_{past}) = p(X_{future}|X_{now}), p(X_{past}|X_{now}, X_{future}) = p(X_{past}|X_{now})$.

Definition 2.3: **The Markov process is said to be *time invariant* if the conditional probability** $p(X_{n+1}|X_n)$ **does not depend on** n**, i.e., for** $n = 1, 2, ...$

$$p\{X_{n+2} = b | X_{n+1} = a\} = p\{X_2 = b | X_1 = a\} \tag{2.6}$$

The above three concepts can be easily extended to two-dimension random fields. The Markovianity for two-dimensional random field can be defined as follows:

Definition 2.4: A two-dimensional random field is said to be a ***Markov Process*** if

$$\begin{aligned} p\{X_i | X_b, X_o\} = p\{X_i | X_b\} \\ p\{X_o | X_b, X_i\} = p\{X_o | X_b\} \end{aligned} \tag{2.7}$$

The natural concepts of 'the past' and 'the future' will be changed by the concepts of 'the boundary', 'the inner' and 'the outer', as shown in Figure 2.2. Definition 2.4 tells us that the knowledge of the boundary X_b decouples the inner X_i and the outer X_o.

The two-dimensional Markov concept above is elegant and intuitive, but still too general to implement in real applications. The simplified version of the above

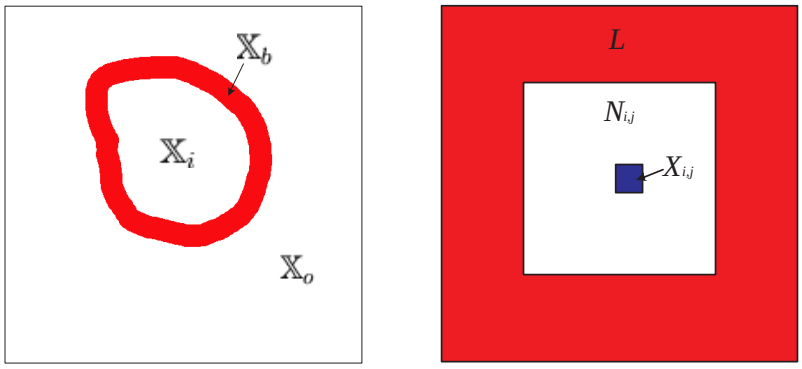


Figure 2.2: Two-dimensional Markov Process

concept is to consider a single element $X_{i,j}$ conditioned on a local neighborhood $N_{i,j}$, as shown in Figure 2.2. The neighborhood $N_{i,j}$ in Figure 2.2 has a noncausal structure, which is more typical and with less limitation compared with the casual structure [22]. Hence in the following discussion, the neighorbood structure is noncasual.

$$sp\{X_{i,j}|X_{k,l},(k,l)\in L,k\neq i,l\neq l\}=p\{X_{i,j}|X_{k,l},(k,l)\in N_{i,j}\} \tag{2.8}$$

The concept of time invariance for one-dimensional Markov random field will be replaced by 'position invariance', which means the conditional probability $p\{X_i|X_b\}$ does not depend on where X_i happens.

2.3 Gaussian-Markov Random Field & Gibbs Random Field

A simplifying assumption of the random field X as a Gaussian distribution leads to a famous statistical image model: Gaussian-Markov Random Field (GMRF) [23]. The conditional density is given by the expression

$$p\{X_{i,j}|X_{k,l},(k,l)\in N_{i,j}\}=\frac{1}{\sqrt{2\pi\sigma^2}}exp\{-\frac{1}{2\sigma^2}[X_{i,j}-\sum_{(k,l)\in N_{i,j}}g_{i,j;k,l}X_{i,j}]^2\} \tag{2.9}$$

where σ is variance of the zero-mean process $X_{i,j}$ and $g_{i,j;k,l}$ are the model parameters. This model is also known as the conditional autoregressive (AR) model. More details are introduced in [23]. The significant advantage of GMRF comes with

the famous property of a Gaussian Process, for which uncorrelatedness means independence. Thus, given the GMRF $X_{i,j}$ with the neighborhood $N_{i,j}$, the Bayesian estimate $E[X_{i,j}|L] = E[X_{i,j}|N_{i,j}]$ becomes a linear expectation:

$$X_{i,j} = \sum_{(k,l)\in N_{i,j}} g_{i,j;k,l} X_{k,l} + \xi_{i,j} \qquad E[X_{k,l}\xi_{i,j}] = 0 \qquad \forall (k,l) \in N_{i,j} \tag{2.10}$$

Gibbs Random Field (GRF) was originally used in statistical physics to study the characteristics of particle interactions, GRF were introduced to the image processing by J. Besag[24]. In the GRFs representation of images, the joint probability distribution of X is given by Gibbs distribution:

$$p\{X = x\} = \frac{1}{Z} e^{-\frac{U(X)}{T}} \tag{2.11}$$

$$U(X) = \sum_{i\in C} V_i(X) \tag{2.12}$$

$$Z = \sum_{X} e^{-U(X)} \tag{2.13}$$

where U is the energy function, C is the set of cliques, V_i is the clique potential, T is the temperature constant.

Definition 2.5: **A clique C is a subset of neighborhood N. Single pixels are also considered cliques. The set of all cliques on a grid is called C.**

Theorem 2.1: **X is an GMRF with respect to the neighborhood system N if and only if X is a Gibbs distribution with respect to C, where C is the set of cliques based on the neighborhood N.**

2.4 Multiscale Transform and Statistics

During the development of the statistical representation for images, the inadequacy of the Gaussian model was apparent. There are several reasons which triggered the development of the multiscale statistical model:

1) Many signals, especially images, have sparse multiscale representation [25];

2) The dependencies between phases of different frequencies in the Fourier domain do not seem to be well captured by the GMRF [26];

3) The property of the multi-resolution transform implies coarse-to fine dependencies among the coefficients [25].

2.4.1 Marginal Models

The many developed statistical model can be divided into two groups: marginal statistical models and joint statistical models. This section reviews several important marginal statistical models, which share a common assumption that the transform coefficients are independent.

1. Double-exponential dustrubution [43]

$$p(w) = \frac{1}{2\sqrt{2\sigma}} e^{-2\sqrt{2\sigma}|w|}$$

2. Generalized Laplacian [44]

$$p(w) = e^{-|w/s|^p} 2\frac{s}{p}\Gamma(1/p)$$

where $p \in [0.5, 0.8]$ and s varies with the scale variances.

3. Mixture of Gaussians [45]

$$p(w) = \sum_{j=1}^{M} p(w|j)p(j)$$

where $\sum_{j=1}^{M} p(j) = 1$ and $p(w|j) = N(\mu_j, \sigma_j)$

4.Mixture of a Gaussian and a point mass function [46]

$$p(w) = kN(\mu, \sigma) + (1-k)\delta(\mu)$$

where $\delta(\mu)$ is a point mass function at μ.

5.Generalized Gaussian distribution (GGD) [47]

$$p(w) = C(\sigma_w, \beta) e^{-[\alpha(\sigma_w,\beta)|w|]^\beta}$$

where $\alpha(\sigma_w, \beta) = \sigma_w^{-1}[\frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}]^{1/2}$ $C(\sigma_w, \beta) = \frac{\beta\alpha(\sigma_w,\beta)}{2\Gamma(1/\beta)}$, and $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$

The above models are proposed to describe the heavy-tailed distribution of the transform coefficients. In Figure 2.3, the original image 'bikes'[2] is decomposed into seven scales. In order to depict the marginal distribution of wavelet coefficients in the third scale and horizontal direction, the coefficients are converted into 32 bins. The solid line is the Gaussian mixture model, and the dashed line is the Gaussian distribution with the same variance as the Gaussian mixture model.

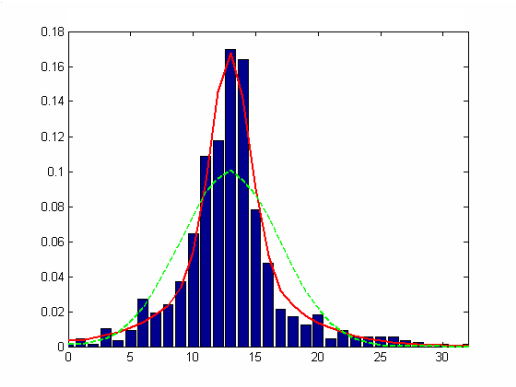


Figure 2.3: Heavy-tailed Distribution

2.4.2 Joint Models

The marginal model is significantly more powerful than the classic Gaussian model. For instance, in denoising, the use of the Generalized Gaussian model yields significant improvements over the Gaussian model [27]. Despite the successes brought by the muti-resolution marginal models, it is also easy to see the disadvantage of marginal models because of the independence assumption. For instance, in the following case, the task is to improve the detail in the retina image for the doctors. The left part in Figure 2.4 is the original retina image, and the right part is the enhanced retina image. The wavelet transform is applied on the original image and the coefficients in each scale and direction are considered as independent. In the right figure, besides the enhanced details, such as small vessels, there are also some irregular textures appearing. The artificial effect injected into the enhanced image comes from the independence assumption.

Experimental tests expose the structure similarity/correlation across the adjacent scales and different orientations. The figure gives an intuitive example for the intra-scale and inter-scale joint statistical properties of the transform coefficients. The image 'bike' is decomposed by db4 wavelet and the figure shows the amplitude of the coefficients. It's easy to note that the large-magnitude coefficients tend to occur near each other within subbands (intra-scale), and also occur at the same relative spatial location in subbands at adjacent scales and orientations (intra-scale).

In the past twenty years, researchers have proposed various joint models. The following section will review the popular joint models:

a. Gaussian Scale Mixtures Model (GSMs) [26]

b. Hidden Markov Models (HMMs) [28],[29],[30]

Figure 2.4: Marginal models and its enhancement application[13]

1)Gaussian Scale Mixtures (GSMs) [26]

a. Definition:

Mathematically, a Gaussian scale mixture vector is defined as:

$$X = \sqrt{z}u \tag{2.14}$$

where the Gaussian random vector u is distributed as $N(0, C_u)$ and the positive scalar random variable z is the multiplier or mixing variable. From Equation (2.14) the vector X is an infinite mixture of Gaussian vectors, thus the density of a GSM vector can be represented as an integral of a Gaussian distribution and scalar variable.

$$\begin{aligned} p(X) &= \int p(X|z)p(z)dz \\ &= \int \frac{exp(-X^T(zC_u)^{-1}X/2)}{(2\pi)^{N/2}|zC_u|^{1/2}}p(z)dz \end{aligned} \tag{2.15}$$

b. Properties

The GSM family includes the α-stable family (including the Cauchy distribution), the generalized Gaussian family, and the symmetrized Gamma family [26]. According to the definition of the GSM vector, GSM density has the following properties:

1. Symmetric:

$$\begin{aligned} p(-X) &= \int \frac{exp(-(-X)^T(zC_u)^{-1}(-X)/2)}{(2\pi)^{N/2}|zC_u|^{1/2}}p(z)dz \\ &= \int \frac{exp(-X^T(-I)(zC_u)^{-1}(-I)X/2)}{(2\pi)^{N/2}|zC_u|^{1/2}}p(z)dz \\ &= \int \frac{exp(-X^T(-I)(-I)(zC_u)^{-1}X/2)}{(2\pi)^{N/2}|zC_u|^{1/2}}p(z)dz \\ &= \int \frac{exp(-X^T(zC_u)^{-1}X/2)}{(2\pi)^{N/2}|zC_u|^{1/2}}p(z)dz \\ &= p(X) \end{aligned}$$

2. Zero-mean:

$$\begin{aligned} E\{X\} &= \int_X Xp(X)dX \\ &= \int_X Xp(-X)dX \\ &= -\int_{-X}(-X)p(-X)d(-X) \\ &= -E\{X\} \end{aligned}$$

$$\Rightarrow E\{x\} = 0$$

3. Heavy-tailed distributions:

The following example shows a simple GSM density with only two Gaussian distribution parts: $p(z) = 0.3\delta(1) + 0.7\delta(4), C_u = 1$

$$\begin{aligned} p(X) &= \int p(X|z)p(z)dz \\ &= \int p(X|z)(0.3\delta(1) + 0.7\delta(4))dz \\ &= 0.3N(0, C_u) + 0.7N(0, 4C_u) \\ &= 0.3N(0, 1) + 0.7N(0, 4) \end{aligned}$$

4. Normalized vector $\frac{X}{\sqrt{u}}$ is whitened.

c. GSM modeling for the image coefficients in the Wavelet domain

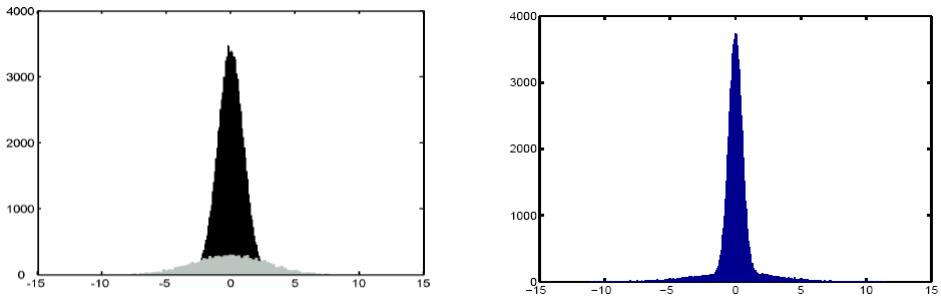


Figure 2.5: Heavy-tailed distribution

In the left figure, the grey part is N(0,4) and the black part is N(0,1); the right figure is the distribution of Gaussian scale mixture.

