Image Quality Assessment Using Frequency Domain Transforms

Mireille Sendashonga

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

October 2006

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Engineering.

© 2006 Mireille Sendashonga

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-28621-0
Our file Notre référence
ISBN: 978-0-494-28621-0

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Measurement of image quality plays a central role in optimization and evaluation of imaging systems. The most straight-forward way to assess image quality is subjective evaluations by human observers, where the mean value of their scores is used as the quality measure. However, objective (quantitative) measures are needed because subjective evaluations are impractical and expensive. The aim of this thesis is to develop simple and low-complexity metrics for quality assessment of digital images.

Traditionally, the most widely used quantitative measures are the mean squared error and measures that model the human visual system. The proposed method uses the Discrete Cosine Transform and the Discrete Wavelet Transform to divide images into four frequency bands and relates the visual quality of the distorted images to the weighted average of the mean squared error between original and distorted images within each band.

The performance of the metrics presented in this thesis is tested and validated on a large database of subjective quality ratings. Simulations show that the proposed metrics accurately predict visual quality and outperform current state-of- the-art methods with simple and easily implemented processing steps.

Extensions of the proposed image quality metrics are investigated. More particularly, this thesis explores image quality assessment when the reference image is only partially available (reduced reference settings), and presents a method for successfully quantifying the quality of distorted images in such settings.

Sommaire

La mesure de la qualité d'image joue un rôle central dans l'optimisation et l'évaluation des systèmes d'imageries. La façon la plus simple d'évaluer la qualité d'image est de solliciter l'opinion de sujets humains (évaluations subjectives) et d'utiliser la moyenne de leurs scores individuels comme mesure de qualité. Toutefois des mesures objectives (quantitatives) de qualité sont nécessaires puisque les évaluations subjectives sont côuteuses. L'objectif de cette thèse est de développer des mesures simples pour estimer la qualité d'images digitales.

Traditionnellement, la qualité des images est évaluée par une mesure de la dégradation en terme d'erreur moyenne quadratique ou par des méthodes qui consistent à mesurer l'erreur de visibilité entre une image dégradée et une image de référence en simulant les propriétés connues du système visuel humain. La méthode proposée dans cette thèse utilise la Transformée en Cosinus Discrète et la Transformé en Ondelettes Discrète pour diviser les images en quatre bandes de fréquence; la qualité visuelle de l'image dégradée est proportionnelle à la moyenne pondérée des erreurs (différences) entre l'image de référence et l'image dégradée dans chaque bande de fréquence.

La performance des mesures de qualité proposées dans cette thèse est évaluée et validée par des expériences sur une large base de données d'images. Les simulations montrent que les mesures proposées dans cette thèse prédisent la qualité visuelle d'images dégradées de façon précise tout en étant simples et faciles à implémenter.

Cette thèse examine également le problème de la mesure de la qualité d'images dans les cas où l'image de référence n'est que partiellement disponible et présente une méthode, validée par des simulations, pour estimer la qualité d'image dans ces cas.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Fabrice Labeau, for his guidance, valuable advice, feedback and financial support during the course of my graduate studies.

Thanks to all my fellow graduate students in the TSP Lab for providing a friendly and supportive work environment.

I would like to acknowledge Dr. Hamid Rahim Sheikh of the University of Texas at Austin for giving me access to the LIVE quality assessment database.

On a more personal note, I'd like to thank my friends Joëlle, Rahel, Ahmed and Jean-Aimé for their companionship, moral support and encouragement.

Last but not least, I am deeply indebted to my parents, my husband, my brother and sister and the rest of my family for their unconditional love, everlasting encouragement and unwavering belief in me throughout my endeavors.

Contents

List of Figures

List of Tables

List of Acronyms

CC	Correlation Coefficient
CSF	Contrast Sensitivity Function
DMOS	Difference Mean Opinion Score
FR	Full Reference
HVS	Human Visual System
LGN	Lateral Geniculate Nucleus
MAE	Mean Absolute Error
MOS	Mean Opinion Score
MSE	Mean Squared Error
NR	No Reference
OR	Outlier Ratio
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
QA	Quality Assessment
RMSE	Root Mean Squared Error
RR	Reduced Reference
SROCC	Spearman Rank Order Correlation Coefficient
SSIM	Structural Similarity Index Metric
VQEG	Video Quality Experts Group

Chapter 1

Introduction

1.1 Importance of Image Quality Assessment

Humans are highly visual creatures. A large part of the brain's neurological resources is devoted to visual perception and humans rely heavily on visual information to transmit information: when it comes to communicating, an image is worth a thousand words. It comes as no surprise that images, specifically digital images, are an integral part of modern life. The applications range from digital photography to medical imaging, from image-based web search to satellite images in weather broadcast. This proliferation of digital images has been fueled by developments in the technologies underlying the capture, transfer, storage and reproduction of digital media.

As these images move through imaging systems, from the point where they are captured to the point where they are viewed on a screen or on paper by a human viewer, each intervening processing stage can introduce visible degradations in the final image output. Acquisition systems are not perfect and sometimes introduce perceptible distortions (e.g. optical blur caused by lens). Improper use of equipment is another source of image degradation at the acquisition stage (e.g. motion blur caused by camera movement). Once captured, digital images are often encoded and compressed to reduce the bandwidth needed to store and transmit them. Lossy compression algorithms introduce compression artifacts which may further deteriorate the perceived quality of images. Modern transmission channels, such as wireless channels or the Internet, are prone to interference, transmission errors, delays and packet losses which also affect the overall quality of images at the receiver side. Given the large number of potential sources of quality degradation that plague practical imaging systems, image quality assessment becomes a necessity.

Measurement of perceived quality plays a central role in the specification, design, testing and operation of imaging systems and has both offline and online applications. Image quality measures may be used to assess the improvement of quality of image restoration and enhancement algorithms. They can also be used in the design of image coders: image coding is essentially an optimization procedure that attempts to maximize perceived image quality with a limited number of bits and image quality measures can serve as guides for bit assignment. They can be used to benchmark image processing systems and algorithms (e.g. comparison between different compression algorithms). They can also be used in a real-time framework to monitor and control image quality over a network.

For the last three decades, researchers have attempted to develop methods to accurately assess image quality [1].

1.2 Measures of Image Quality

Since humans are the ultimate end users of most image processing applications, the most straight-forward way to assess image quality is to solicit the opinion of human observers. The process of collecting and processing the opinion of human subjects to obtain a mean opinion score (MOS) has been standardized [2]: if executed according to established guidelines, subjective evaluations are perhaps the closest we can get to the truth about perceived quality. However, conducting subjective quality evaluations is a complex and cumbersome process that requires a large number of human observers, strictly controlled experimental conditions and several lengthy test sessions. It should also be noted that subjective rating results may not be reproducible as the observers' ratings can be influenced by factors such as environmental conditions, motivation and mood. While these tests can be considered as a benchmark for image quality measurement, they cannot be used for practical purposes. It is not economical, or even possible, to solicit human opinion each time an image has to be evaluated. Furthermore, because they are time-consuming, subjective evaluations cannot be incorporated in real-time applications.

Objective quality metrics eliminate the need for expensive subjective studies by automating the quality estimation process. The ultimate goal of objective image quality assessment research is to develop quality metrics that are closely related to quality as experienced by human viewers. Ideally, these metrics should be generic: they should give accurate quality estimation regardless of the image content and type of the distortion.

1.3 Objective Image Quality Assessment

1.3.1 Full, reduced and no-reference image quality assessment

Objective quality metrics can be classified into three categories. Full reference (FR) metrics compare a distorted image against a reference image, which is considered to be distortion-free. Reduced reference (RR) metrics are designed to predict the quality of distorted images with only partial information about the reference images: a reduced set of features or descriptors are extracted from the reference and distorted images and quality is estimated based on this reduced amount of information. No reference (NR) metrics evaluate image quality blindly (without any reference information).

Although humans can judge image quality without any explicit reference, objective NR image quality assessment is a very difficult task and there is no generic NR quality metric in literature: most proposed metrics limit their scope to specific distortion types that must be known *a priori* such as blocking artifacts [3, 4, 5, 6] or blur and ringing artifacts [7, 8]. The development of RR metrics has been limited by the ability to find good general-purpose RR features that yield accurate image quality prediction with minimal data rate. Only a handful of RR methods have been proposed in the literature, and most they of them can only be used for specific distortion types (e.g. JPEG and JPEG2000 compression artifacts in [9]). To the best of our knowledge, only one general-purpose RR image quality assessment metric exists in literature [10]. The applicability of FR measures is much wider and therefore these metrics have received the most attention.

1.3.2 Approaches to full-reference image quality assessment

FR algorithms evaluate the quality of an image by comparing it against a reference image. In other words, they measure the similarity or fidelity between two images. The most obvious way to measure the similarity between two images is to compute an error signal by subtracting the test signal from the reference and then computing the average energy of the error signal: the resulting value is conventionally known as the Mean Squared Error (MSE) and is the most widely used FR metric. However, it does not correlate well with subjective image quality. This has led researchers to adopt a psychophysical viewpoint where the comparison between the test and reference images is done in a way that mimics the different processing stages of the human visual system. The ultimate goal of "bottom-up" approach of the QA problem is to build systems that functions the same way as the HVS. On the other hand, some researchers have taken a "top-down" approach to bypass the challenges of modeling the HVS and developed QA methods that are based on the hypothesized overall functionalities of the entire HVS.

1.3.3 Performance evaluation of objective quality assessment metrics

The ultimate goal of objective quality assessment metrics is to make quality predictions that are in agreement with the subjective opinion of human observers. Therefore objective metrics are evaluated against benchmark data sets (databases) of mean opinions scores (MOS) which are image quality ratings by human subjects. Qualitatively, objective image quality metrics are typically evaluated with respect to three attributes:

- Prediction accuracy - the ability to predict subjective quality ratings with low error
- Prediction monotonicity - the degree to which the predictions made by the objective quality metric agree with the relative magnitudes of subjective quality ratings
- Prediction consistency - the degree to which the objective quality metric maintains prediction accuracy over a wide range of impairments

Because subjective quality assessment studies are cumbersome and time-consuming, most quality assessment researchers have considered it sufficient to assess the performance of their metrics based a limited set of subjective quality rating scores. For example in [11], the entire data set was derived from only three reference images (distorted by compression distortion only). However, more extensive subjective quality rating data sets are needed in order for the performance assessment of image QA metrics to be statistically significant. To the best of our knowledge, apart from video quality studies conducted by the Video Quality Experts Group (VQEG), the largest database of subjective quality ratings is the LIVE database [12]. This publicly available database, which will be used as a benchmark for the simulated and proposed metrics in this thesis, is the result of an extensive subjective quality assessment study by the Laboratory for Image and Video Engineering (LIVE) and Center for Perceptual Systems (CPS) at the University of Texas at Austin. This study used 982 images distorted using five different distortions types (JPEG2000 compression, JPEG compression, white Gaussian noise, Gaussian blur and bit errors in a JPEG2000 bitstream transmitted over a fast-fading Rayleigh channel) and involved more than 20,000 human quality evaluations [13].

1.4 Thesis Description and Organization

This thesis presents a low complexity image quality assessment method based on frequency domain transforms. The organization of the thesis is as follows. Chapter 2 presents the background of objective full reference image quality assessment. It begins with a brief overview of the anatomy and properties of the human visual system. Different objective methods used in image quality assessment are reviewed and categorized according to their underlying principles. Chapter 3 presents the proposed full reference objective image quality assessment method. Two image quality measures based on frequency domain transformations are introduced: Q_{DCT} and Q_{DWT} are based on the Discrete Cosine Transform and (DCT) Discrete Wavelet Transform (DWT) respectively. These metrics express the quality of distorted images numerically as a scalar value and graphically as a quality map. The validity of the proposed metrics is investigated by correlating their estimation results with subjective scores and comparing them with the MSE and a state-of-the-art image quality metric. Chapter 4 extend the scope of the proposed method to reduced reference frameworks. Chapter 5 summarizes the thesis, highlights the contributions and suggests some recommendations for future work.

Chapter 2

Full Reference Image Quality Assessment: Background

Researchers have primarily focused on full reference image quality metrics because of their wider applicability: they can be used to estimate a wide spectrum of distortions. In this chapter, we present some full-reference methods that have been proposed in the literature. The simplest measure is the mean squared error (MSE) and more advanced metrics can be divided in two categories: those that rely on modeling the human visual system and those that use arbitrary signal fidelity criteria. A brief introduction to the human visual system (HVS) and the HVS properties relevant for quality assessment purposes is provided in Section 2.1. This will lay a foundation for better understanding of the material in subsequent sections.

2.1 The Human Visual System

Human vision is a complex process that requires numerous optical, synaptic, photochemical, and electrical components to work together. Fig. 2.1 shows the components of the the early stages of the HVS. The functionality of the higher layers of the HVS (i.e. how the human brain extracts higher-level cognitive information from the visual stimulus) is far from being well understood and is currently an active research topic in several disciplines including biology, anatomy, psychology, and physiology. However, the functions of each of the lower level components of the HVS (eyes, lateral geniculate nucleus and primary visual cortex) are fairly well understood and their description could fill volumes. A detailed description of the HVS may be found in [14] and [15]. This section is not intended to be a thorough review of the HVS: the anatomy and properties of the lower level components of the HVS are discussed only to the extent to which they impact image quality.

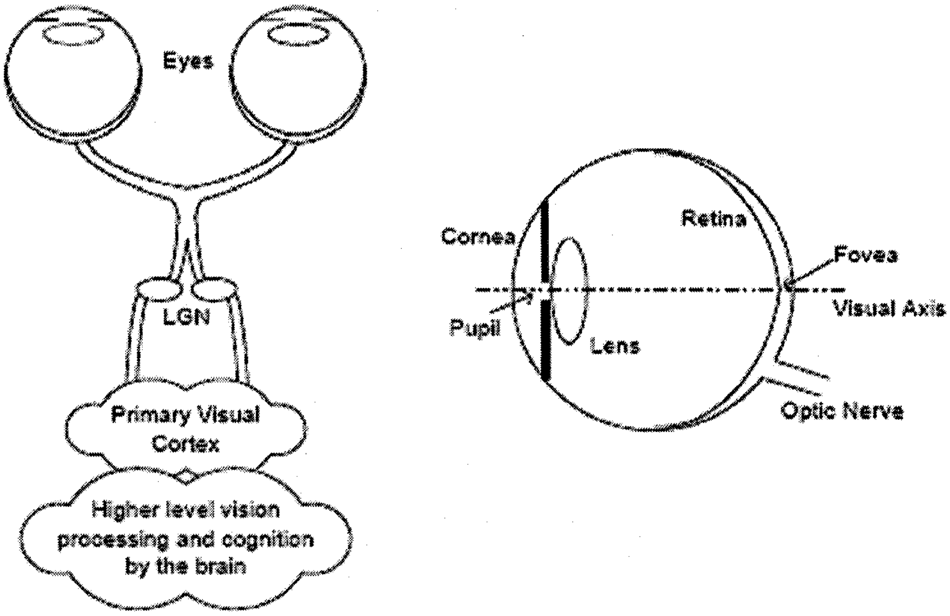


Fig. 2.1 Schematic diagram of the human visual system [13]

2.1.1 Anatomy of the human visual system

From an image processing point of view, the early HVS can may divided into four stages: optical processing, retinal processing, LGN processing and cortical processing.

The optical system of the eye is composed of three main components: the cornea, the pupil and the lens. Visual stimuli in the form of light rays enter the eyes through the cornea. The light rays pass through the pupil which controls the amount of light entering the eye. The lens focus the light rays onto the retina at the back of the eye.

Photoreceptor cells on the retina capture, sample and encode the focused image by converting the visual stimulus (light) into neural signals (electrical impulses). There are two types of photoreceptors cells: cones and rods. Rods are responsible for vision at low light levels (scotopic conditions). However, they do not distinguish between colors and have low visual acuity (a measure of detail) and are therefore generally neglected in

HVS modeling for quality assessment purposes. Cones are responsible for vision in normal light conditions (photopic conditions) and provide humans with basic color vision. There are three different types of cones each sensitive to a different portion of the visible light spectrum: L-cones, M-cones and S-cones, sensitive to long, medium and short wavelengths (different colors) respectively. The three types of cones split the image into three visual streams which can be crudely approximated to the Red, Green and Blue color components.

The neural signals generated in the cones pass through several layers of neurons in the retina before being carried off to the brain by the optic nerve. These signals are reorganized in the optic chiasm and the lateral geniculate nucleus (LGN): the visual signals from the left visual field are projected onto the right LGN while signals from the right visual field are projected onto the left LGN. The fibers from the LGN enter the visual cortex where vision processes such as detection and discrimination are performed by the neurons which are tuned to various aspects of the incoming streams such as spatial and temporal frequencies, orientations and directions of motion. The visual streams generated in the cortex are carried off into higher layers of the brain for further processing which involves processes such as motion sensing and cognition to arrive at a single interpretation (i.e. quality measurement or decision regarding the visibility of artifacts).

2.1.2 Properties of the human visual system

Neurophysiological and psychophysical studies are the primary source of information regarding the overall functionality of the HVS. These studies have highlighted several important features and properties of the human visual system some of which are relevant to image quality assessment.

Intra-eye blurring

Due to inherent limitations and imperfections of the eye optics (e.g. refraction, diffraction), the retinal image turns out to be a distorted version of the input. The most noticeable distortion is blurring. This low-pass blur is typically modeled as a linear space-invariant filter characterized by a Point Spread Function (PSF).

Foveal and peripheral vision

The density of photoreceptors cells varies across the retina. They are more tightly packed on the point that lies on the visual axis in the center of the retina (fovea). The result of this unequal distribution is that whenever a human observer fixates a point, the region around the fixation point is resolved with the highest resolution (foveal vision) while surrounding regions are resolved with progressively lower resolution (peripheral vision). Most image quality assessment models concentrate their modeling efforts on foveal vision and neglect peripheral vision.

Light adaptation and contrast sensitivity

The HVS operates over a very wide range of light levels from a dark night to a bright sunny day. Yet the dynamic range of neurons is nowhere near this. Light adaptation occurs in order to adjust to these changing conditions of illumination. On one hand, the pupil adjusts its size to control the amount of light entering the eye: when the light impinging upon the eye increases, the pupil diameter reduces and conversely in low-light condition the pupil gets larger letting more light into the eye. On the other hand, the retina copes with variable lighting conditions by encoding the contrast of the visual stimulus instead of encoding the absolute light intensities: this non-linear transformation that maintains the contrast sensitivity of the HVS over a wide range of background light intensities in known as *Weber's law.*

Spatial frequency sensitivity and orientation sensitivity

Psychovisual studies have shown that contrast sensitivity varies with spatial frequency. The HVS is much more sensitive to lower spatial frequencies than high ones. This is typically modeled by a contrast sensitivity function (CSF). The CSF is slightly band-pass in nature but most image quality assessment model of the HVS implement a low-pass version. It has also been shown that the HVS is not isotropic: the response of the HVS is maximum at horizontal and vertical directions and decreases to a minimum at an angle of 45 degrees. In reality the CSF is a multivariate function of spatial frequency, temporal frequency (which is irrelevant for image quality assessment but is modeled for video quality assessment), orientation, viewing distance and color direction.