Consider an image decomposed into oriented subbands at multiple scales. The fine layers contain the detailed information, and the coarse layers contain the fundamental information, which is largely structural information. In this section, GSM model is applied to describe the statistical properties of the fine layers and coarse layers. $x^c_{s,o}(m,n)$ is denoted as the decomposed coefficient at the position (m,n) in the scale s and the orientation o. And $x_{s,o}(m,n)$ is denoted as the neighborhood of coefficients surrounding $x^c_{s,o}(m,n)$, as shown in Figure 2.6.

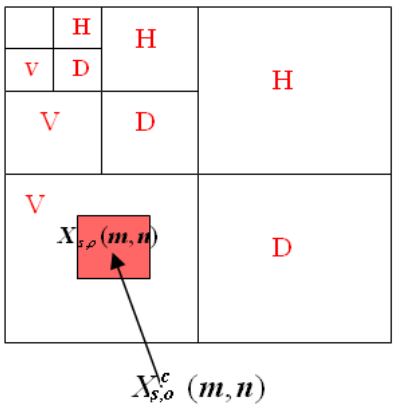


Figure 2.6: Gaussian Scale Mixture Model

In general, the neighborhood can come from the same subband, adjacent scale or a different orientation in the same scale. In the implemetation, the neighborhood located in the same subband is used, as in Figure 2.6. The subband is partitioned

into non-overlapping blocks of N coefficients each, and each block is modeled as a vector X with Gaussian Scale Mixture model. To be specific, a GSM vector X is modeled by the product of two independent random variables, namely a positive scalar random variable z known as the multiplier or mixing variable and a Gaussian random vector u is distributed as $N(0, C_u)$.

Theoretically, there is correlation within each subband at different scales and correlation across different subbands. As a compromise between accuracy and computational cost,we partition a subband into several non-overlapping blocks with the following assumption: each partitioned block is independent of the others, and sample z with logarithmically uniform spacing, which we have observed to require fewer samples, for the same quality, than linear sampling. Simulation results show that the approximation is acceptable. Without loss of generality, we can assume $E\{z\} = 1$, which implies $C_x = C_u$.

Then the GSM model is

$$Pr_{X^i_{s,o}} = \sum_j [P(z_j) \times N(0, z_j C_{X^i_{s,o}})] \tag{2.16}$$

where $X^i_{s,o}$ is the ith block in the subband of the scale s and the orientation o, $Pr_z(z)$ is the density of the mixing variable, and N is the dimension of the random vector $X^i_{s,o}$.

d. GSM model parameters estimation

As depicted above, the Gaussian Scale Mixture Model is represented as $N(0, zC_x)$. As a compromise between accuracy and computational cost, in the implementation z is sampled with logarithmically uniform spacing which is observed to require fewer samples, for the same quality, than linear sampling. Only 13 samples of $log(z)$ are considered in GSM model over an interval $[log(z_{min}), log(z_{max}))]$ using steps of size 1. $log(z_{min}) = -3$. $log(z_{max}) = 9$. Hence, the GSM model as implemented can be represented as:

$$\begin{aligned} Pr_{X^i_{s,o}} &= \sum_{j=1}^{j=13} [P(z_j) \times N(0, z_j C_{X^i_{s,o}})] \\ Pr_{X^i_{s,o}} &= \sum_{j=1}^{j=13} [P(z_j) \times f_j(X^i_{s,o})] \end{aligned} \tag{2.17}$$

where $f_j = N(0, z_j C_x)$, $z_j = -3, -2, ..., 8, 9$

C_X can be easily computed as:

$$C_X = \frac{1}{N}\{X - \frac{1}{N}\sum_{n=1}^{n=M} X_n\} \times \{X - \frac{1}{N}\sum_{n=1}^{n=M} X_n\}^T \tag{2.18}$$

From the observation data $\mathbf{X_1}, \mathbf{X_2}, ...\mathbf{X_M}$, the GSM model parameters are estimated by maximizing the likelihood of the observed set of wavelet coefficients. To be specific, the EM method is applied to yield $P(z_j)$, $j = 1, 2, ..., 13$. :

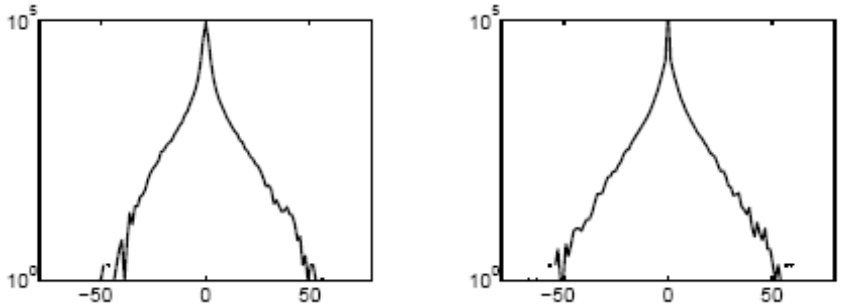
a. Initialize $Pr_0(z_j) = 1/13, j = 1, 2, 3..., 13$

b. $Pr_{q+1}(z_j) = \frac{1}{N}\sum_{n=1}^{n=M} \frac{Pr_q(z_j) f_j(\mathbf{X_n})}{\sum_{j=1}^{j=13} Pr_q(z_j) f_j(\mathbf{X_n})}$

c. Iterate step b until $Pr_{q+1}(z_j)$ converges

e. Performance of GSM models

The following example shows the performance of a GSM for modeling wavelet coefficients statistics. Figure 2.7 shows a comparison of coefficient statistics from an example image subband (left colum) with those from the simulation of the GSM model (right colum). Model parameters (covariance matrix and the multiplier z density) are estimated by the method presented above.



Original Wavelet coefficients histogram Simulated histogram from GSM

Figure 2.7: Marginal histograms of original wavelet coefficients and GSM simulated [26]

2) Hidden Markov Model (HMMs) [28]

A hidden Markov model is one of the most important concepts proposed in stochastic processing. In the past several years, it has made great accomplishments in speech processing, image processing, information theory and other areas. Y. Xu [28], M. Crouse [29] J. Romberg [30], Choi [31],Nowak [32] and G. Fan [33] extended hidden Markov model multi-resolution image statistics and made great effort to develop it into a powerful analytical tool for images.

Generally, HMMs are designed to model the coefficient statistics by proposing coefficients' hidden state dependencies through intrascale or across inter-scales. Typically HMMs are based on an assumption that given the state of parent, the child is decoupled from the tree on the other side of its parents (hidden Markov property). HMMs depict statistics by a Markov chain, but unfortunately Markov chains for image coefficients are not stationary. There is no analytical representation for such Markov chains in information theory, and it will be extremely time-consuming to compute the entropy for all nonstationary Markov chain. Hence, in the following chapter, we will focus on GSM and its applications in the theoretic information criteria for image quality assessment.

2.5 Chapter Summary

This chapter answered several significant questions. At the beginning, the random filed is introduced and followed by an important concept: a Markov Random Field. In the rest of this chapter, a literature review about image modeling is given, which especially focuses on Gaussian Scale Mixture Model(GSM). In the following chapters, GSM model is applied to represent the images . A method is proposed to compute the 'perceptual information' contained in an image based on the human visual system described in the next chapter.

Chapter 3

Human Visual System Model

3.1 HVS Model Definition

Chapter 1 gave a brief introduction about the human visual system and its psychophysical features. One fundamental psychophysical property is contrast sensitivity, as shown in Figure **??**.

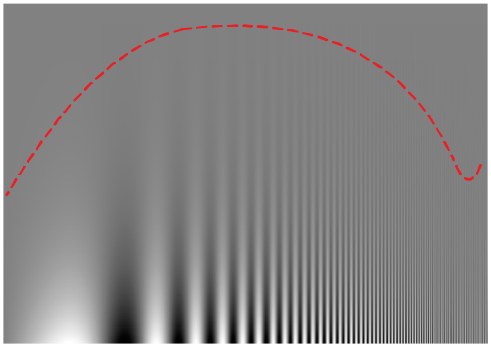


Figure 3.1: Contrast Sensitivity Chart

Campbell-Robson contrast sensitivity chart (Campbell and Robson, 1968). The spatial CSF appears as the envelope of visibility of the modulated pattern.

In the previous quality evaluation algorithms, contrast sensitivity was first implemented as a low-pass filter in [35]. Contrast sensitivity is also modeled as weight function in [36],[37].

As shown in Figure **??**, human eyes are unable to perceive the contrast difference around the dash line. This phenomenon implies that human eyes only perceive

light energy above certain minimum amounts. This is also the basic idea in quantum physics. In 1905, Einstein proposed that electro magnetic radiation (light) is quantized and exists in elementary amounts (quanta) that we now call photons. This gives us an idea to explain the contrast sensitivity as the quantization process with the step size defined as the luminance threshold at different frequencies. In the HVS model here the light is converted to 'photons' by human eyes. From the definition below, this quantization process acts as a nonlinear operation. The details about the implementation and its benefits will be discussed in the following.

Definition 3.1: If the image coefficient is X and the quantization step is $\triangle$, then the perceived signal in the human brain is $X_{\triangle}$

$$X_{\triangle} = \frac{X}{\triangle} \tag{3.1}$$

3.2 HVS Model Implementation

This section will discuss how to implement a Human Visual System Model. The model parameters are the quantization step in each subband. The method given here is based on the assumption that the quantization step in each subband is constant. The following steps will help us to compute the quantization step in the X_{s_0,o_0} subband:

i) Constrain other subbands to zero, e.g. $X_{s',o'} = 0$, for $s' \neq s_0$,$o' \neq o_0$; reconstruct the image with inverse wavelet transform, e.g., $I' = IWT(0, 0, ..., X_{s_0,o_0}, 0, ...)$

ii) Change the coefficients in the current subband until the viewer can detect the change of the reconstructed image, e.g., $I'' = IWT(0, 0, ..., X_{s_0,o_0} + \triangle, 0, ...)$, $max(|I'' - I'|) \geq threshold$. In our implementation, $threshold = 4$ gray levels, for the orginal images with 256 gray levels. The simulation result shows that the threshold between 3 to 6 gray levels will achieve the best performance.

iii) Treat $\triangle_{s_0,o_0} = augmin\{\triangle | max(|I'' - I'|) \geq threshold\}$ as the quantization step in the subband X_{s_0,o_0}

To define the *threshold* in the above steps, consider 3.2 first. The left figure is the original image 'Lena'; the middle figure is the changed image. It is very difficult for human eyes to detect the difference between these two figures. In the right figure, the red boxes indicate the areas in which the pixels values are changed. How large a pixel value change can we detect the difference? For an 8-bit gray level image, the threshold in the spatial domain is 2 bit, namely for an image pixel values belong to $[0, 255]$, $threshold = 4$.

Figure 3.2: An example for detectable threshold in spatial domain

Based on psychovisual tests, human eyes are more sensitive to contrast change in the low frequency range. In the following examples, Lena is decomposed into 6 scales, and Figure 3.3 depicts the quantization step in each subband: the quantization steps in the finest scales are bigger than those in the coarse scales. This indicates that the proposed HVS model with quantization steps is consistent with the properties of human eyes.

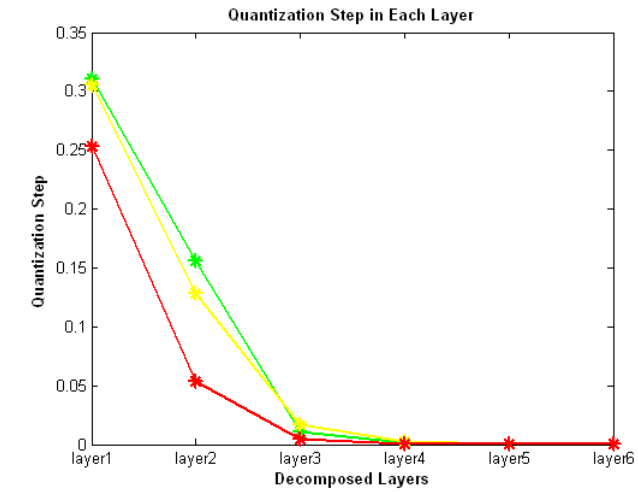


Figure 3.3: Threshold for HVS model

HVS model: quantization step in each decomposed layer. The quantization step in the fine layers (high frequency domain) is higher than the step in the coarse layers (low frequency domain). This implies our HVS model is sensitive to contrast stimuli of low frequency.