Multi-channel structure of cortical neurons

Each neuron in the primary visual cortex is tuned to specific frequencies and orientations (band-limited response). This is equivalent to having several independent visual channels. The collection of these channels spans the full range of visual spatial frequencies and orientations. The CSF is the overall response of the ensemble of these neurons.

Masking

Masking characterizes the response of the HVS to a combination of several signals (image components). A stimulus is perceived differently as a function of the background (mask) onto which it lies. The visibility of artifacts (or errors) is decreased (masked) in active areas of an image (strong edges, strongly textured areas). Similarly, errors in smooth or homogeneous areas are more easily detected. It has been shown that the masking effect is strongest when the stimulus and the mask are closely coupled (in terms of frequency and orientation).

2.2 Mean Squared Error

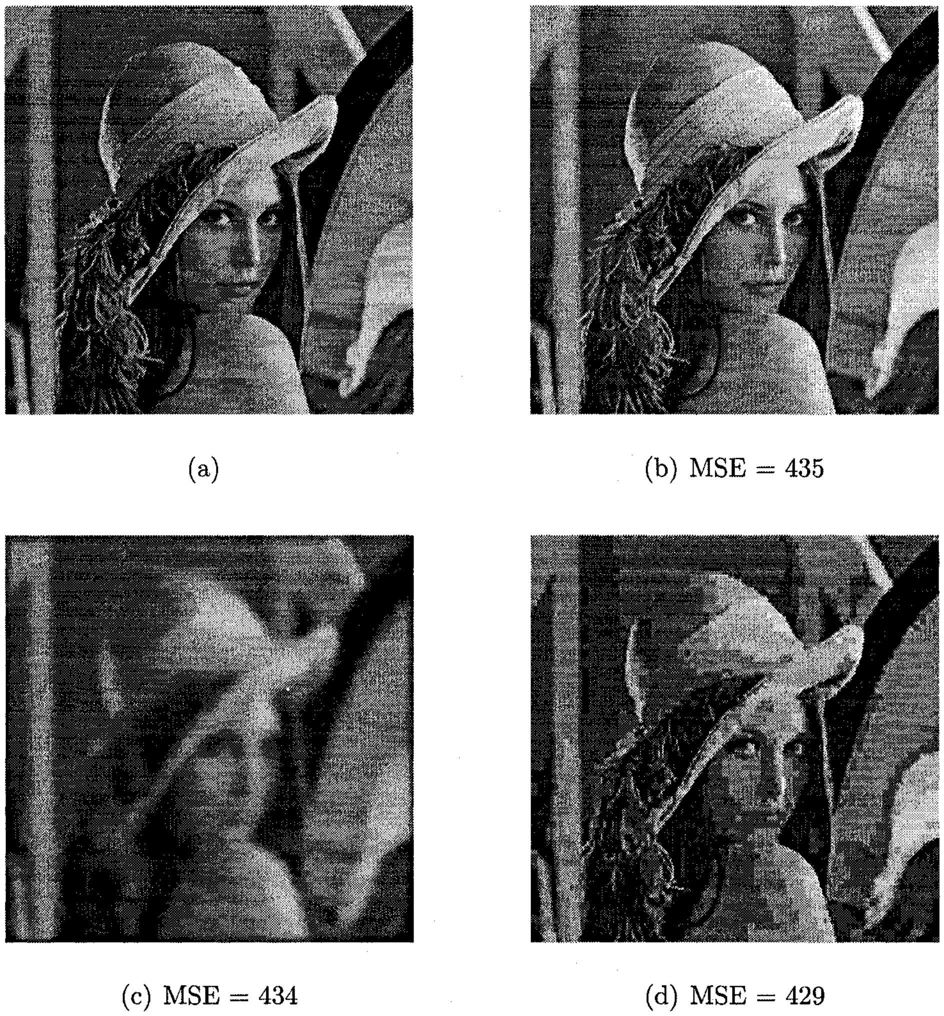
The simplest and still most widely used objective full reference objective image quality metrics are the mean squared error (MSE) and the related peak signal-to-noise ratio (PSNR), which are defined as:

$$\mathrm{MSE} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} (X_{ij} - Y_{ij})^2 \tag{2.1}$$

$$\mathrm{PSNR} = 10 \log_{10} \frac{L^2}{\mathrm{MSE}} \tag{2.2}$$

where X is the reference image, Y is the distorted image and M, N are the dimensions of the images. L is the dynamic range of the pixel values (for standard 8 bits/pixel images L is equal to 255).

The MSE and PSNR are appealing measures because they are easy to compute and are mathematically simple to deal with for optimization purposes. However, the predictive performance of the MSE and PSNR relative to human perception is poor. Fig. 2.2 clearly shows that equal values of MSE for two images does not necessarily imply equivalent quality: the three distorted images with the same amount of error energy have different structure of errors and hence different perceptual quality. A slight spatial shift of an image produces a large numerical MSE, but no significant visual distortion. Conversely, a small MSE can result in a very noticeable visual artifact if the total error is concentrated in a small area, which also happens to be the primary region of interest in the image. The problem lies in the fact that the MSE is based on simple pixel-to-pixel difference calculations and treats all errors equally regardless of their type [16]; it does not take into account the perceptual properties of the HVS discussed in Section 2.1. For instance, the MSE does not consider the fact that the sensitivity of the human visual system is different for different types of errors depending on their spatial frequency, orientation, spatial position and visual context. Also, simple error summation, as in Eq. 2.1, may differ considerably from the way the primary cortex and higher level cognitive processes pool errors to arrive at an assessment of the perceived distortion.



(a)

(b) MSE = 435

(c) MSE = 434

(d) MSE = 429

Fig. 2.2 Lena image altered with different types of distortions: (a) Original image 512 x 512, 24 bits/pixel; (b) Mean-shifted image; (c) Blurred image; (d) JPEG compressed image. Images (b)-(d) have similar MSE but radically different visual quality.

In [11] and [17], a number of variants of the MSE (peak MSE, Laplacian MSE, normalized MSE, normalized absolute error) and other pixel-based metrics are investigated. It is shown that although some of these measures correlate well with subjective evaluations results for a given compression technique, they are not reliable across different techniques.

Image quality metrics in the last three decades have tried to improve upon the MSE by incorporating knowledge about perceptual properties of the human visual system.

2.3 Image Quality Assessment Based on Human Visual System Modeling

In an attempt to improve upon the MSE, researchers have developed image quality metrics that simulate the processes of human vision from the eye to the visual cortex and try to sequence them in the way they occur in the HVS. These metrics come in different flavors based on tradeoffs between accuracy in modeling the HVS and computational feasibility. They can be classified into two types, namely single channel and multiple channel models. Single channel models do not decompose the image into several channels and assess quality based on single-scale representations of the image. Multiple channel models decompose the image into multiple spatial frequency and/or orientations bands and perform CSF and masking separately for each channel before pooling the results to obtain either a single numerical value that quantifies the perceptual dissimilarity between the reference and distorted image or a number than represents the probability that a human eye will detect a difference between the two images. Alternatively, the output can be a map of perceptual dissimilarities or a map of detection probabilities. The aim of this section is not to present an exhaustive review and detailed description of HVS-based methods but we discuss a few of the most representative methods. An extensive review of HVS-based metrics may be found in [18].

The first attempt to incorporate HVS modeling was made by Mannos and Sakrison [19] in 1974. This single channel model consists of a luminance non-linearity that models light adaptation followed by a linear space invariant element to represent the spatial frequency sensitivity of the HVS. The quality rating is the mean squared error between the model outputs of the original and distorted images. Despite its simplicity, this work pioneered the field of image quality assessment because it was the first work that linked the field of image processing with the field of vision science.

Subsequent HVS based full-reference quality metrics employ more sophisticated models. All of these metrics share a similar pipeline structure shown in Fig. 2.3.

The pre-processing stage ensures that the reference and distorted images are properly

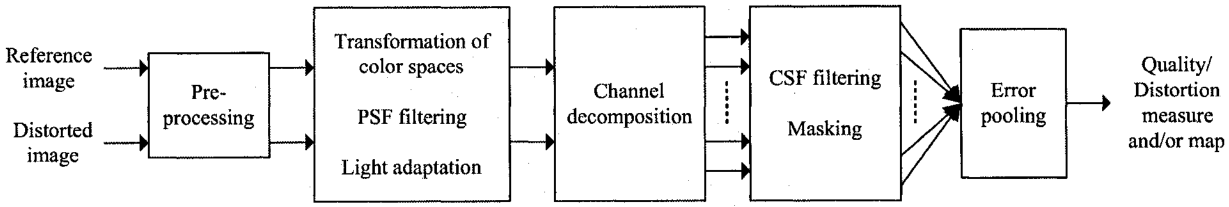


Fig. 2.3 Framework of HVS-based quality assessment systems

calibrated and aligned. Input images may have undergone a number of different transformations and may originate from different display devices. Most quality metrics require that the digital pixel values stored in computer memory be calibrated for display devices by converting them to luminance values of pixels on the display device. Registration establishes point-to-point correspondence between the reference and distorted images.

In the second stage, the images may be converted to a color space that conforms better to the HVS. The traditional Red, Green, Blue (RGB) color space may not be the best way since the RGB channels are highly correlated which is due to significant overlap between the L-, M- and S-cone sensitivities. Most algorithms designed for color image convert the RGB values to other color spaces (e.g. opponent color space, YUV). At the stage, a low-pass filer simulating the PSF may be applied. Finally, the reference and distorted images need to be converted into corresponding contrast stimuli to simulate light adaptation. One way to achieve this is by a nonlinear transformation. Commonly used transformation include conversion to density (log) and various power laws (e.g. cube root). However, some metrics choose to implement the contrast calculation later in the system, during or after the channel decomposition.