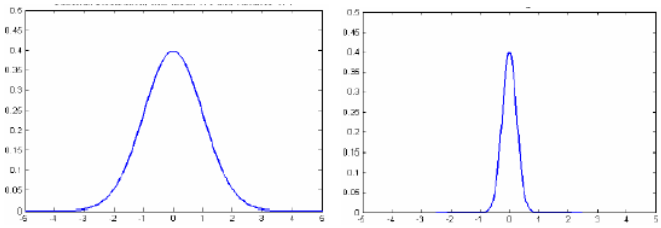
For common distortions: contrast degradation, blurring, compression degradation, noising, the human visual system model proposed above can explain the lost information due to contrast degradation, blurring and compression degradation.

Essentially, these three kinds of degradation share the same distortion model:

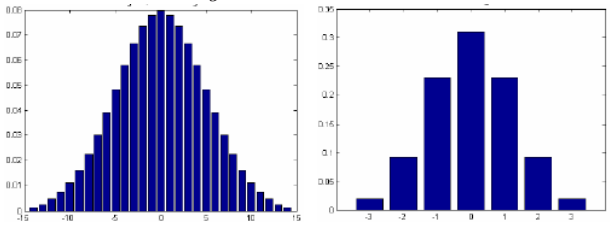
$$X_{distorted} = g \times X_{original} \tag{3.2}$$

where g is distortion parameter, and $0 \leq g \leq 1$

In the following, a simple example to motivate the HVS model is given. $X_{distorted} = 1/4 \times X_{original}$. Through the HVS model, $X_{original\triangle}, X_{distorted\triangle}$ are derived. If we are only given $X_{distorted\triangle}$, it is impossible to completely reconstruct the original quantized signal $X_{original\triangle}$ from $X_{distorted\triangle}$, even if a very complicated adaptive filter is applied, which means some information is lost. Or we can understand the information lost in the quantization process in a simpler way. Suppose we only have $X_{distorted\triangle}$. It is impossible for us to retrieve the original signal $X_{original\triangle}$ and the information is lost.



Original signal $X_{original}$ and distorted signal $X_{distorted}$



Quantized signal $X_{original\triangle}$ and Quantized signal $X_{distorted\triangle}$

Figure 3.4: An example for the lost information due to HVS model

3.3 Chapter Summary

This section built a computational human visual system to describe human vision: contrast sensitivity. Hence, mathematically a continuous signal in the real world is processed to derive the corresponding discrete signal which is represented as the 'perceptual' signal in human brain. The current question is how to compute the information contained in that 'perceptual' signal. The next chapter will

first introduce the foundation of information theory and describe the information computation in detail.

Chapter 4

Perceptual Information

When people are watching a picture or watching a video, how do they gather the information? How much information can we abstract from a perfect image, namely an image which is clear? How much information can we get from a distorted image, namely an image with noise, compression distortion or other distortion? How can we compute the information? Can we simply take the difference of the information in the perfect image and the distorted image as the indicator for the evaluated image? In this chapter, we will define a human visual model and propose a method to compute the information in one image.

4.1 Foundation of Information Theory

Remember the beginning part in Chapter II, where we define the number of binary questions needed to describe the information contained in an image. Although this idea is simple and intuitive, it is inconvenient and time-consuming to count the question number as the information. Researchers in the past several decades have devoted huge efforts to look for and develop proper methods to describe the information within signals. In 1948, Claude E. Shannon published *A Mathematical Theory of Communication* [38] in the Bell System Technical Journal. In the paper, Shannon depicted the information of a source composed of a finite number of symbols by measuring entropy. The basic idea of entropy is to compute the uncertainty of the signal source based on the probability distribution of the source components.

Definition 4.1: The entropy $H(X)$ of a discrete random variable X is defined by:

$$H(X) = -\sum_{X \in \mathcal{X}} p(X) log_2(p(X)) \tag{4.1}$$

Definition 4.2: The joint entropy $H(X_1, X_2)$ of a pair of discrete random variables (X_1, X_2) with a joint distribution $p(X_1, X_2)$ is defined by:

$$H(X_1, X_2) = -\sum_{X_1 \in \mathcal{X}_1} \sum_{X_2 \in \mathcal{X}_2} p(X_1, X_2) log_2(p(X_1, X_2)) \tag{4.2}$$

Definition 4.3: Consider two random variables X_1 and X_2 with a joint probability mass function $p(X_1, X_2)$ and marginal probability mass functions $p(X_1)$ and $p(X_2)$. The mutual information $I(X;Y)$ is defined as:

$$I(X_1; X_2) = \sum_{X_1 \in \mathcal{X}_1} \sum_{X_2 \in \mathcal{X}_2} log_2(\frac{p(X_1, X_2)}{p(X_1)p(X_2)}) \tag{4.3}$$

The definition of the mutual information is not direct, but we can understand the mutual information as the shared information between X_1 and X_2.

Theorem 4.1: Let X_1, X_2,..., X_n be drawn according to $p(X_1, X_2..., X_n)$. Then:

$$H(X_1, X_2, ...X_n) \leq \sum_{i=1}^{n} H(X_i) \tag{4.4}$$

with equality if and only if the X_i are independent.

Theorem 4.1 tells us if two X_1, X_2 random variables have nothing to do with each other, the total information contained in the joint random variable (X_1, X_2) is just the summation of the information contained in each variable.

Definitions 4.1 to 4.3 discuss the discrete random variable and its information entropy and is the foundation of information theory. Recall the mutliresolution transform and GSM in Chapter II where the statistical model used to represent the natural scenes is continuous. In the next section the continuous random variable and its information entropy: differential entropy, is introduced.

Definition 4.4: The differential entropy $H(X)$ of a continuous random variable X with a density $p(X)$ is defined as

$$H(X) = \int_{\mathcal{X}} p(X) log_2(p(X)) dX \tag{4.5}$$

where $\mathcal{X}$ is the support set of the random variable.

From the above definition, if we have Gaussian distribution $p(X) = N(u, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{X^2}{\sigma^2}}$. Then the differential entropy of X is:

$$H(X) = -\int_{-\infty}^{+\infty} p(X) log_2(p(X)) dX = \frac{1}{2} log_2(2\pi e\sigma^2) \tag{4.6}$$

Consider if X is an N dimensional Gaussian distribution vector with the covariance matrix C_X, which can be decomposed as:

$$\begin{aligned} C_X &= P\Lambda Q \\ QC_XP &= \Lambda \end{aligned} \tag{4.7}$$

where Λ is a diagonal matrix, P is composed of eigenvectors, and $P = Q^T$. The new Gaussian distribution vector $Y = PX$ will have the independence property: its components $y_1, y_2, ..., y_N$ are independently Gaussian distributed. The corresponding variances of $y_1, y_2, ..., y_N$ are the diagonal coefficients in Λ. $y_1 \sim N(0, \lambda_1)$, $y_2 \sim N(0, \lambda_2)$,..., $y_N \sim N(0, \lambda_N)$

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ . & . & . & . \\ 0 & 0 & ... & \lambda_N \end{bmatrix}$$

$$\begin{aligned} H(Y) &= \sum_{i=1}^{i=N} H(y_i) && (4.8) \\ &= \sum_{i=1}^{i=N} H(N(0, \lambda_i)) && (4.9) \\ &= \sum_{i=1}^{i=N} \frac{1}{2} log_2(2\pi e\lambda_i) && (4.10) \\ &= \frac{1}{2} log_2(\prod_{i=1}^{i=N} (2\pi e\lambda_i)) && (4.11) \end{aligned}$$

$$= \frac{1}{2} log_2((2\pi e)^N \prod_{i=1}^{i=N} (\lambda_i)) \quad (4.12)$$

$$= \frac{1}{2} log_2((2\pi e)^N det(\Lambda)) \quad (4.13)$$

$$= \frac{1}{2} log_2((2\pi e)^N det(QC_X P) \quad (4.14)$$

$$= \frac{1}{2} log_2((2\pi e)^N det(Q) det(C_X) det(P)) \quad (4.15)$$

$$= \frac{1}{2} log_2((2\pi e)^N det(C_X)) \quad (4.16)$$

where (4.7) follows from Theorem 4.1 (independence); (4.9) from Formula (4.6); (4.10) from the property of logarithm function; (4.12) from the property of diagonal matrix: the product of its diagonal coefficients is its determinant; (4.13) from (4.7); (4.16) from $det(P) = det(Q) = 1$, because P and Q are composed of eigenvectors.

Hence, finally the differential entropy of a Gaussian distribution vector X is:

$$H(X) = \frac{1}{2} log_2((2\pi e)^N det(C_X)) \quad (4.17)$$

Definition 4.5: The differential entropy of a set X_1,X_2,...,X_n of random variables with density $p(X_1, X_2, ..., X_n)$ is defined as:

$$H(X_1, X_2, ..., X_n) = \int p(X_1, X_2, ..., X_n) log_2(p(X_1, X_2, ..., X_n)) dX_1 dX_2 ... dX_n \quad (4.18)$$

Definition 4.6: For $(X, Y) \sim p(X, Y)$, then the conditional entropy $H(Y|X)$ is defined as:

if X and Y are discrete random variables.

$$H(Y|X) = \sum_X p(X) \sum_Y p(Y|X) log_2(Y|X) \quad (4.19)$$

if X and Y are continuous random variables.

$$H(Y|X) = \int_X p(X) \int_Y p(Y|X) log_2(Y|X) \quad (4.20)$$

if X is a discrete random variable, and Y is a continuous random variable.

$$H(Y|X) = \sum_X p(X) \int_Y p(Y|X) log_2(Y|X) \quad (4.21)$$

This definition is not straight forward. The conditional entropy can be explained as the unknown information contained in the variable Y if X is known to us.

Definition 4.7: Consider two random variables X and Y with a joint probability function $p(X,Y)$ and marginal probability functions $p(X)$ and $p(Y)$. The mutual information $I(X;Y)$ is the relative entropy between the joint distribution and the product distribution $p(X)p(Y)$: if X and Y are discrete random variables.

$$I(X;Y) = \sum_X \sum_Y p(X,Y) log_2(\frac{p(X,Y)}{p(X)p(Y)}) \tag{4.22}$$

if X and Y are continuous random variables.

$$I(X;Y) = \int_X \int_Y p(X,Y) log_2(\frac{p(X,Y)}{p(X)p(Y)}) \tag{4.23}$$

The concept of mutual information can be understood as the shared information between X and Y.

Theorem 4.2: Let X_1, X_2,...,X_n be drawn according to $p(X_1, X_2, ..., X_n)$. Then the joint entropy

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, ..., X_1) \tag{4.24}$$

For example: $H(X,Y) = H(X) + H(Y|X)$, $H(X,Y) = H(Y) + H(X|Y)$. The relationship between random variable entropy, joint entropy, conditional entropy and mutual entropy:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (4.25)\\ I(X;Y) &= H(Y) - H(Y|X) && (4.26)\\ I(X;Y) &= H(X) + H(Y) - H(X,Y) && (4.27)\\ I(X;Y) &= I(Y;X) && (4.28) \end{aligned}$$

The above relationship can be easily expressed in a Venn diagram in Figure 4.1.

Theorem 4.3: Let X be a continuous random variable with the variance σ, the differential entropy

$$H(X) \leq \frac{1}{2} log_2(2\pi e\sigma^2) \tag{4.29}$$

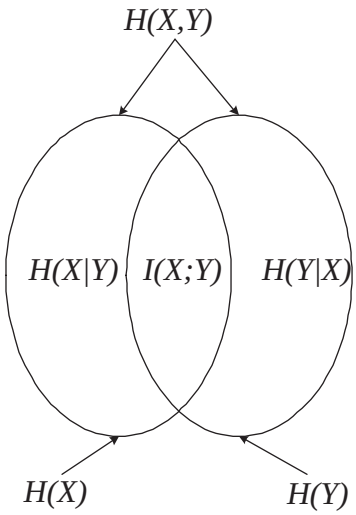


Figure 4.1: Relationship between entropy and mutual information

with equality if and only if the X is Gaussian.