Channel decomposition models the frequency and orientation selective channels in the HVS: the input images are divided into different spatial and orientation subbands. In an effort to accurately model cortical neurons, some quality assessment algorithms implement sophisticated channel decomposition (Cortex Transform, Gabor decomposition). However, simpler decompositions such as the discrete cosine transforms or wavelet transform are also used because of their suitability for certain applications (coder-specific applications).

In the fourth stage, CSF filtering is typically implemented as weighting factors for each subband. The errors between the reference and distorted images are computed for each channel. Most models implement masking in the form of a gain-control mechanism that weights the error in a channel by a visibility threshold for that channel. If the magnitude of the error is less than the threshold (also called the Just Noticeable Distortion or JND level), the error in that particular channel will be indistinguishable.

In the final stage, the errors from the various frequency and orientation selective streams (channels) are pooled into a single number for each pixel (yielding a quality/distortion map) or a single numerical value E for the whole image using Minkowski pooling:

$$E = (\sum_{l} \sum_{k} |e_{l,k}|^{\beta})^{1/\beta} \tag{2.3}$$

where $e_{l,k}$ is the error of the k-th coefficient in the l-th subband (channel) and β is a constant with value typically between 1 and 4.

Daly's Visible Differences Predictor (VDP) [20], developed for the evaluation of high quality imaging systems, is probably the most elaborate HVS-based image quality metric. This multiple channel model produces a probability of detection map between the reference and distorted images. Each point on the map describes the probability that a human observer will perceive a difference between the reference and distorted images at that point. Initially, a nonlinear response function is applied to each of the input luminance images to account for light adaptation and the non-linear response of retinal neurons. The transformation is done using cube-root power law at low luminance levels and logarithmic dependence at high luminance levels. The images are then converted into the frequency domain and weighted by a contrast sensitivity function (CSF). An orientation and frequency selective Cortex Transform splits the images into five frequency and six orientation channels. Combined with an orientation independent base frequency band this gives a total of 31 channels. To model masking in each channel, a threshold elevation map is computed from the mask contrast in that channel. A psychometric function converts error strengths (weighted by the threshold elevations) into a probability-of-detection map for each channel. Pooling is carried out across the channels to obtain an overall detection map.

Lubin's model [21, 22] was developed for display evaluation. Similarly to Daly's model, it estimates a detection probability of the differences between the original and distorted images. First, the input images are blurred to model the PSF. They are then re-sampled to reflect the unequal photoreceptor sampling in the retina (peripheral vision). Channel decomposition is achieved through a Laplacian pyramid which decomposes the images into seven spatial frequency resolutions followed by a set of steerable orientation filters which decompose the signal in four orientations yielding a total of 28 channels. The filtered images are then converted to units of contrast. The CSF is modeled by normalizing the output of each frequency-selective channel by the baseline contrast sensitivity for that channel. A transformation by a sigmoid non-linearity models masking in each channel. Errors in each channel are pooled into a distortion map using Minkowski pooling across frequency. A single number for the entire image may be obtained by applying an additional pooling stage.

Teo and Heeger [23] developed a multi-channel metric that incorporates PSF, luminance masking and combines contrast sensitivity and masking into a single step called contrast normalization. The channel decomposition process used quadrature steerable filters with four spatial frequency resolutions and six orientation levels. The contrast normalization takes the output of channels at all orientations at a particular frequency. The contrast normalization model uses parameters which were chosen to fit the authors' experimental data. Therefore, this model is tailored to a specific set of conditions and would require additional optimization to be adapted to a new set of conditions.

Some HVS-based models have been developed specifically for image compression applications. They have the same general structure as the algorithms described above, the only difference being that they the adopt frequency decomposition of a given coder. They are considerably simpler than the models discussed above since they only have to consider the properties of the HVS that are relevant to image compression.

Watson's DCT metric [24] is based on the 8 × 8 DCT commonly used in image compression standards (e.g. JPEG). The first step is to convert the reference and distorted images into a luminance/chromincance color space. The luminance components are then partitioned into 8 × 8 pixel blocks and transformed to the frequency domain using the DCT. A visibility threshold is computed for each of the 64 subbands within each block. The visibility threshold is determined by three factors: the baseline contrast sensitivity associated with the DCT component (determined empirically in [25]), luminance masking and contrast/texture masking. These thresholds are used to weight the error in each subband. The errors in each subband are pooled spatially using Minkowski pooling. Then, the errors are pooled across frequency to obtain a single distortion value. A distortion map may be obtained by skipping the spatial error pooling step and performing the frequency pooling on each block independently.

Safranek & Johnston's perceptual image coder and metric uses the same strategy that Watson's DCT metric the only difference being that the channel decomposition uses a generalized quadrature mirror filter (GQMF) bank which splits the frequency spectrum into 16 uniform subbands.

Bradley's wavelet visible difference predictor (WVDP) [26] uses a model based on the wavelet transform which is intended for use with wavelet-based coders (e.g. JPEG2000). This model is a simplification of Daly's VDP described above. The modifications include the use of a separable wavelet transform instead of the Cortex Transform, the application of a wavelet contrast sensitivity function (CSF), and a simplified definition of subband contrast that allows one to predict the noise visibility directly from the wavelet coefficients.

The HVS-based methods described in this section follow the same fundamental design philosophy. They approach the image quality assessment problem from a "bottom-up" viewpoint: they simulate the functionality of each relevant component in the HVS and combine in a way that mimics the stages in the HVS from the eye to the brain. However, these methods are plagued by the complexity of the HVS models making them impractical for inclusion in real-time image processing systems. Uncertainties about the actual processing of visual information in the human brain (particularly at higher levels of the brain) complicate the design of HVS models. Furthermore, they often require robust and accurate calibration for specific viewing conditions. These drawbacks have led some researchers to believe that the HVS-based framework might not be the best way to approach the image quality problem [1].

2.4 Image Quality Assessment Based on Arbitrary Image Fidelity Criteria

Recently, researchers have explored novel approaches to the QA problem that are not based on models of the HVS. In contrast, they approach the image QA from the "top-down" where arbitrary signal criteria are used which assess quality in a manner that is quite different from that of the HVS. This approach is not concerned with accurately modeling the HVS provided that image quality is predicted accurately.

The Structural Similarity Index Metric

A recent paper [27] presents a new numerical image quality measure called the Structural Similarity Index Metric (SSIM). This state-of-the art metric has proved successful on large scale studies and will be used as a benchmark in this thesis. The fundamental paradigm underlying this method is that the human visual system is highly adapted to extract structural information (relative spatial covariance) from the viewing field and therefore, a measurement of structural information loss can provide a good approximation of the perceived image distortion. The structural information in an image is defined as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast. The metric separates the task of similarity measurement into three comparisons: luminance, contrast and structure.

Given two images (or images patches) $\mathbf{x}$ and $\mathbf{y}$ of size N to be compared, luminance is estimated as the mean of each image

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2.4}$$

contrast is estimated using the standard deviation as

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2, \tag{2.5}$$

and structure is estimated from the image vector $\mathbf{x}$ by removing the mean and normalizing by the standard deviation

$$\varsigma_x = \frac{\mathbf{x} - \mu_x}{\sigma_x}. \tag{2.6}$$

These measurements are combined using a luminance comparison $l(x, y)$, a contrast comparison function $c(x, y)$ and a structure comparison function $s(x, y)$ to give a composite measure of structural similarity:

$$\text{SSIM} = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{2.7}$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components. The SSIM values range from 0 to 1, where zero corresponds to a loss of all structural similarity and one corresponds to having an exact copy of the original image.

The comparison function are given as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2.8}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{2.9}$$

$$s(x,y) = \frac{\langle \varsigma_x, \varsigma_y \rangle + C_3}{\sigma_x\sigma_y + C_3} = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{2.10}$$

where $\langle\rangle$ is the inner-product operator defining the correlation between the structure of the two images and the constants C_1, C_2 and C_3 are non-negative constants included to avoid instability when $\mu_x^2 + \mu_y^2$, $\sigma_x^2 + \sigma_y^2$ and $\sigma_x\sigma_y$ respectively are very close to zero.

In this thesis, we follow the guidelines given in [27] by setting $\alpha = \beta = \gamma = 1$ and $C_3 = C_2/2$ yielding:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.11}$$

2.5 Summary

In the last three decades, the dominant paradigm in full-reference image quality assessment has been the "bottom-up" approach which bases itself on modeling the human visual system: metrics based on this paradigm simulate the functionality of each relevant component of the HVS and use these components as basic buildings blocks to assess image quality by combine them together in a pipeline structure. Recent state-of-the art metrics, such as the SSIM, follow a "top-down" approach by basing themselves on the hypothesized overall functionality of the entire HVS and treat it as a black-box system where only the input-output relationship is of concern.

It is interesting to note that the boundary between the "bottom-up" and "top-down" categories is somewhat blurred because many QA algorithms contain elements from both categories. For instance, even though the SSIM does not make use of explicit models or measurements of HVS sensitivities, it implicitly accounts for important HVS properties such as light adaption and masking, in addition to the perception of image structure.

The image QA metrics based on the HVS, particularly those that aim to be general-purpose metrics (Daly's VDP [20], Lubin [21, 22], Teo and Heeger [23]) are so elaborate that they tend to be difficult to implement, computationally intensive and difficult to match to a given set of conditions. On the other hand, the effectiveness of "top-down" methods such as the SSIM highly depends of the validity of the hypotheses they are based upon and the standard MSE, despite its appealing simplicity, does not perform well as an image QA metric for the reasons stated in Section 2.2. Our goal in this thesis is to design simple, practical and computationally efficient metrics. We choose to focus on image compression applications. We argue that a simple modification of the MSE by attaching weights (derived from prior knowledge about the image compression parameters) to the image samples in a frequency domain will achieve our goal. The metrics developed in this thesis are influenced by the particular application they are developed for but it will be shown that they have general applicability.