Now, in the following section, we will discuss the relationship between the entropy of continuous signal and the entropy of its quantized signal. 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 4.2:

$$p(\tilde{X}_i)\triangle = \int_{i\triangle}^{(i+1)\triangle} p(X)dX \tag{4.30}$$

Consider the quantized random variable:

$$X_\triangle = \tilde{X}_i \qquad if \quad i\triangle \leq X \leq (i+1)\triangle \tag{4.31}$$

Then the probability that $X_\triangle = \tilde{X}_i$

$$P_i = \int_{i\triangle}^{(i+1)\triangle} p(X)dX = p(\tilde{X}_i)\triangle \tag{4.32}$$

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

$$\begin{aligned} H(X_\triangle) &= -\sum_{i=-\infty}^{+\infty} P_i log_2(P_i) \\ &= -\sum_{i=-\infty}^{+\infty} p(\tilde{X}_i) \triangle log_2(p(\tilde{X}_i)\triangle) \\ &= -\sum_{i=-\infty}^{+\infty} p(\tilde{X}_i) \triangle log_2(p(\tilde{X}_i)) - \sum_{i=-\infty}^{+\infty} p(\tilde{X}_i) \triangle log_2(\triangle) \\ &= -\sum_{i=-\infty}^{+\infty} p(\tilde{X}_i) \triangle log_2(p(\tilde{X}_i)) - log_2(\triangle) \end{aligned}$$

since $\sum_{i=-\infty}^{+\infty} p(\tilde{X}_i)\triangle = \int p(X) = 1$, the first term will approach the differential entropy $H(X) = \int p(X) log_2(p(X)) dX$

Hence, we derive the following theorem about the relationship of the differential entropy and discrete entropy.

Theorem 4.4: Consider a continuous variable X and its quantized random variable $X_\triangle$

$$H(X_\triangle) \to H(X) - log_2(\triangle) \qquad \triangle \to 0 \tag{4.33}$$

where $\triangle$ is the quantization step.

From Theorem 4.4, the approximation of information entropy contained in the quantized random variable $X_\triangle$ can be represented as:

$$H(X_\triangle) \simeq H(X) - log_2(\triangle) \tag{4.34}$$

If $log_2(\triangle) \geq H(X)$, $H(X_\triangle) = 0$, hence the approximation formula should be

$$H(X_\triangle) = \begin{cases} H(X) - log_2(\triangle) & \text{if} \quad log_2(\triangle) \leq H(X) \\ 0 & \text{if} \quad log_2(\triangle) > H(X) \end{cases} \tag{4.35}$$

For an N dimensional continuous random vector X, the quantized random vector $X_\triangle$ can be computed as:

$$H(X_\triangle) = \begin{cases} H(X) - Nlog_2(\triangle) & \text{if} \quad Nlog_2(\triangle) \leq H(X) \\ 0 & \text{if} \quad Nlog_2(\triangle) > H(X) \end{cases} \tag{4.36}$$

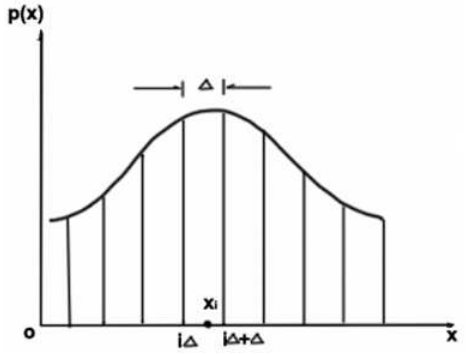


Figure 4.2: Quantization of a continuous random variable

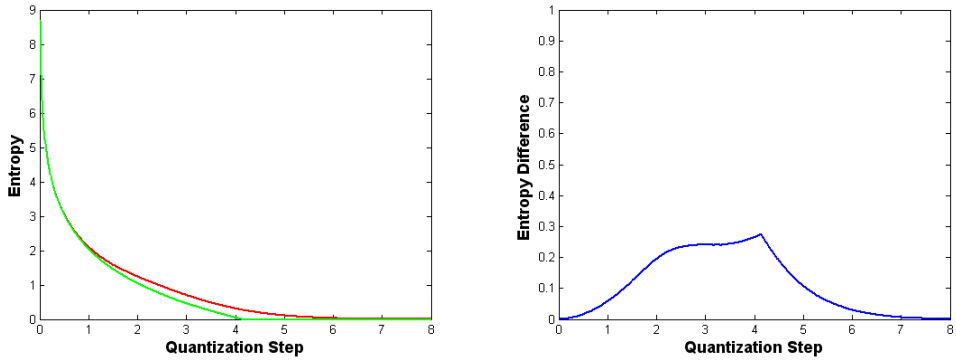


Figure 4.3: Performance of the approximation expression

Figure 4.3 gives the performance of the approximation expression. In the left figure, the light line is our approximation, and the dark line is the exact differential entropy in the continuous variable X, which is Gaussian distribution with variance $\sigma = 1$. The right figure shows the difference between the approximation and the exact entropy. In the simulation comparison, the quantization step varies from 0.01 to 8. The simulation result shows that the approximation method is close to the exact entropy in the large entropy area. Although the approximation difference reaches the maximum (0.2658 bit) when the quantization step is around 4, in this case, the exact entropy is only 0.3123 bit and this amount of entropy will not play an important role in the total entropy contained in an image (total entropy is over 10^7 bits).

4.2 Perceptual Information

In the previous sections, we introduced the image statistical model, HVS model (quantization procedure) and the foundation of information theory. Now we are ready to propose a method to compute the perceptual information contained in an image. Recall the GSM model in Chapter II.

The density of a GSM vector can be represented as an integral of Gaussian distribution and a scalar variable. Theoretically, there is correlation within each subband at different scales and correlation across different subbands. For the compromise between accuracy and computational cost,we partition a subband into several non-overlapping blocks with the following assumption: each partitioned block is independent of the others, and sample z with logarithmically uniform spacing. The simulation result shows that the approximation is acceptable. Without loss of generality, we can assume $E\{z\} = 1$, which implies $C_x = C_u$.Then the GSM model is

$$Pr_{X^i_{s,o}} = \sum_j [p(z_j) \times N(0, z_j C_{X^i_{s,o}})] \tag{4.37}$$

where $X^i_{s,o}$ is the ith block in the subband of the scale s with orientation o, $Pr_z(z)$ is the density of the mixing variable, and N is the dimension of the random vector $X^i_{s,o}$.

According to the independence assumption, the subband in the scale s with orientation o is partitioned into $n_{o,s}$ non-overlapping blocks and each partitioned

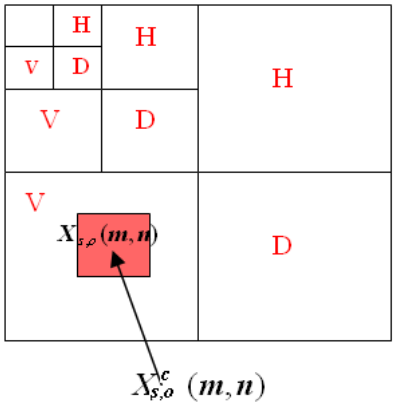


Figure 4.4: Gaussian Scale Mixture Model

block is independent of the others. The perceptual information contained in the image is:

$$H(X) = \sum_{s} \sum_{o=h,d,v} \sum_{i}^{n_{o,s}} H(X_{s,o}^{i}) \tag{4.38}$$

where $X_{s,o}^{i}$ is the N dimensional random vector for the ith block in the scale s with orientation o.

In the following part, we derive how to compute $H(X_{s,o}^{i})$. To simplify the notation, B is used instead of $X_{s,o}^{i}$ in the derivation, and z is the corresponding scalar random variable for the block B.

$$
\begin{aligned}
H(B) &= H(B,z) && (4.39)\\
&= H(B|z) + H(z) && (4.40)\\
&= \sum_{j=1}^{j=13} p(z_j) H(B|z_j) + H(z) && (4.41)\\
&= \sum_{j=1}^{j=13} p(z_j) H(N(0, z_j C_B)) - H(z) && (4.42)\\
&= \sum_{j=1}^{j=13} p(z_j) \frac{1}{2} log_2(det(2\pi e z_j C_B)) - H(z) && (4.43)\\
&= \sum_{j=1}^{j=13} p(z_j) \frac{1}{2} log_2(det(2\pi e z_j C_B)) - \sum_{j=1}^{j=13} p(z_j) log_2(p(z_j)) && (4.44)
\end{aligned}
$$

where (4.27) follows from the fact that the information of z is totally contained in B, because we can estimate the distribution of z from B according to the method in Chapter II. This fact means that the joint entropy of (B, z) will not bring any 'new' information. (4.28) follows Theorem 4.2 (chain rule); (4.29) from the definition of conditional entropy (Definition (4.6));(4.30) from the definition of the Gaussian scale mixture model,$B|z_j \sim N(0, z_j C_u)$; (4.31) from the differential entropy of Gaussian distribution (4.6); and (4.32) from the definition of discrete random variable.

The perceptual information contained in Block B, can be computed as:

$$
H(B_\Delta) = \sum_{j=1}^{j=13} p(z_j) \frac{1}{2} log_2(2\pi e det(z_j C_B)) - \sum_{j=1}^{j=13} p(z_j) log_2(p(z_j)) - N log_2(\Delta) \tag{4.45}
$$

The total perceptual information contained is the summation of the perceptual information in each block.

4.3 Minimum Perceptual Information between the Reference Image and Distorted Image

Now it is possible for us to compute the remaining perceptual information after degradation. Because of the different degradations, such as noising, blurring and

compression distortion, it is very difficult to design a mathematical degradation model. Although in [39], Niranjan proposed the famous degradation model:

$$Y = gX + v \tag{4.46}$$

where $0 \leq g \leq 1$ and v is a white noise, this model can explain noising, blurring. But other degradation: such as compression distortion and fast fading distortion are out of the range of this model. Niranjan tried to use the above degradation model and compute the mutual information between the reference image and the distorted image. Their algorithm is not only limited to certain degradations, but also requires the estimation on the parameters of the degradation. Furthermore, because there are no proper models for image enhancement techonolgies, it is also impossible to compute the mutual information between the reference image and the enhanced image. In this paper we define the minimum perceptual information remaining in the distorted/enhanced image as the criterion for image quality.

The basic idea is explained below:

Theorem 4.4: If X is the original one dimensional random signal, Y is the distorted signal from X, and distance between two signals is $D = E[(X-Y)^2]$ the minimum mutual information between X and Y:

$$H(X) - \frac{1}{2}log_2(2\pi eD) \leq I_{min}(X;Y) \leq \frac{1}{2}log_2(2\pi e\sigma^2) - \frac{1}{2}log_2(2\pi eD) \tag{4.47}$$

where σ is the variance of the signal X.

Theorem 4.5: If X is the original N-dimensional random vector, Y is the distorted vector from X, and distance between two vectors is $D = E[(X-Y)(X-Y)^T]$ the minimum mutual information between X and Y:

$$H(X) - \frac{1}{2}log_2((2\pi e)^N det(D)) \leq I_{min}(X;Y) \leq \frac{1}{2}log_2((2\pi e)^N det(C_X)) - \frac{1}{2}log_2((2\pi e)^N det(D)) \tag{4.48}$$

where C_X is the covariance matrix of X.

The foundation discussion introduced the basic idea of mutual information: the information shared between two information sources. This gives us the idea to compute the minimum perceptual information between the reference image and

distorted image. What's the relationship between the minimum mutual information and the minimum perceptual information? Roughly speaking:

$$H(X) - \frac{1}{2}log_2((2\pi e)^N det(D)) \leq I_{min}(X_\triangle; Y_\triangle) \leq \frac{1}{2}log_2((2\pi e)^N det(C_X)) - \frac{1}{2}log_2((2\pi e)^N det(D))$$

If the quantization step is $\triangle$, the minimum perceptual information:

$$I_{min}(X_\triangle; Y_\triangle) \leq \begin{cases} \frac{1}{2}log_2((2\pi e)^N det(C_X)) - \frac{1}{2}log_2((2\pi e)^N det(D)) \\ \quad if \frac{1}{2}log_2((2\pi e)^N det(C_X)) \geq log_2(\triangle^{-N}) \, , \frac{1}{2}log_2((2\pi e)^N det(D)) \geq log_2(\triangle^{-N}); \\ \frac{1}{2}log_2((2\pi e)^N det(C_X)) - Nlog_2(\triangle) \\ \quad if \frac{1}{2}log_2((2\pi e)^N det(C_X)) \geq log_2(\triangle^{-N}) \, , \frac{1}{2}log_2((2\pi e)^N det(D)) \leq log_2(\triangle^{-N}); \\ 0 \\ \quad else \end{cases}$$

$$I_{min}(X_\triangle; Y_\triangle) \geq \begin{cases} H(X) - \frac{1}{2}log_2((2\pi e)^N det(D)) \\ \quad if H(X) \geq log_2(\triangle^{-N}) \, , \frac{1}{2}log_2((2\pi e)^N det(D)) \geq log_2(\triangle^{-N}); \\ H(X) - Nlog_2(\triangle) \\ \quad if H(X) \geq log_2(\triangle^{-N}) \, , \frac{1}{2}log_2((2\pi e)^N det(D)) \leq log_2(\triangle^{-N}); \\ 0 \\ \quad else \end{cases}$$

The above expression can be easily explained as the following:

1) If the distance D is so small that human eyes are unable to detect the distortion, namely $\frac{1}{2}log_2(2\pi e det(D)) \leq log_2(\triangle^{-N})$, then $I_min(X_\triangle; Y_\triangle)$ belongs to the range $[H(X), \frac{1}{2}log_2(2\pi e det(C_x))]$;

2) If the distance D is large enough to be detected by human eyes, namely $\frac{1}{2}log_2(2\pi e det(D)) \geq log_2(\triangle^{-N})$, then the minimum perceptual information will decrease;

3) If the original signal is very weak, namely $H(X) - \frac{1}{2}log_2(2\pi e det(D))$, human eyes will detect no information from X. Hence, there will be no mutual perceptual information counted.