Chapter 3

Full Reference Image Quality Assessment Using Frequency Domain Transforms

Image QA can be viewed from a purely general point of view or it can be analyzed in the context of specific tasks. FR image quality assessment assumes that the undistorted reference image is fully available. In practical applications, FR metrics are used in off-line applications. They are typically used either for optimization purposes during the design stage of image processing systems or for comparative analyzes between different image processing systems and algorithms (i.e. to determine which of them provides the best quality results). In such frameworks image quality assessment is the measure of degradation when an image is distorted from processing. Therefore, quality metrics need not necessarily rely on sophisticated general models of the HVS but on a priori knowledge about the image processing system under consideration.

Of particular interest is the design and evaluation of image compression schemes. Most image compression schemes consist of three closely connected components namely transformation, quantization and and encoding as shown in Fig 3.1: compression is accomplished by applying a linear transform to decorrelate the image data, quantizing the resulting transform coefficients, and entropy coding the quantized values.

To facilitate the exploitation of psychovisual redundancies, the pictures are transformed to a domain where different frequency ranges with varying sensitivities of the human visual

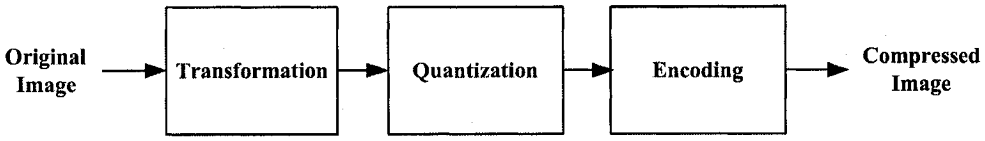


Fig. 3.1 A typical image compression system

system can be separated. After the transformation, the numerical precision of the transformed data is reduced in order to decrease the number of bits in the stream. The degree of quantization applied to each coefficient is usually determined by the visibility of the resulting distortion to a human observer. Quantization is the stage that is responsible for quality degradation. After the data has been quantized into a finite set of values, it is be encoded by exploiting the redundancy between the quantized coefficients in the bitstream. Entropy coding, which relies on the fact that certain symbols occur much more frequently than others, is often used for this process.

3.1 Description of the Proposed Quality Assessment Method

The proposed QA method is composed of three stages:

- Color space conversion
- Frequency domain transformation
- Error weighting and pooling

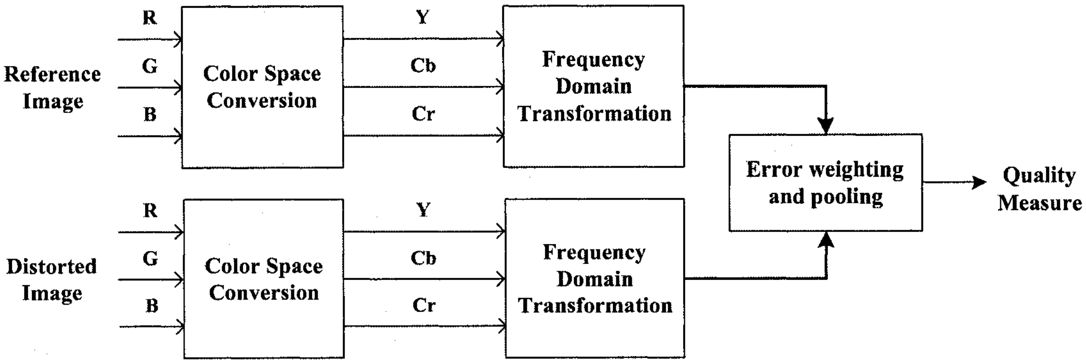


Fig. 3.2 Block diagram of the proposed image QA method

3.1.1 Color space conversion

In the first stage, the reference and distorted images are converted to a luminance/chrominance color space. The base images are in RGB (Red, Green, Blue) format which stores each color's value ([0,255] range) for each pixel. Most colors in the visible spectrum can be recreated by a combination of the RGB components. As mentioned in Section 2.3 the RGB color space is not the best choice for digital image processing tasks since the red, green and blue components are highly correlated. Therefore, the input images are converted the the YCbCr color space. The conversion is done through a linear transformation of the RGB components, as shown in Eq. 3.1, which produces a luminance signal (Y) and a pair of chrominance signals $(Cb$ and $Cr)$.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \tag{3.1}$$

The luminance (Y) signal conveys color brightness levels and provides the grayscale version of the images, and each chrominance signal gives the difference between a color and a reference white at the same luminance: Cb encodes the difference between the blue primary and luminance, and Cr the difference between the red primary and luminance. Only the luminance layer is considered for QA purposes. This choice is based on evidence that luminance is more important than chrominance from a perceptual standpoint: the accuracy of the brightness information of the luminance channel has far more impact on the image discerned than that of the other two. This fact is used in many image compression systems where chrominance data is assessed and processed at a lower resolutions than luminance data.

3.1.2 Frequency domain transformation

In the second stage, the luminance images are transformed using the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT). The DCT and DWT are orthogonal transforms may be represented by Eq. 3.2:

$$X = UAV^T \tag{3.2}$$

where A is the input image ($M \times N$), U and V are transition matrices ($M \times M$ and $N \times N$ respectively) and X is the transformed data ($M \times N$).

The Discrete Cosine Transform

For the DCT, the orthogonal transition matrix U is defined Eq. 3.3:

$$U(i,j) = \begin{cases} \sqrt{\frac{1}{M}} & i = 1, \\ \sqrt{\frac{2}{M}} \cos\left[\frac{\pi(2j-1)(i-1)}{2M}\right] & 2 \leq i \leq M. \end{cases} \tag{3.3}$$

The orthogonal matrix V is defined similarly by replacing M with N in Eq. 3.3.

The Discrete Wavelet Transform

Wavelet transforms decompose signals through the use of scaling functions and wavelet functions. The one-dimensional scaling function $\varphi(x)$ is the solution of the following two-scale equation:

$$\varphi(x) = \sum_{n=0}^{L} h_n \sqrt{2} \varphi(2x - n) \qquad x \in \Re \tag{3.4}$$

where $\{h_n\}$ is a finite sequence of real numbers (scaling coefficients) and $L < M$. There exists various types of wavelets transforms (e.g. Haar wavelet transform, Daubechies wavelet transform) each with well defined coefficient sequences $\{h_n\}$. The one-dimensional wavelet function $\psi(x)$ is given by the two-scale expression:

$$\psi(x) = \sum_{n=0}^{L} g_n \sqrt{2} \psi(2x - n) \qquad g_n = (-1)^n h_{L-n} \tag{3.5}$$

In two-dimensional wavelet analysis, one uses a scaling function $\varphi(x)\varphi(y)$ and three two-dimensional wavelets functions $\psi(x)\varphi(y)$, $\varphi(x)\psi(y)$ and $\psi(x)\psi(y)$. The orthogonal transition matrices U and V are defined by the following procedure:

1. Each row of the upper $M/2 \times M$ part of U consists of the sequence $\{h_n\}$. The first row is $h_0, h_1, \ldots, h_L, 0, \ldots$; the second row is the first row shifted to the right by two places i.e. $0, 0, h_0, h_1, \ldots, h_L, 0, \ldots$; the third row is the first row shifted to the right

by four places, etc. When h_L reaches the last column, the portion overflowing into the right end moves to the left end periodically

2. Each row of the lower $M/2 \times M$ part of U consists of the sequence $\{g_n\}$. As for the upper part, each row is the double right shift of the previous row.
3. The orthogonal matrix V is constructed in the same way by replacing M with N.

The DCT and DWT transform the images from the spatial (pixel) domain to the frequency domain using a transformation Each M x N image is transformed to an M x N matrix populated with coefficients that describe the horizontal and vertical spatial frequency characteristics of the image. The images are separated into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). These transforms are analogous to the the multiple channel models of the HVS and are used extensively in image compression schemes. In this thesis, the coefficients are grouped in 4 quadrants: low frequency (LL) coefficients which are clustered in the left top corner, mid frequency coefficients in the top right (HL) and bottom left (LH) corners and high frequency (HH) coefficients in the bottom right corner as shown in Fig. 3.3.

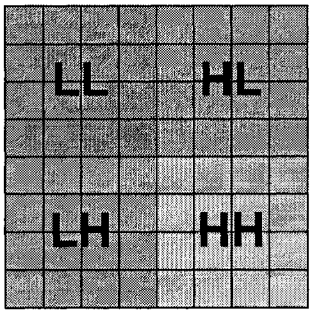


Fig. 3.3 Frequency subbands of transform coefficients

3.1.3 Error weighting and pooling

In the third stage, the MSE between the transform coefficients of the reference and distorted images are computed for each quadrant (LL, HL, LH and HH). The errors in each quadrant are pooled using a weighted mean as shown in Eq. 3.6 below.

$$\mathrm{Q} = (w_{\mathrm{LL}}\ \mathrm{MSE_{LL}} + w_{\mathrm{HL}}\ \mathrm{MSE_{HL}} + w_{\mathrm{LH}}\ \mathrm{MSE_{LH}} + w_{\mathrm{HH}}\ \mathrm{MSE_{HH}})^{1/2} \tag{3.6}$$

The numerical number is labeled $\mathrm{Q_{DCT}}$ or $\mathrm{Q_{DWT}}$ depending on the linear transform that was used in the second stage.

There is evidence that the human eye is much more sensitive to errors in low frequencies (which contain the basic image structural information) than in higher frequencies (which correspond to image details and noise) [14]. This fact is exploited extensively for compression purposes: in the quantization stage, compression algorithms quantize high frequency coefficients more coarsely (higher quantization factors q) than low frequency coefficients as illustrated in Fig. 3.4(a) which shows the standard quantization matrix used in JPEG compression. Since quantization is the stage that is responsible for quality degradation, quantization steps can be used to derive a weighing assignment as shown in Eq. 3.7 below.

$$w_{ij} = [q_{ij}(q_{\mathrm{LL}}^{-1} + q_{\mathrm{HL}}^{-1} + q_{\mathrm{LH}}^{-1} + q_{\mathrm{HH}}^{-1})]^{-1} \tag{3.7}$$

where $\{ij\} = \{\mathrm{LL,HL,LH,HH}\}$ and $\sum w_{ij} = 1$. q_{LL}, q_{HL}, q_{LH} and q_{HH} are the average quantization steps for each quadrant. As an example, if the quantization matrix under consideration is the one in Fig. 3.4(a) then q_{LL} is the average of the 16 quantizations steps in the top-left corner i.e. $q_{\mathrm{LL}} = 16.1875$.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

(a) Luminance quantization matrix

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

(b) Chrominance quantization matrix

Fig. 3.4 Standard luminance and chrominance quantization matrices for JPEG

3.2 Experimental Results

In this section we present results on validation of the proposed image quality metrics on the LIVE database [12] and comparisons with other quality assessment algorithms namely the PSNR and the SSIM [27].

3.2.1 LIVE database details

Our experiments are conducted on the LIVE data set [12] which contains a total of 982 images distorted using five different distortions types: JPEG2000 compression, JPEG compression, white Gaussian noise, Gaussian blur and bit errors in a JPEG2000 bitstream transmitted over a fast-fading Rayleigh channel as shown in Table 3.1 (details of the distortion parameters may be found in the LIVE database documentation [12]). Each image in the database is associated with the corresponding Difference Mean Opinion Score (DMOS) which is the average subjective rating given to the image by human subjects (the average number of subjects used to evaluate each image is 22.8 [13]). The DMOS is expressed in a 0-100 scale, where 0 indicates imperceptible quality degradation. The average standard deviation between the individual ratings is $\sigma_{\text{DMOS}} = 6.524$.

Table 3.1 Details of the LIVE database

	Number of images
JPEG2000	227
JPEG	233
White Noise	174
Gaussian Blur	174
Fast-fading Rayleigh channel	174
Total	982

3.2.2 Simulation parameters

The weights used in our experiments are given in Table 3.2. For Q_{DCT}, these weights are computed from the quantization factors of the standard DCT luminance quantization matrix used for JPEG compression: the quantization step for each quadrant (LL, HL, LH and HH) is set as the average of the quantizer steps in the respective DCT region. The weights used for Q_{DWT} are computed from the wavelet quantization factors described in [28].

Table 3.2 Weighting assignments used in experiments (luminance layer Y)

DCT	q_{ij}	w_{ij}	DWT	q_{ij}	w_{ij}
LL	16.1875	0.5779	LL	14.049	0.4066
HL	54.8125	0.1707	HL	23.028	0.2481
LH	59.1250	0.1582	LH	23.028	0.2481
HH	100.3750	0.0932	HH	58.756	0.0972

The SSIM parameters C_1 and C_2 in Eq. 2.11 are set to 6.5025 and 58.5255 respectively as suggested in [27].

The type of wavelet transform used in the simulation is the Cohen-Daubechies-Favreau (CDF) 9/7 wavelet transform. This wavelet is an especially effective biorthogonal wavelet and is used in the JPEG2000 standard.

3.2.3 Performance metrics and calibration of objective scores

As discussed in 1.3.3, the performance of objective image quality metrics is typically evaluated with respect to three attributes namely prediction accuracy, prediction monotonicity and prediction consistency. These attributes are evaluated through five performance measures which are specified by the Video Quality Experts Groups [29] and listed in Table 3.3 below.

Table 3.3 Performance measures

Prediction accuracy		
	Root Mean Squared Error	Eq. 3.10
	Mean Absolute Error	Eq. 3.11
	Pearson correlation coefficient	Eq. 3.12
Prediction monotonicity		
	Spearman rank order correlation coefficient	Eq. 3.13
Prediction accuracy		
	Outlier ratio	Eq. 3.15

The proposed and comparison metrics (Q_{DCT}, Q_{DWT} and PSNR, SSIM respectively) are applied to each image in the data set. It is generally acceptable for a QA metric to stably predict subjective quality within a non-linear mapping. Since the mapping can be easily compensated for and is likely to depend upon the subjective validation/application scope and methodology, it is best to leave it to the final application, and not to make it part of the QA algorithm. For each metric, the resulting objective quality scores denoted x_i (where $i = 1, ..., N$ and N is the size of the data set) are mapped to a set of predicted DMOS ($\mathrm{DMOS_p}$) denoted $p_i = g(x_i)$ (where $i = 1, ..., N$ and $g(x)$ is the function used for mapping). This mapping is done to facilitate comparison between the subjective ratings and objective scores in a common analysis space. The function $g(x)$ is typically a non-linear logistic function. Non-linear mapping is chosen over linear mapping to account for the non-linear characteristics of subjective scores at the extremes of test ranges: at the extremes of the scale (corresponding to very high and very low quality) the distribution of subjective scores tends to be quite skewed. In this thesis, a five-parameter logistic function with additive linear term is used as shown in Eq. 3.8 below.

$$p_i = g(x_i) = \beta_1 \mathrm{logistic}(\beta_2, (x_i - \beta_3)) + \beta_4 x + \beta_5 \tag{3.8}$$

where

$$\mathrm{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(x\tau)} \tag{3.9}$$

The five parameters β_1, β_2, β_3, β_4, β_5 are chosen in such a way that $p_i = g(x_i)$ better fits the experimental data. The fitting was done using MATLAB's *fminsearch* function using all of the experimental data.

The first performance metric is the RMSE between the DMOS and the predicted subjective subjective scores ($\mathrm{DMOS_p}$). The RSME is also know as the standard error of estimation and is representative of the size of a "typical" error.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - p_i)^2} \tag{3.10}$$

where s_i is the DMOS of the i-th image in the data set.

The second performance metric is the MAE between DMOS and $\mathrm{DMOS_p}$:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|s_i - p_i| \tag{3.11}$$

The Pearson correlation coefficient (CC), is used to measure the association between DMOS and $\mathrm{DMOS_p}$. It measures prediction accuracy by characterizing the degree of scattering of data pairs (s_i,p_i) around a linear function.

$$\mathrm{CC} = \frac{\sum_{i=1}^{N}[p_i - \overline{p}][s_i - \overline{s}]}{\sqrt{\sum_{i=1}^{N}[p_i - \overline{p}]^2}\sqrt{\sum_{i=1}^{N}[s_i - \overline{s}]^2}} \tag{3.12}$$

where $\overline{s}$ and $\overline{p}$ denote the mean of vectors $[s_1, ..., s_N]$ and $[p_1, ..., p_N]$ respectively.

The Spearman rank order correlation coefficient (SROCC) is a nonparametric correlation measure which describes prediction monotonicity by quantifying whether changes (increase or decrease) in DMOS are followed by changes (increase or decrease) in $\mathrm{DMOS_p}$. Ideally (SROCC = 1), the difference between a metrics rating of two images should always have the same sign as the differences between the corresponding subjective ratings.

$$\mathrm{SROCC} = \frac{\sum_{i=1}^{N}[P_i - \overline{P}][S_i - \overline{S}]}{\sqrt{\sum_{i=1}^{N}[P_i - \overline{P}]^2}\sqrt{\sum_{i=1}^{N}[S_i - \overline{S}]^2}} \tag{3.13}$$

where S_i and P_i denote the ranks of s_i and p_i respectively in the ordered data series and $\overline{S}$ and $\overline{P}$ are the midranks (average of ranks) of the respective data sets.

The outlier ratio (OR) measures prediction consistency by computing the number of outliers (N_o). An outlier is defined as a data point for which the absolute prediction error $|e_i| = |s_i - p_i|$ is greater than a certain threshold. We set the threshold at twice the standard deviation of the subjective ratings:

$$|e_i| = |s_i - p_i| > 2\sigma_{\mathrm{DMOS}} \tag{3.14}$$

The outlier ratio is then given by:

$$\mathrm{OR} = \frac{N_o}{N} \tag{3.15}$$

3.2.4 Results

Since the proposed metrics are primarily intended for evaluation of compressed images, they are first evaluated on the subset of the LIVE database containing JPEG2000 and JPEG images. The results in Tables 3.4 and 3.5 show that the Q_{DCT} and Q_{DWT} outperform the PSNR. When compared to the SSIM, both metrics provide comparable results. The Q_{DCT} is not strongly biased towards JPEG, and similarly the Q_{DWT} is not biased towards JPEG2000.