The proof for the above theorems and formulas is complicated and listed in Appendix.

From the independence assumption we made in Chapter II, The minimum perceptual information contained in the image is:

$$I_{min}(X;Y) = \sum_{s} \sum_{o=h,d,v} \sum_{i}^{n_{o,s}} I_{min}(X_{s,o}^{i}; Y_{s,o}^{i}) \tag{4.49}$$

where $X_{s,o}^{i}$ is the N dimensional random vector for the ith block in the scale s and the orientation o, $Y_{s,o}^{i}$ is at the same spatial position as $X_{s,o}^{i}$.

4.4 Information Theoretic Criteria for Image Quality Assessment

$min|_D\{I(X;Y)\}$ can be our information fidelity criterion that quantifies the statistical information that is shared between the source and the distorted images. But perceptual information contained in images may vary a lot, which might cause confusion when it is applied to different images. Accordingly, we normalize the minimum mutual entropy and define the information theoretic criterion as:

$$ITC = \frac{min|_D\{I(X;Y)\}}{H(X)} \tag{4.50}$$

Sheikh proposed an information theoretic criterion (VIF) in [39] . Basically, Sheikh implemented a degradation model:

$$Y = g \times X + v$$

where g is the degradation scale and v is white noise. In his paper, Sheikh tried to use the above model to approximate the relationship between the source signal and distorted signal. That model is proper for white noise injection if the wavelet function is orthogonal. Meanwhile it is a good approximation of Gaussian Blurring. Furthermore, the implementation of degradation model limits the application of Sheikh's algorithm to evaluate enhancement technologies. The method proposed in this thesis does not need a degradation model but only computes the average distance between the reference image and the test image. It is possible for our method to evaluate the image quality and the performance of enhancement technologies.

4.5 Chapter Summary

Chapter 4 starts with the question of how to define the 'perceptual' information in an image. In the following section, the basic concepts of information theory are introduced. From the relationship of differential entropy and discrete entropy, we connect the natural scenes statistics with human visual system model to compute the perceptual information contained in an image. This chapter also solves the problem of how to describe the distorted image if its reference image is given. Instead of independently computing the information in the reference image and distorted image, we compute the shared information (mutual information) as a possible criterion for image quality assessment. Finally, a normalized mutual information is defined as the information theoretic criterion. In the next chapter, validation and discussion are given.

Chapter 5

Results and Discussions

To test the performance of the proposed ITC algorithm, it was applied on a database set from the University of Texas, Austin [2]. The simulation results were compared with several famous image quality metrics. The results of experiments using different model parameters, such as the size of neighborhood and wavelet function are also discussed in this chapter. Meanwhile this chapter illustrates the impact of implementing the HVS model in Chapter 3, and proposes a new method to optimize the HVS model performance.

5.1 Algorithm Framework

The framework of the information theoretic criterion is shown in Fig 5.1. The test (distorted) image and the reference image are first decomposed into corresponding subbands T and R in the wavelet domain. Second, the distance $d(T, R)$ is computed as mean absolute distance or mean square distance. In the following step, reference subbands R and test image subbands T are represented as random fields X and Y by GSM. Third, through the HVS model, we derive the perceived signal $X_{\triangle}$ and $Y_{\triangle}$. The lower bound of mutual entropy between $X_{\triangle}$ and $Y_{\triangle}$ under the condition of distance $d(T, S)$ is computed as the perceptual information lost in the distorted images. Finally, the information theoretic criterion ITC is defined as the normalized mutual information.

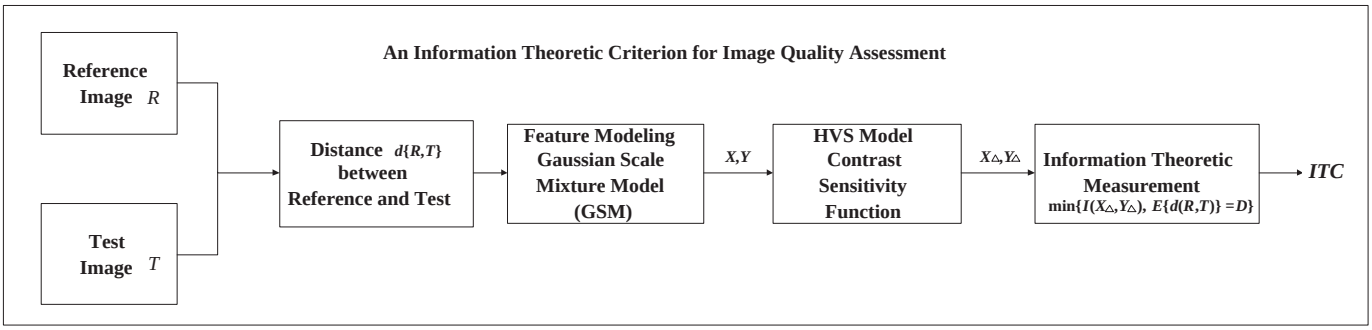


Figure 5.1: The framework of the information theoretic criterion

5.2 Experimental Results Using Different Wavelet Functions

As described in the previous section, the reference image and the test image are first decomposed into the wavelet domain. This section will give us a sense of the performances of ITC using different wavelet functions.

From the simulation results in Table 5.1 and Table 5.2, it is found that the wavelet function will not affect the algorithm performance much. For J2k compression distortion, the correlation between the subjective judgments and objective judgment varies between 0.9502 and 0.9787 using different wavelet functions; for blurring degradation, the correlation varies between 0.9612 to 0.9757; for white noise injection, the correlation varies between 0.9842 to 0.9901; and evaluating various distortions, the over-all correlation varies between 0.9203 and 0.9792.

Linear Correlation	J2k	Blurring	Noise	Over-all
rbio 1.3	0.9657	0.9589	0.9881	0.9587
rbio 1.5	0.9747	0.9643	0.9892	0.9366
rbio 2.2	0.9767	0.9678	0.9866	0.9696
rbio 2.4	0.9785	0.9716	0.9883	0.9742
rbio 2.6	0.9774	0.9704	0.9887	0.9762
rbio 2.8	0.9778	0.9729	0.9887	0.9771
rbio 3.1	0.9747	0.9738	0.9886	0.9673
rbio 3.3	0.9787	0.9738	0.9890	0.9777
rbio 3.5	0.9786	0.9710	0.9900	0.9792
rbio 3.7	0.9784	0.9726	0.9892	0.9776
rbio 3.9	0.9787	0.9757	0.9889	0.9781
rbio 4.4	0.9784	0.9682	0.9884	0.9725
rbio 5.5	0.9770	0.9675	0.9892	0.9741
rbio 6.8	0.9767	0.9685	0.9892	0.9741
db5	0.9592	0.9610	0.9903	0.9444
db7	0.9627	0.9664	0.9892	0.9485
sym6	0.9502	0.9612	0.9901	0.9203
coif3	0.9604	0.9691	0.9890	0.9357
bior 3.3	0.9771	0.9703	0.9842	0.9690
bior 3.5	0.9736	0.9646	0.9893	0.9679

Table 5.1: The linear correlation between subjective judgments and objective judgments

Mean Absolute Error	J2k	Blurring	Noise	Over-all
rbio 1.3	0.0897	0.0715	0.0785	0.0808
rbio 1.5	0.1041	0.0760	0.0802	0.0884
rbio 2.2	0.0605	0.0494	0.0865	0.0650
rbio 2.4	0.0799	0.0530	0.0851	0.0733
rbio 2.6	0.0939	0.0640	0.0829	0.0815
rbio 2.8	0.891	0.0654	0.0825	0.0818
rbio 3.1	0.0662	0.0571	0.0823	0.0741
rbio 3.3	0.0771	0.0571	0.0831	0.0729
rbio 3.5	0.0889	0.0716	0.0794	0.0808
rbio 3.7	0.0831	0.0658	0.0817	0.0774
rbio 3.9	0.0745	0.0611	0.0828	0.0730
rbio 4.4	0.0842	0.0602	0.0842	0.0770
rbio 5.5	0.0898	0.0723	0.0813	0.0819
rbio 6.8	0.1022	0.0739	0.0808	0.0872
db5	0.1945	0.0865	0.0768	0.1165
db7	0.1725	0.0825	0.0819	0.1262
sym6	0.2070	0.0830	0.0772	0.1302
coif3	0.1746	0.0751	0.0820	0.1165
bior 3.3	0.1120	0.0559	0.0955	0.0740
bior 3.5	0.9736	0.0737	0.0790	0.0904

Table 5.2: The mean absolute error between subjective judgments and objective judgments

	J2k	Blurring	Noise	Over-all	Computation Time
1×1	0.1233	0.0715	0.0785	0.0808	3.950s/pic
3×3	0.9784	0.9682	0.9884	0.9725	11.93s/pic
5×5	0.9788	0.9695	0.9839	0.9741	32.17s/pic
7×7	0.9780	0.9668	0.9844	0.9702	59.95s/pic

Table 5.3: Different block size VS Performance (Linear Correlation)

	J2k	Blurring	Noise	Over-all	Computation Time
1×1	0.1233	0.0715	0.0785	0.0808	3.950s/pic
3×3	0.0842	0.0603	0.0840	0.0769	11.93s/pic
5×5	0.0663	0.0484	0.0975	0.0703	32.17s/pic
7×7	0.0728	0.0535	0.0961	0.0740	59.95s/pic

Table 5.4: Different block size VS Performance (Absolute Error)

5.3 Experimental Results Using Different Sizes of Neighborhood

In the Gaussian Scale Mixture modeling, the bottleneck in the implementation of ITC is the computation time and the size of the neighborhood used in GSM. In the experiments, 1×1,3×3,5×5 and 7×7 are used in GSM to derive the statistical properties of images. The above table tells us the average computation time of different neighborhood size and the corresponding performance of ITC. ITC using 1×1 neighborhood consume the least time to evaluate the image, but the objective judgments are not as good as expected. As the neighborhood size increases, better results are derived, however the computation time increases exponentially.

5.4 Experimental Results using Decomposition in One Orientation

The default configuration used in the implementation is to model the statistical properties of the image in three directions: horizontal, vertical and diagonal directions. It is interesting to take a look at the performances of ITC using only one orientation. The table below demonstrates the efficiency of the method.

Generally, the method using horizontal data has better correlation with human judgement, although using vertical data will give us better results for J2k distortion.

5.5 Impact of Human Visual System Model on the Information Theoretic Criterion

In this section,the impact of the human visual system model proposed in this thesis will be discussed. As described in Chapter 3, the human visual system model is implemented as a quantization procedure. The reason is that vision has different psychophysical sensitivity in different frequency bands. Roughly speaking, if the changes are too small the eye may not detect these tiny distortions. A simple example of that phenomenon is given in Figure 3.2. The quantization step in each subband indicates the sensitivity of human eyes. We make the assumption, namely the quantization step is fixed, to derive a good approximation of the human vision system. Figure 5.2 gives us a clear illustration about the impact of the human visual system model to ITC. In the left part, the quantization step in each subband is set to 1, which means we ignore that the contrast sensitivity of human eyes is not same in different frequencies. The simulation results shows that the objective judgments have poor consistency with human evaluations. The right figure shows the experimental results which are consistent with objective judgments.