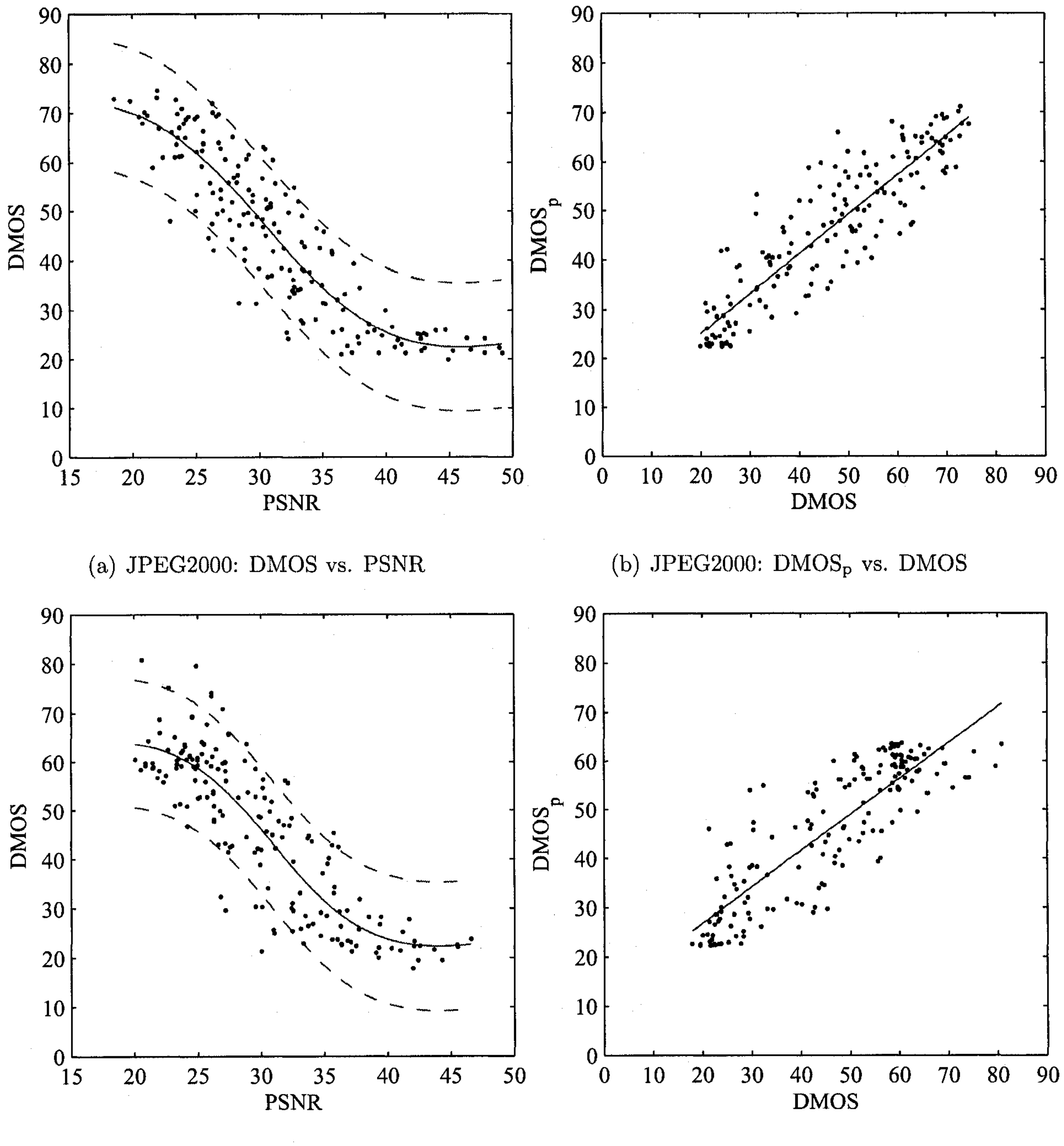
Graphical results in the form of scatter plots for the PSNR, SSIM, Q_{DCT} and Q_{DWT} are shown in Fig. 3.5 to 3.8 respectively. In each plot, each point represents one distorted image. The y-axis represents the subjective DMOS in a 0-100 scale (where 0 indicates imperceptible quality degradation). On the left side plots, the x-axis plots represents the quantitative measure by each method and the solid line represents the fitting with the logistic function in Eq. 3.8. The dashed line represents the outlier point limit. On the right side plots, the x-axis plots represents the predicted DMOS on a 0-100 scale (same scale as the subjective DMOS). The solid line represents the fitting with a linear function; the CC is a measure of the degree of scattering of the points around that linear function.

Table 3.4 Prediction performance of PSNR, SSIM, Q_{DCT} and Q_{DWT} on JPEG2000 compressed images

Model	RMSE	MAE	CC	SROCC	OR
PSNR	7.1805	5.5313	0.8964	0.8894	0.0711
SSIM	6.0268	4.5145	0.9690	0.9710	0.0529
Q_{DCT}	6.2655	4.7311	0.9665	0.9613	0.0396
Q_{DWT}	6.3344	4.7986	0.9657	0.9614	0.0529

Table 3.5 Prediction performance of PSNR, SSIM, Q_{DCT} and Q_{DWT} on JPEG compressed images

Model	RMSE	MAE	CC	SROCC	OR
PSNR	8.1720	6.3658	0.8595	0.8413	0.1086
SSIM	6.1691	4.5791	0.9671	0.9576	0.0429
Q_{DCT}	6.4021	4.6235	0.9646	0.9462	0.0601
Q_{DWT}	6.7855	4.9024	0.9601	0.9415	0.0730



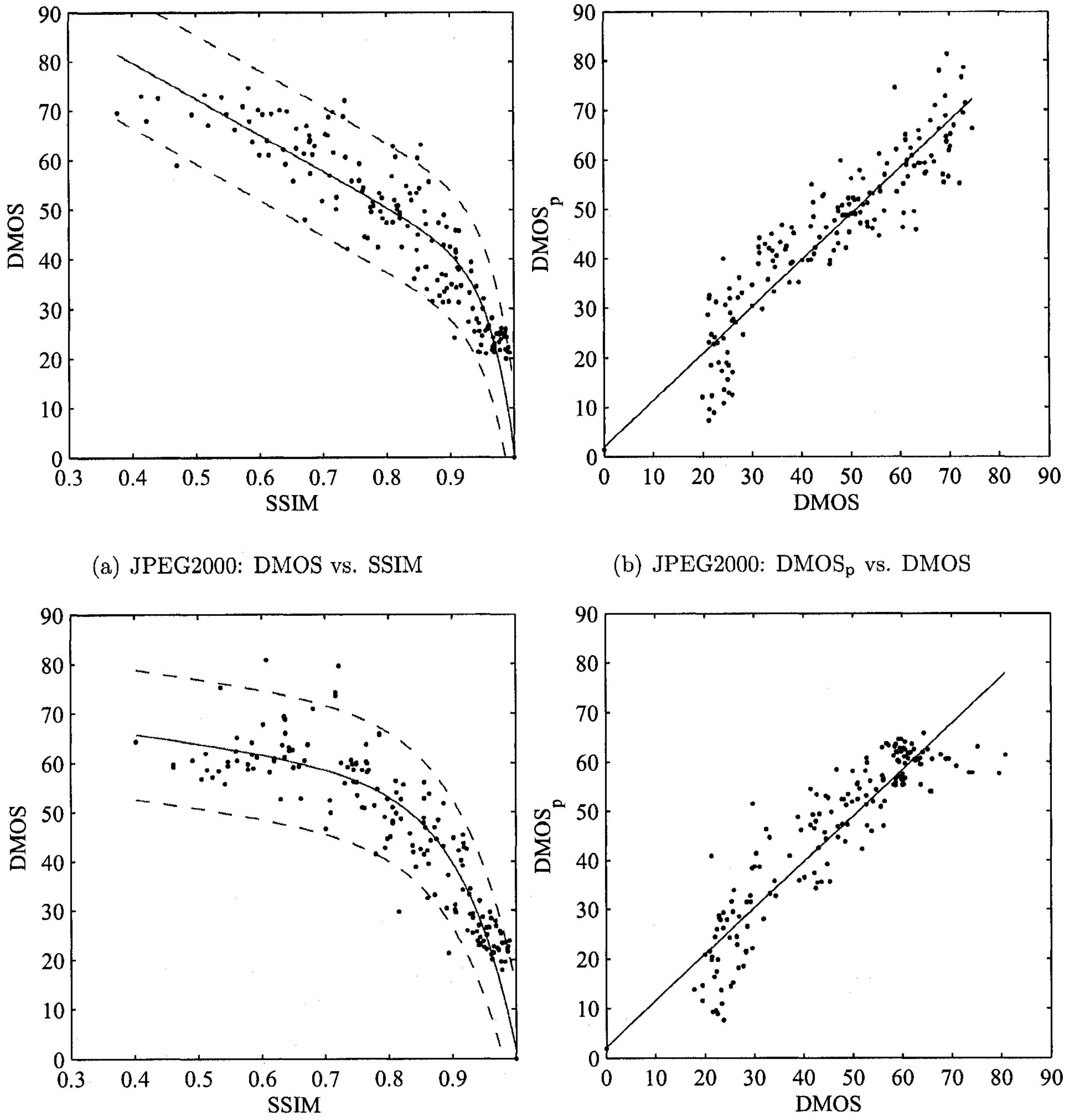
(a) JPEG2000: DMOS vs. PSNR

(b) JPEG2000: $\mathrm{DMOS_p}$ vs. DMOS

(c) JPEG: DMOS vs. PSNR

(d) JPEG: $\mathrm{DMOS_p}$ vs. DMOS

Fig. 3.5 Scatter plots for the quality prediction of JPEG2000 and JPEG compressed images by PSNR