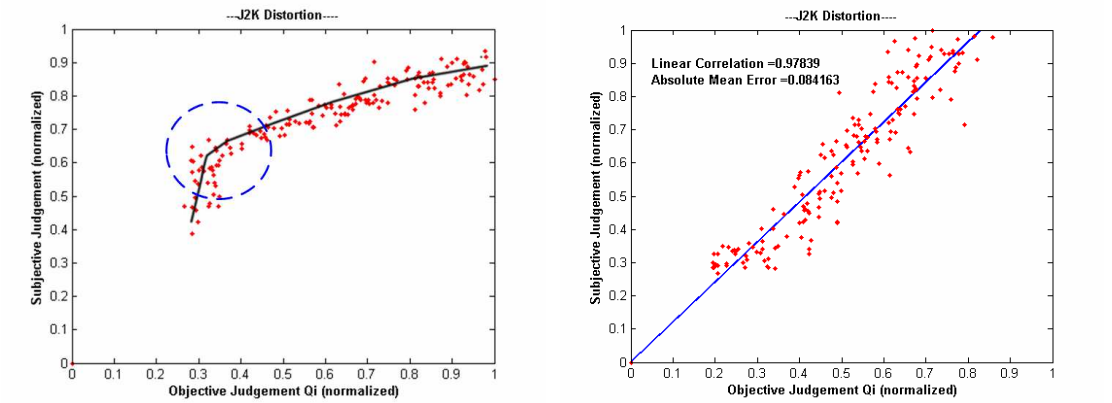


Figure 5.2: Impact of Human Visual System Model on the Information Theoretic Criterion

5.6 Optimization of Human Visual System Model

With the assumption that the quantization step in each subband is constant, the following steps will help us to compute the quantization steps in the subbands:

Suppose C is a M dimensional vector composed of the corresponding quantization steps for M subbands, $C(i)$ is the quantization step for the ith subband:

a) Initialize the quantization step vector C_1 and C_2, for example $C_1(i) = min\{$coefficients in ith subband $\}$, $C_2(i) = max\{$coefficients in ith subband $\}$;

b) Compute the objective judgment of information theoretic criterion ITC_1 and ITC_2 for the test images using C_1 and C_2 quantization step vector;

a) Compute the linear correlation L_1 between ITC_1 and subjective evaluations $DMOS$, and the linear correlation L_2 between ITC_2 and $DMOS$;

d) Compute the new generation of $C_3 = W_1 \times C_1 + W_1 \times C_1$, where

$W_1 = \frac{\frac{1}{1-L_1}}{\frac{1}{1-L_1} + \frac{1}{1-L_2}}$, $W_2 = \frac{\frac{1}{1-L_2}}{\frac{1}{1-L_1} + \frac{1}{1-L_2}}$;

e) Compute the ITC_3 using C_3 quantization step vector;

f) Compute the linear correlation L_3 between ITC_3 and DMOS;

g) Discard one quantization step vector, of which the linear correlation between the objective judgments and subjective judgments is smallest; for example: if $L_3 > L_2 > L_1$, then C_1 discarded;

h) Refresh the two remaining quantization vectors as newly generated C_1 and C_2;

i) Iterate Step a) to Step h) until the correlation reaches the maximum.

In Step d) the weight W_1 and W_2 satisfy the property that $W_1 + W_2 = 1$. The correlation parameters L_1 and L_2 belong to $[0, 1]$. If $L_1 > L_2$, so $W_1 > W_2$ if the new generated quantization vector C_3 is close to C_1.

Tables 5.5 and 5.6 show us the optimization results. Although the above method proposed above improves ITC, it is still very difficult to achieve the global optimization.

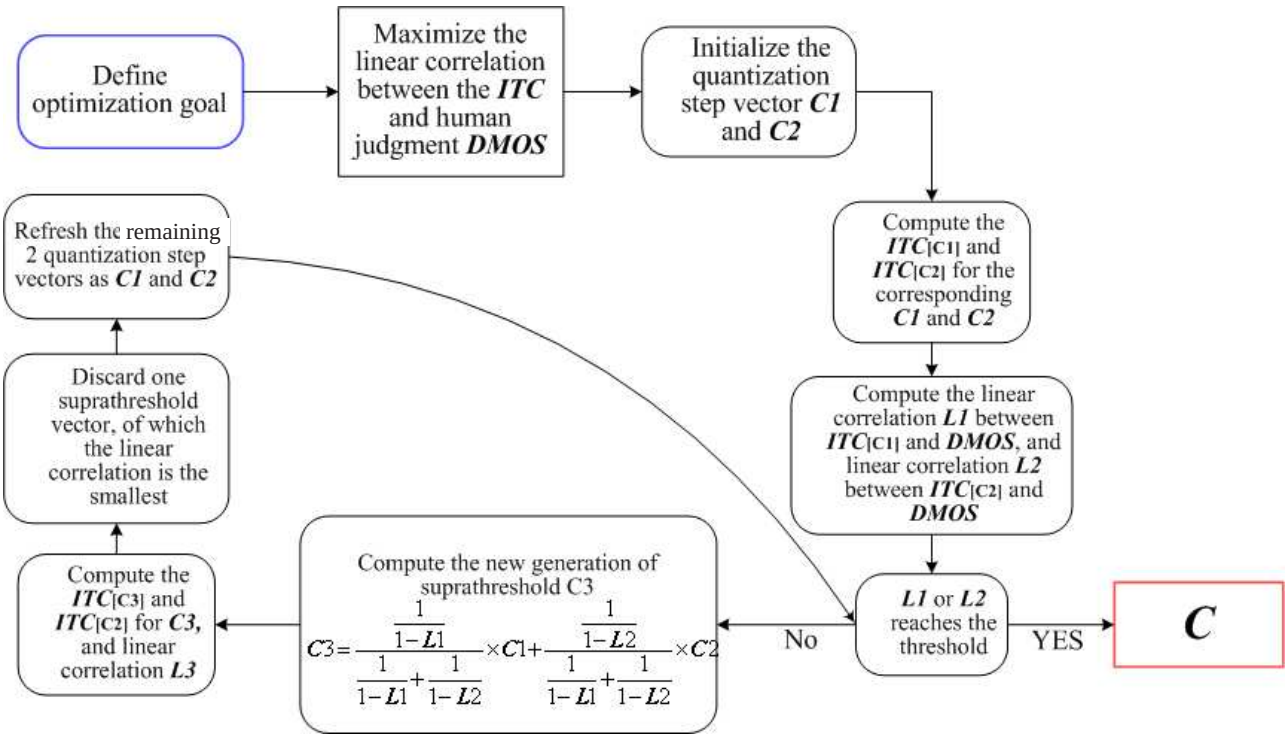


Figure 5.3: Optimization of Human Visual System Model

	Gaussian Noise	J2k Distortion	Gaussian Blurring	Overall
$\text{ITC}_{Original}$	0.9832	0.9773	0.9710	0.9723
$\text{ITC}_{Optimized}$	0.9900	0.9792	0.9740	0.9775

Table 5.5: The performance of optimized ITC (Linear Correlation)

	Gaussian Noise	J2k Distortion	Gaussian Blurring	Overall
$\text{ITC}_{Original}$	0.0599	0.0490	0.0700	0.0605
$\text{ITC}_{Optimized}$	0.0423	0.0448	0.0494	0.0450

Table 5.6: The performance of optimized ITC (Absolute Error)

	Gaussian Noise	Gaussian Blurring	J2K Distortion	Overall
PSNR	0.8916	0.7827	0.5409	0.6354
MSE	0.9147	0.6400	0.7247	0.7884
MSSIM	0.9457	0.8785	0.8969	0.8730
Universal Image Quality Assessment	0.9551	0.9304	0.9342	0.9188
Non-reference Image Quality Assessment	-*	-	0.8969	-
VIF	0.9795	0.9781	0.9751	0.9747
ITC	0.9900	0.9792	0.9740	0.9775

Table 5.7: The performance of ITC vs other metrics (Linear Correlation)

5.7 Comparison of ITC with Other Image Quality Metrics

To validate the information theoretic criterion proposed in this thesis, several existing image quality criteria were applied to the image database set. The results from different image quality criteria, including PSNR, MSE, Mean Structure Similarity (MSSIM) [41], Universal Image Quality Assessment [42], visual information fidelity (VIF) [40] , and Information Theoretic Criterion (ITC) are shown below. In Tables 5.7 and 5.8, tests results are listed to illustrate the efficiency of different image quality measurements. Table 5.7 lists the most important feature of the measurement: linear correlation between objective judgments and subjective measurements and Table 5.8 lists the mean absolute error. The simulation results show that ITC outperforms the other image quality metrics. Visual Information Fidelity (VIF) is proposed by Sheikh, Bovik and Veciana. By implementing a degradation model to compute the mutual information between the reference and test images, VIF first estimated degradation parameters and then computed mutual information computation as quality criteria. Note that the method in this thesis does not need a degradation model - it has the ability to evaluate any distortion. From the simulation results in Tables 5.7 and 5.8 , ITC has better consistency to human judgment in Gaussian noise and Gaussian Blurring (absolute difference), but VIF has better performance in J2k distortion.

its

	Gaussian Noise	Gaussian Blurring	J2K Distortion	Overall
PSNR	0.1673	0.2896	0.3241	0.2577
MSE	0.4380	0.2858	0.3318	0.3519
MSSIM	0.1007	0.1974	0.1214	0.1122
Universal Image Quality Assessment	0.0810	0.1341	0.1214	0.1122
Non-reference Image Quality Assessment	-*	-	0.1290	-
VIF	0.0649	0.0842	0.0577	0.0693
ITC	0.0423	0.0448	0.0494	0.0450

Table 5.8: The performance of ITC vs other metrics (Absolute Error)

5.8 ITC based on Reduced Test Image

Sometimes, it is impossible to access the full image. For instance, during the transmission, partial information regarding the 'perfect version' is available, then how to estimate the image quality based on the limited reference information is a practical problem. Hence evaluation of the ITC based on a reduced image will be discussed in this section. In the simulations, the reduced test image is locate at the upper left part of the test image, and the size varies from 64×64 to 512×512. In the following table, the performance of ITC corresponding to different size and its computation time is listed. Although the smaller partial image will greatly reduce the computation time, the ITC performance will get worse.

ITC. decomposition

5.9 ITC using the Subbands in One Orientation

As each wavelet subband is treated independent of every other, in this section ITC is validated using the subbands in horizontal, vertical and diagonal directions individually. From the simulation results, it is clear that the performance of ITC changes slightly using different orientation. But there is no evidence to prove that using certain direction is better than other directions. For instance, ITC in the horizontal direction has better correlation with human judgments for Gaussian Noise

	Gaussian Noise	Gaussian Blurring	J2K Distortion	Overall
64×64	0.9162	0.8978	0.8835	0.8934
128×128	0.9598	0.9156	0.9210	0.9211
256×256	0.9643	0.9403	0.9658	0.9491
512×512	0.9771	0.9642	0.9734	0.9650
Full Size	0.9900	0.9792	0.9740	0.9775

Table 5.9: The performance of ITC based on reduced test image (Linear Correlation)

	Gaussian Noise	Gaussian Blurring	J2K Distortion	Overall
64×64	0.0887	0.0856	0.1084	0.0921
128×128	0.0612	0.0779	0.0898	0.0769
256×256	0.0594	0.0677	0.0617	0.0632
512×512	0.0497	0.0507	0.0528	0.0510
Full Size	0.0423	0.0448	0.0494	0.0450

Table 5.10: The performance of ITC based on reduced test image (Absolute Error)

Linear Correlation	Gaussian Noise	Gaussian Blurring	J2K Distortion
Horizontal	0.9900	0.9792	0.9700
Vertical	0.9831	0.0642	0.9740
Diagonal	0.9822	0.9700	0.9717

Table 5.11: The performance of ITC based on reduced test image (Linear Correlation)

Absolute Error	Gaussian Noise	Gaussian Blurring	J2K Distortion
Horizontal	0.0423	0.0448	0.0494
Vertical	0.0467	0.0460	0.0528
Diagonal	0.0430	0.0457	0.0533

Table 5.12: The performance of ITC based on reduced test image (Absolute Error)

and Gaussian Blurring, while ITC in the vertical direction has better consistency for J2k distortion.

5.10 Chapter Summary

The information theoretic criterion ITC introduced in Chapter 4 is evaluated in this chapter. Using the database set of 729 test images, ITC has been shown to be a reliable measurement tool for image quality: ITC has been shown to perform well over the wide range of scenes using different parameters of GSM. Meanwhile its prediction performance is equivalent to or even superior to other advanced image quality metrics, such as VIF, MSSIM. This chapter also proposed a new method to derive the parameters in HVS model in Chapter 3. Finally, we tried to find a tradeoff point between performance and running time.

Chapter 6

Conclusions and Future work

The goal of this thesis was to propose an image quality assessment method using an information theoretic framework. This chapter draws some conclusions by pointing out the contributions. Furthermore, this chapter explains how limitations of ITC raises the directions for future works.

6.1 Contributions and Limitations

The primary goal of this thesis is to propose a possible method to compute the perceptual information contained in images and try to evaluate the image quality based on perceptual information. The goal has been met with the following contributions:

6.1.1 Contributions of ITC

1. **HVS model**: This thesis presented a new human visual system model according to one essential human psychophysical feature. In the new model, human contrast sensitivity as the quantization procedure with the step size defined as the luminance threshold in different frequency domain. Hence the natural (continuous) signals will be represented as the perceptual (discrete) signals in the human brain. In mathematics:

If the image coefficient is X and the quantization step is $\triangle$, then the perceived signal in human brain is $X_{\triangle}$

$$X_{\triangle} = \frac{X}{\triangle}$$

2. **Perceptual information**: Based on a GSM, wavelet coefficients in each subband are modeled as a continuous statistical signal X by a mixture of Gaussian distributions.

$$X = \sqrt{z}u$$

Through the HVS model proposed in Chapter 3, the natural (continuous) signals will be processed to derive the perceptual signals. This thesis derives an analytical representation of the differential entropy contained in natural signals, which gives the relationship between the information in natural signal X and the information in perceptual signals $X_{\triangle}$

$$H(X_{\triangle}) = H(X) - log_2(\triangle)$$

in which the differential entropy of X is as derived in Chapter 4:

$$H(X) = H(X|z) + H(z)$$

3. **Minimum mutual information**: How to describe the relationship between a test image and its reference image? Minimum mutual information between the distorted signal and its reference signal is proved to be a proper description for two images. Generally speaking, the mutual information tells us how much information is still shared between the test image and its reference after various degradations. The problem with mutual information is the complication to depict the exact connections between the test image and its reference image, such as fade channel distortion in wireless communication. This makes it impossible to derive the mutual information or conditional information. In this thesis minimum mutual information is chose to act as the approximation of the relationship between a test image and its reference image.

4. **Information theoretic criterion**: As Chapter 4 provided an approximation for the common information shared by the test and reference images, this thesis defined an information theoretic criterion as normalized minimum mutual information:

$$ITC = \frac{I(X;Y)}{H(X)}$$

Various experiments show that ITC has good consistency to human judgments. Furthermore several simplified but efficient information theoretic criteria were proposed.

6.1.2 Limitations of ITC

Each image quality assessment method has its limitations. Generally speaking, ITC has several parts: wavelet decomposition, statistical modeling in each frequency band, human vision system implementation and perceptual information estimation. In each part, there are some assumptions made to achieve a good performance with limited computation resources and the limited knowledge we know about human vision.

In the statistical modeling part, one important assumption is that each subband is independent of the others, while that is not the fact. The coefficients in each subband have relationships with the other coefficients in the other subbands within the same scale. Furthermore, the coefficients have correlations with the ones in the other subbands across the scales. There are some statistical models which depict the statistical properties of an image acorss the scales, however as stated in Chapter II, the definition and computation of the perceptual information will be out of our scope.

In the implementation of the human vision system model, the contrast sensitivity in each frequency band is treated as a certain constant. It will not be adaptive to different types of images or viewing conditions. Moreover, the methods proposed to compute the parameters of human vision system is not perfect. Even the optimized method can just search the local optimization.

In the last part of ITC, we tried to find a method to get a good approximation of information defined based on the human vision introduced in Chapter 3. The method has good performance when large information is contained in the signal, while it has bigger estimation error for the signal with fewer information. Finally the lost perceptual information due to degradation is banded to avoid computing the exact lost information, for in the most cases there is no way to know the relationship between the original signal and the degraded one if not informed ahead, and it is also impossible to compute the exact lost information under some circumstances, such as fade channel distortion in the practical communications. The method proposed in this thesis is supposed to be a universal method for all kinds distortions and does not need to be informed of the degradation type. Then the tradeoff between the automation and performance of the metric must be achieved.

6.2 Future Research Direction

6.2.1 Blind Reference Image Quality Assessment

In this thesis, a literature review about the full reference image quality assessment is given, and an information theoretic metric is proposed and it achieved excellent consistency with human judgments. The method needs the original images namely references to compute the lost information. However, in practical applications, sometimes there is no way to access the original data, image quality has to be estimated based on the degraded one. This task is called blind reference image quality assessment. The task should be pay more attention in the future, and it will lead us to the better understanding of human vision system and statistical properties of images. In the next section, an introduction about the statistics across different scales is given to get us a sense of future research.

6.2.2 Correlation of Statistic Properties in Multi-resolution

There is high correlation between the wavelet coefficients in adjacent scales with same orientations. The images used to gather such properties are of high resolution (512x768) and each image is divided into blocks of 64x64 pixels. Hidden Markov Model is applied to each 64x64 pixel block.

Here is a brief introduction to Hidden Markov Model:

In each subband, the wavelet coefficients are described by two-state Gaussian mixture model. Each wavelet coefficient w_i is associated with an unobserved hidden state variable $s_i \in S, L$. The value of s_i dictates the components in the mixture model. State S corresponds to a zero-mean, low-variance Gaussian:

$$g(x, \mu, \sigma_S^2) = \frac{1}{\sqrt{2\pi}\sigma_S} exp\{-\frac{(x-\mu)^2}{2\sigma_S^2}$$

so for the other state L:

$$g(x, \mu, \sigma_L^2) = \frac{1}{\sqrt{2\pi}\sigma_L} exp\{-\frac{(x-\mu)^2}{2\sigma_L^2}$$

Then the distribution of a wavelet coefficient based on two-state Gaussian mixture model can be written as:

$$\begin{aligned} f(w_i) &= p_i^S f(w_i|s_i = S) + p_i^L f(w_i|s_i = L) \\ where f(w_i|s_i = S) &= g(x, \mu, \sigma_{S;i}^2) \\ f(w_i|s_i = L) &= g(x, \mu, \sigma_{L;i}^2) and \sigma_{L;i} > \sigma_{S;i} \end{aligned} \tag{6.1}$$

The following section shows some results about the relationship between the wavelet coefficients in different layers and orientations. In Figure 6.1, X axis is: $log_2(2\pi\sigma_L^2)$ in one layer, which means the information contained in State L, and Y axis is the information contained in State L in another adjacent layer.

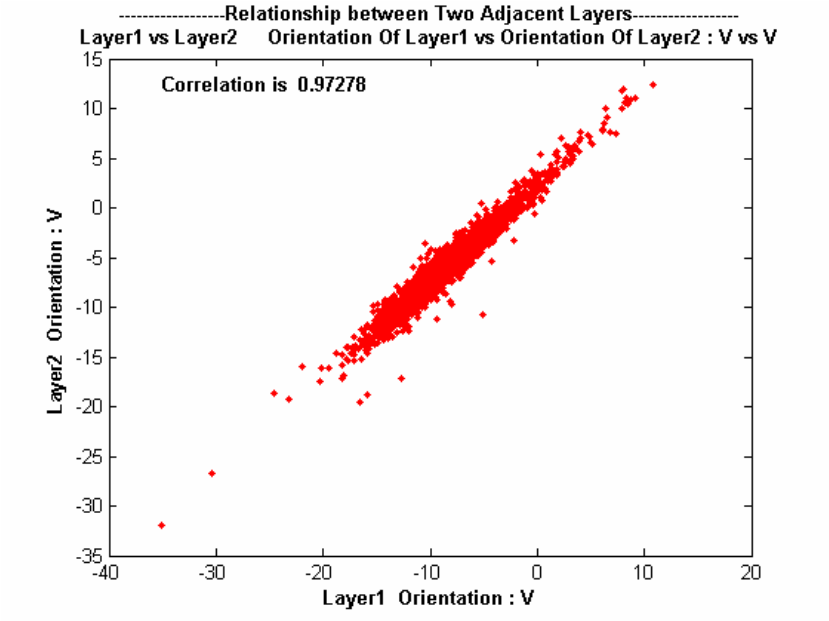


Figure 6.1: Relationship between Two Adjacent Layers

Our simulation results are listed in the following two tables which give us a sense of the correlations across different layers and different orientations.

The strong correlation between the wavelet coefficients within the two adjacent layers triggers an idea to access the blind assessment. Suppose the information is lost in one layer, based on the correlation it is possible to retrieve the lost information from its neighbor frequency bands. How to take the advantage of the strong correlation across layers is out of the scope of this thesis.

6.3 Chapter Summary

This chapter summarized the contents of the thesis, and outlined the advantages and limitations of ITC compared with other existed image quality metrics. The rest of this chapter made some efforts to depict the correlations across different frequency bands and explore the blind quality assessment.

	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	0.2963	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	0.1863	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

Figure 6.2: The relationship between the information contained in State L of wavelet coefficients in different layers and orientations

	1H	2H	3H	4H	1D	2D	3D	4D	1V	2V	3V	4V
1H	1	0.9705	0.9239	0.8355	0.0518	0.0043	-0.0708	-0.1276	0.2038	0.1460	0.0735	0.0003
2H		1	0.9551	0.8674	0.0302	0.0246	-0.0414	-0.0993	0.1520	0.1509	0.0793	0.0068
3H			1	0.8987	-0.0008	-0.0032	0.0082	-0.0629	0.0912	0.1063	0.1063	0.0206
4H				1	-0.0159	-0.0155	-0.0131	0.0794	0.0397	0.0424	0.0457	0.1507
1D					1	0.9769	0.9405	0.8807	0.1736	0.1559	0.1284	0.1098
2D						1	0.9663	0.9075	0.1224	0.1542	0.1298	0.1138
3D							1	0.9323	0.0657	0.0985	0.1492	0.1242
4D								1	0.0222	0.0547	0.0938	0.2231
1V									1	0.9719	0.9278	0.8483
2V										1	0.9580	0.8795
3V											1	0.9097
4V												1

Figure 6.3: The relationship between the information contained in State S of wavelet coefficients in different layers and orientations

Appendix I

Chapter I listed several image quality metrics with the simulation results in Table 1.1. It is necessary to make the details of implementations clear.

For Moments of the angels:

$$D_5 = \frac{2}{\pi} cos^{-1} \frac{< X, Y >}{||X|| \cdot ||Y||}$$

in which $< X, Y >$ is the cross product of vector X and Y, which X and Y are composed by the pixel values within two corresponding neighborhoods.

For Kullback-Leibler divergence (relative entropy):

$$D_7 = \int_{x \in \mathcal{X}} p_X(x) log(\frac{p_Y(x)}{p_X(x)}) dx$$

in which the distribution of X or Y is defined based on the pixel value.

For Mutual Information :

$D_{10} = I(X, Y) = \int_{x \in \mathcal{X}, y \in \mathcal{Y}} -p_{XY}(x, y) log(\frac{p_{XY}(x,y)}{p_X(x) p_Y(y)}) dx dy$

In the implementation, because the pixel value is discrete and varies between 0 and 255, the probability density of $p_X(x)$ or $p_{XY}(x, y)$ is changed to the discrete probability, meanwhile the integrals should be changed to summations.

Appendix II

This appendix gives the proofs of several theorems in Chapter 4. In this section, we will first prove a lemma of ration distortion theory, and then apply this lemma to Gaussian Scale Mixture Model used in this thesis.

Theorem 4.4: Consider a continuous variable X and its quantized random variable $X_{\triangle}$

$$H(X_{\triangle}) \rightarrow H(X) - log_2(\triangle) \qquad \triangle \rightarrow 0 \tag{2}$$

where $\triangle$ is the quantization step.

Proof:

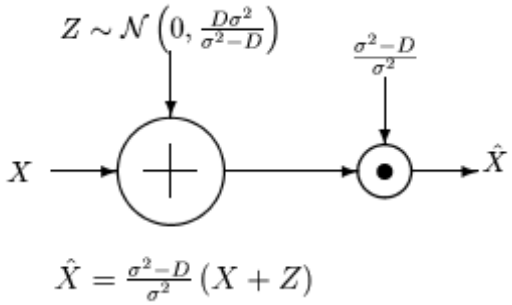
$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) && (1)\\ &= H(X) - H(X-Y|Y) && (2)\\ &\geq H(X) - H(X-Y) && (3)\\ &\geq H(X) - H(N(0, E[(X-Y)^2]) && (4)\\ &= H(X) - \frac{1}{2}log(2e\pi E[(X-Y)^2]) && (5)\\ &= H(X) - \frac{1}{2}log(2e\pi D) && (6) \end{aligned}$$

where (3) follows from the fact that conditioning reduces entropy and (4) follows from the fact that the normal distribution maximizes the entropy for a given second moment(*Theorem 4.3*). It is easy to prove that if the difference between the original signal X and distorted signal Y is Gaussian distributed, the lower bound $H(X) - \frac{1}{2}log(2e\pi D)$ can be achieved. Hence:

$$I(X;Y)_{min} \geq H(X) - \frac{1}{2}log(2e\pi D)$$

To prove the upper bound, let us consider the joint distribution as shown in the figure below, and calculate the distortion and the mutual information between X and Y. Since

$$\begin{aligned} Y &= \frac{\sigma^2 - D}{\sigma^2}(X+Z) && (7)\\ & && (8) \end{aligned}$$



Joint distribution for upper bound

$$
\begin{aligned}
E[(X-Y)^2] &= E(\frac{D}{\sigma^2}X - \frac{\sigma^2 - D}{\sigma^2}Z)^2 && (9)\\
&= (\frac{D}{\sigma^2})^2 E(X^2) + (\frac{\sigma^2 - D}{\sigma^2})^2 E(Z^2) && (10)\\
&= (\frac{D}{\sigma^2})^2 \sigma + (\frac{\sigma^2 - D}{\sigma^2})^2 \frac{D\sigma^2}{\sigma^2 - D} && (11)\\
&= D && (12)
\end{aligned}
$$

Also the mutual information is

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) && (13)\\
&= H(Y) - H(\frac{\sigma^2 - D}{\sigma^2}Z) && (14)
\end{aligned}
$$

Now

$$
\begin{aligned}
E(Y^2) &= (\frac{\sigma^2 - D}{\sigma^2})^2 E[(X+Z)^2] && (15)\\
&= (\frac{\sigma^2 - D}{\sigma^2})^2 (E(X^2) + E(Z^2)) && (16)\\
&= (\frac{\sigma^2 - D}{\sigma^2})^2 (\sigma^2 + \frac{D\sigma^2}{\sigma^2 - D}) && (17)\\
&= \sigma^2 - D && (18)
\end{aligned}
$$

Hence, we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(\frac{\sigma^2 - D}{\sigma^2}Z) && (19)\\ &= H(Y) - H(Z) - log(\frac{\sigma^2 - D}{\sigma^2}) && (20)\\ &\leq H(N(0,\sigma^2 - D)) - \frac{1}{2}log(2e\pi\frac{D\sigma^2}{\sigma^2 - D}) - log(\frac{\sigma^2 - D}{\sigma^2}) && (21)\\ &= \frac{1}{2}log(2e\pi(\sigma^2 - D))\frac{1}{2}log(2e\pi\frac{D\sigma^2}{\sigma^2 - D}) - log(\frac{\sigma^2 - D}{\sigma^2}) && (22)\\ &= \frac{1}{2}log(\frac{\sigma^2}{D}) && (23) \end{aligned}$$

Hence,

$$I(X;Y)_{min} \leq \frac{1}{2}log(\frac{\sigma^2}{D})$$

Combine the upper bound and lower bound, we can get

$$H(X) - \frac{1}{2}log(2e\pi D) \leq I(X;Y)_{min} \leq \frac{1}{2}log(2e\pi\sigma^2) - \frac{1}{2}log(2e\pi D) \quad (24)$$

If the original signal X and distorted signal Y are N dimensional and the distance D between X and Y is defined as $E[(X-Y)(X-Y)^T]$, we apply differential entropy of N-dimensional Gaussian distribution vector (4.17) to Theorem 4.3 to get Theorem 4.5

Theorem 4.5: If X is the original N-dimensional random vector, Y is the distorted vector from X, and distance between two vectors is $D = E[(X-Y)(X-Y)^T]$ the minimum mutual information between X and Y:

$$H(X) - \frac{1}{2}log_2((2\pi e)^N det(D)) \leq I_{min}(X;Y) \leq \frac{1}{2}log_2((2\pi e)^N det(C_X)) - \frac{1}{2}log_2((2\pi e)^N det(D)) \quad (25)$$

where C_X is the covariance matrix of X.

Following the ideas used to prove Theorem 4.3, we can easily get $I_{min}(X_\triangle; Y_\triangle)$. Please notice that each part in (25) corresponds to "a signal". $H(X)$ corresponds to the original signal X. $\frac{1}{2}log_2((2\pi e)^N det(D))$ corresponds to a Gaussian distributed vector G^1 with the covariance matrix D. And $\frac{1}{2}log_2((2\pi e)^N det(C_X))$

corresponds to a Gaussian distributed vector G^2 with the covariance matrix C_X. Apply Theorem 4.4 to each signal above if given the quantization step $\triangle$, then the entropy of each quantized signal is given:

$$H(X_\triangle) = \begin{cases} H(X) - Nlog_2(\triangle) & \text{if} \quad Nlog_2(\triangle) \leq H(X) \\ 0 & \text{if} \quad Nlog_2(\triangle) > H(X) \end{cases}$$

$$H(G^1_\triangle) = \begin{cases} \frac{1}{2}log_2((2\pi e)^N det(D)) - Nlog_2(\triangle) & \text{if} \quad Nlog_2(\triangle) \leq \frac{1}{2}log_2((2\pi e)^N det(D)) \\ 0 & \text{if} \quad Nlog_2(\triangle) > \frac{1}{2}log_2((2\pi e)^N det(D)) \end{cases}$$

$$H(G^2_\triangle) = \begin{cases} \frac{1}{2}log_2((2\pi e)^N det(C_X)) - Nlog_2(\triangle) & \text{if} \quad Nlog_2(\triangle) \leq \frac{1}{2}log_2((2\pi e)^N det(C_X)) \\ 0 & \text{if} \quad Nlog_2(\triangle) > \frac{1}{2}log_2((2\pi e)^N det(C_X)) \end{cases}$$

Combine the above three formulas, we can get the upper bound and lower bound of minimum mutual information between $X_\triangle$ and $Y_\triangle$:

$$I_{min}(X_\triangle; Y_\triangle) \leq \begin{cases} \frac{1}{2}log_2((2\pi e)^N det(C_X)) - \frac{1}{2}log_2((2\pi e)^N det(D)) \\ \quad if \frac{1}{2}log_2((2\pi e)^N det(C_X)) \geq log_2(\triangle^{-N}), \frac{1}{2}log_2((2\pi e)^N det(D)) \geq log_2(\triangle^{-N}); \\ \frac{1}{2}log_2((2\pi e)^N det(C_X)) - Nlog_2(\triangle) \\ \quad if \frac{1}{2}log_2((2\pi e)^N det(C_X)) \geq log_2(\triangle^{-N}), \frac{1}{2}log_2((2\pi e)^N det(D)) \leq log_2(\triangle^{-N}); \\ 0 \\ \quad else \end{cases}$$

$$I_{min}(X_\triangle; Y_\triangle) \geq \begin{cases} H(X) - \frac{1}{2}log_2((2\pi e)^N det(D)) \\ \quad if H(X) \geq log_2(\triangle^{-N}), \frac{1}{2}log_2((2\pi e)^N det(D)) \geq log_2(\triangle^{-N}); \\ \quad H(X) - Nlog_2(\triangle) \\ \quad if H(X) \geq log_2(\triangle^{-N}), \frac{1}{2}log_2((2\pi e)^N det(D)) \leq log_2(\triangle^{-N}); \\ \quad 0 \\ \quad else \end{cases}$$

Bibliography

[1] ITU-R Recommendation BT.500-10. Methodology for the subjective assessment of the quality of television pictures, March 2000.

[2] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE image quality assessment database release 2", http://live.ece.utexas.edu/research/quality.

[3] D. Andreutos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Distance measures for color image retrieval", IEEE International Conference On Image Processing, 1998

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

[5] M. Basseville,"Distance measures for signal processing and pattern recognition", Signal Process. Vol. 18, pp.349-369, 1989.

[6] J. West, J. M. Fitzpatrick, M.Y. Wang, B.M. Dawant, C.R. Maurer, et. al. "Comparison and evaluation of retrospective intermodality brain image registration techniques", Journal of Computer Assisted Tomography, 21(4):554-566, 1997

[7] C. Studholme, D.J. Hawkes, and D.L.G. Hill. "A normalized entropy measure for multi-modality image alignment", Proceedings of the SPIE, Vol 3338, pp.132-143, February 1998

[8] F. Maes, A. Collignon, D. Vandermeulen, and G. Marchaland P. Suetens. "Multimodality image registration by maximization of mutual information", IEEE Transactions on Medical Imaging, 16(20:187-198, April 1997

[9] Sang Hyun Kim and Rae-Hong Park, "A novel approach to scene change detection using a cross entropy", Proceeding of International Conference on Image Processing, Volume: 3, pp.10-13 Sept. 2000

[10] D. Geman, S. Geman, C. Graffigne, and P. Dong, "Boundary detection by constrained optimization", IEEE Trans. Pattern Anal. Machine Intell. 12(7), pp.609-628, 1990.

[11] J. Puzicha, T. Hofmann, and J. Buhmann, "Non-parametric similarity measures for unsupervised texture segmentation and image retrieval", in Proc. CVPR97, pp. 267-272, 1997.

[12] J. Puzicha, Y. Rubner, C. Tomasi, and J. Buhmann. "Empirical evaluation of dissimilarity measures for color and texture. In Int. Conf. on Computer Vision", volume 2, pp. 1165-1173, Corfu, Greece, Sept. 1999.

[13] Di Zhang and Ed Jernigan, Project of Medical Image Enhancement for Retina Image, Jan, 2005.

[14] Zhou Wang, Hamid R. Sheikh and Alan C. Bovik, "Objective video quality assessment",The Handbook of Video Databases, CRC Press, pp. 1041-1078, September 2003

[15] O.D.Faugeras, Fundamentals in Computer Vision : An Advanced Course, Cambridge University Press, 1983

[16] Human eye structure, http://library.thinkquest.org/

[17] EYE AND RETINA, http://thalamus.wustl.edu/course/eyeret.html

[18] Fiona Rowe, Visual Fields via the Visual Pathway

[19] F. W. Campbell, J. G. Robson, Application of Fourier analysis to the visibility of gratings., Journal of Physiology, pp.551-566.

[20] Mannos J., Sakrison D., The effects of a visual fidelity criterion of the encoding of images. IEEE Transactions on Information Theory 20, 4 (1974), pp.525-536.

[21] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images", IEEE Trans. Pattern Anal. Machine Intell., vol. 6, no. 6, pp. 721-741, 1984.

[22] J. Zhang, P. Fieguth and D. Wang, "Random field models", Handbook of Image & Video Processing, A. Bovik, ed., pp. 301-312.

[23] S. Li. "Markov Random Field Modeling in Computer Vision", Sproginer-Verlag, New York 1995.

[24] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", J. Royal Society, Series E, 36: pp. 192-236, 1974.

[25] A. Willsky. "Multi-resolution Markov models for signals and image processing", Proceedings of the IEEE, 90(8): pp. 1396-1458, 2002.

[26] Martin J. Wainwright, Eero P. Simoncelli, "Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images", Applied and Computational Harmonic Analysis, pp. 89-123.

[27] P. Moulin and J. Liu, "Analysis of multi-resolution image denoising schemes using a generalized Gaussian and complexity priors", IEEE Trans. Sig. Proc. 6, pp. 300-303 (1999).

[28] Y. Xu, J. Weaver and D.Healy, "Wavelet transform domain filters: a spatially selective noise filtration technique", IEEE Trans. on Image Processing, 3(6): pp. 747-758, 1994.

[29] M. Crouse, R. Baraniuk, "Contextual hidden Markov models for wavelet-domain signal processing", Proceedings of 31st Asilomar Conference in Signals, Systems and Computers, 1997.

[30] J. Romberg, H. Choi, and R. Baraniuk, "Bayesian tree-structured image modeling using wavelet domain hidden Markov models," in Proc. of SPIE Conf. on Mathematical Modeling, Bayesian Estimation, and Inverse Problems, 1999, vol. 3816, pp. 306-320.

[31] H. Choi and R. Baraniuk, "Multiscale image segmentation using wavelet-domain hidden Markov models", IEEE Trans. IP, vol. 10, no. 9, pp. 1309-1321, 2001.

[32] R. Nowak, "Multiscale hidden Markov models for Bayesian image analysis", in Bayesian Inference in Wavelet Based Models (B. Vidakovic and P. Muller, eds.), Lecture Notes in Statistics 141, Springer-Verlag, 1999.

[33] G. Fan and X.-G. Xia, "Improved hidden Markov models in the wavelet-domain", IEEE Trans. Signal Processing, 449: pp. 115-120, 2001.

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

[35] Q. Lin, "Halftone image quality analysis based on a human vision model", in Proc. SPIE, vol. 1913, Feb. 1993, pp. 378-389.

[36] Ivkovic, G.,Sankar, R., "An algorithm for image quality assessment", Acoustics, Speech, and Signal Processing, 2004. Proceedings, May 2004 Page(s):iii - pp. 713-716 vol.3.

[37] Claude Shannon, "A mathematical theory of communication", Bell System Technical Journal, vol. 27, pp. 379-423 and pp. 623-656, July and October, 1948.

[38] Niranjan Damera-Venkata, and Thomas D. Kite, Wilson S. Geisler, Brian L. Evans, and Alan C. Bovik, "Image quality assessment based on a degradation model",IEEE Transactions on Image Processing, vol. 9, pp. 636-650, Apr. 2000.

[39] Sheikh HR, Bovik AC, "Image information and visual quality", IEEE Transactions on Image Processing, Vol: 15 No: 2, February 2006 pp. 430 - 444.

[40] Zhou Wang, Alan C. Bovik, "Image quality assessment: from error visibility to structural similarity", IEEE Transactions on Image Processing, Vol. 13, No. 4, April 2004.

[41] Z. Wang and A. C. Bovik,"A universal image quality index", IEEE Signal Processing Lett., vol. 9, no. 3, pp. 81-84, Mar. 2002.

[42] S. Golden, "Identifying multiscale statistical models using the wavelet transform", Ma. Sc. Thesis, Massachusetts Institute of Technology, 1991.

[43] B. Ripley, "Statistical inference for spatial processes", Cambridge University Press, 1988.

[44] S. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising", IEEE Trans. on Image Processing, 9: pp. 1522-1531, 2000.

[45] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach", J. R. Statis. Soc. B, 60: pp. 725-749, 1998.

[46] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges", IEEE Trans on PAMI, 14: pp.710-732, 1992.

[47] Foveal and Peripheral Vision, http://webexhibits.org/colorart/ag.html