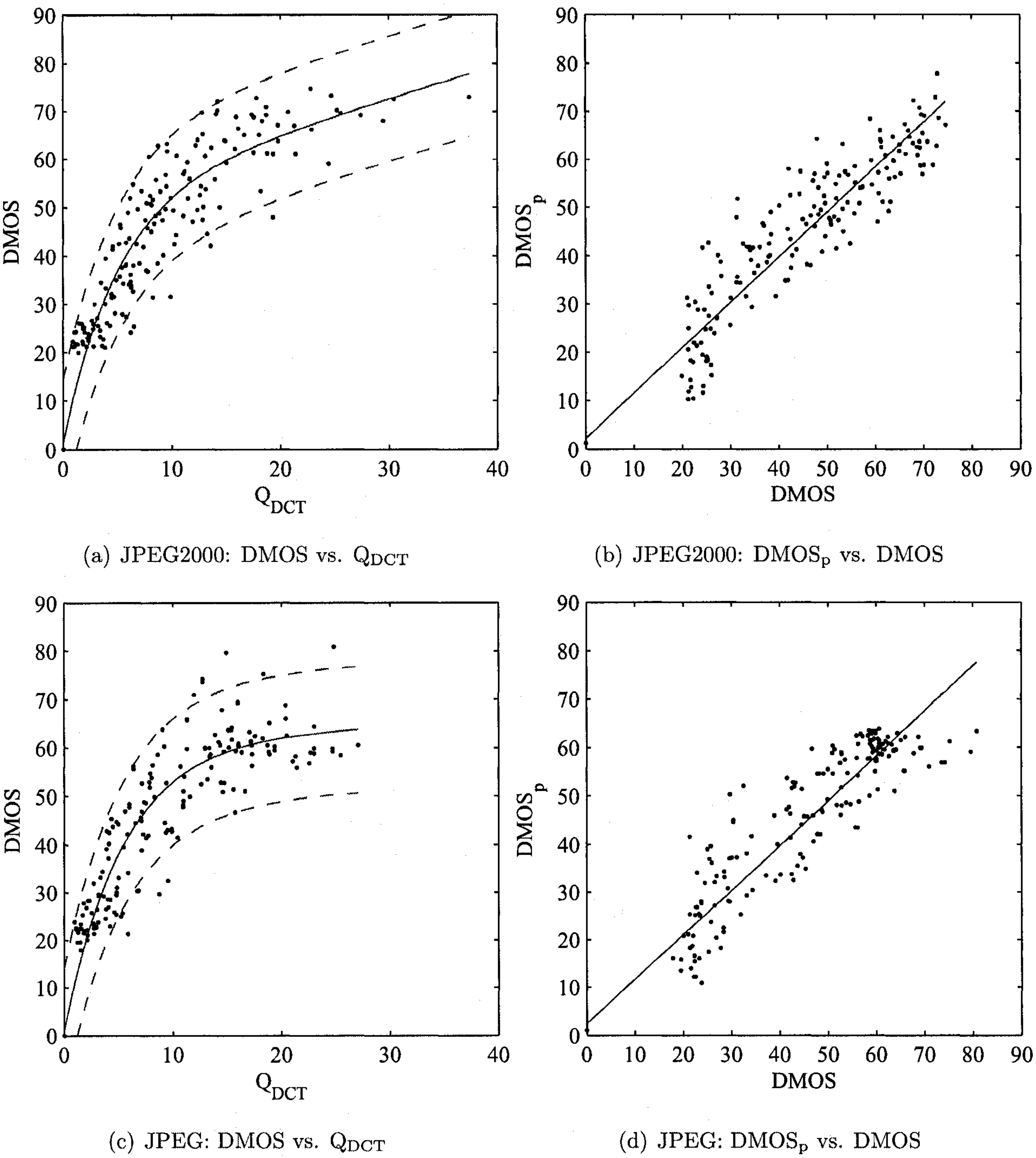
(a) JPEG2000: DMOS vs. SSIM

(b) JPEG2000: DMOS$_p$ vs. DMOS

(c) JPEG: DMOS vs. SSIM

(d) JPEG: DMOS$_p$ vs. DMOS

Fig. 3.6 Scatter plots for the quality prediction of JPEG2000 and JPEG compressed images by SSIM



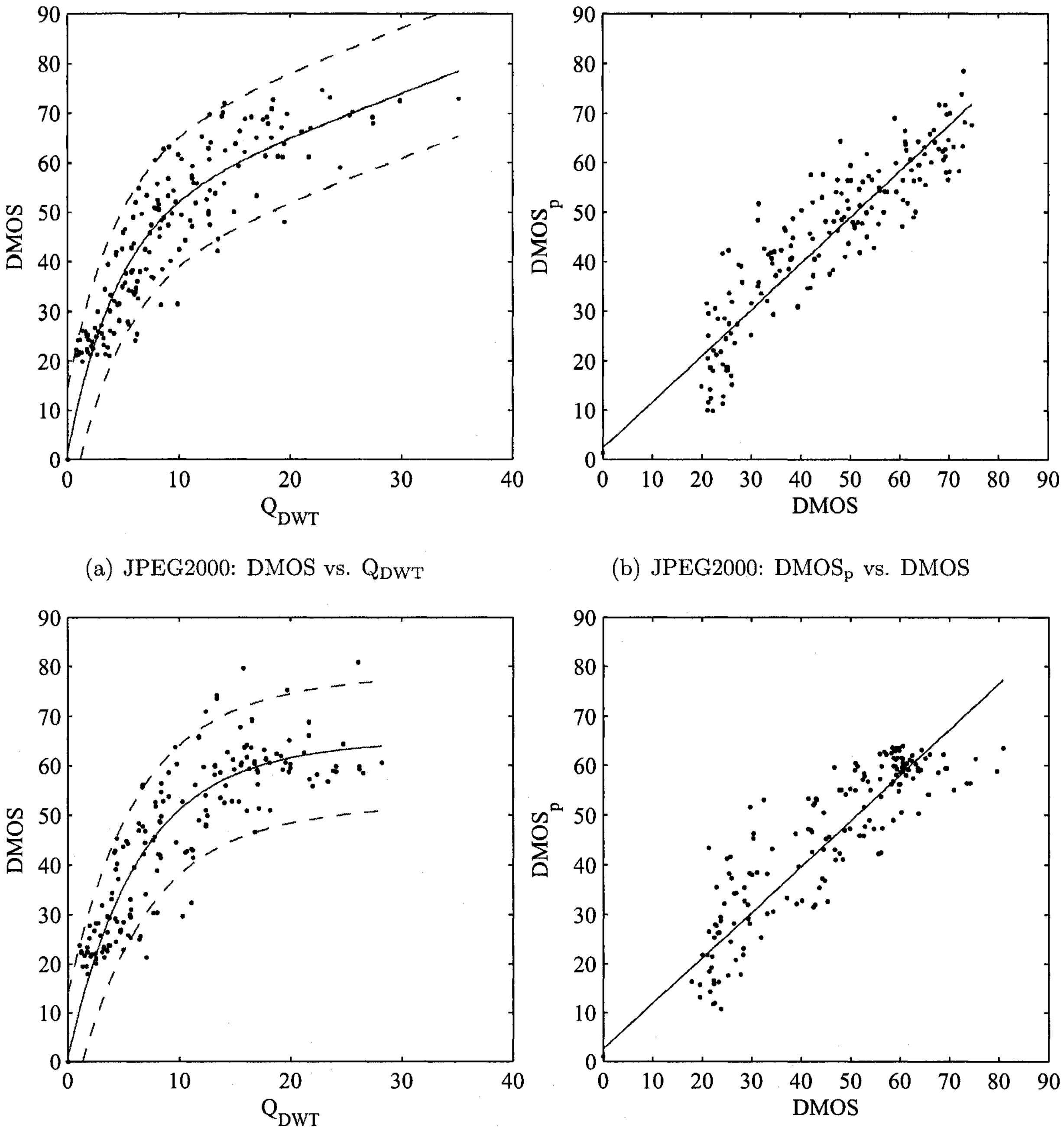
(a) JPEG2000: DMOS vs. Q_{DCT}

(b) JPEG2000: DMOS$_p$ vs. DMOS

(c) JPEG: DMOS vs. Q_{DCT}

(d) JPEG: DMOS$_p$ vs. DMOS

Fig. 3.7 Scatter plots for the quality prediction of JPEG2000 and JPEG compressed images by Q_{DCT}



(a) JPEG2000: DMOS vs. Q_{DWT}

(b) JPEG2000: $DMOS_p$ vs. DMOS

(c) JPEG: DMOS vs. Q_{DWT}

(d) JPEG: $DMOS_p$ vs. DMOS

Fig. 3.8 Scatter plots for the quality prediction of JPEG2000 and JPEG compressed images by Q_{DWT}

To investigate the performance of Q_{DCT} and Q_{DWT} as general-purpose QA metrics, they were tested on the complete LIVE database. The performance of the proposed and comparison metrics on the entire LIVE data set is summarized in Table 3.6 and illustrated in Fig. 3.9 to 3.12. Both proposed metrics clearly outperform both the PSNR and the SSIM.

Prediction performance results pertaining to specific distortion types are given in Tables 3.7 to 3.11. The Q_{DCT} and Q_{DWT} outperform the PSNR not only in overall performance but also within each distortion type. When compared to the SSIM, both metrics provide comparable results.

Table 3.6 Prediction performance of PSNR, SSIM, Q_{DCT} and Q_{DWT} on LIVE images

Model	RMSE	MAE	CC	SROCC	OR
PSNR	9.0868	7.2725	0.8256	0.8197	0.1566
SSIM	8.0485	6.2793	0.9374	0.9240	0.1202
Q_{DCT}	7.9504	6.0169	0.9390	0.9215	0.1018
Q_{DWT}	7.9938	6.0898	0.9383	0.9208	0.1090

Table 3.7 Prediction accuracy (RMSE) of PSNR, SSIM, Q_{DCT} and Q_{DWT} for various distortion types

Distortion	PSNR	SSIM	Q_{DCT}	Q_{DWT}
JPEG2000	7.1805	6.0268	6.2655	6.3344
JPEG	8.1720	6.1691	6.4021	6.7855
White Noise	2.4692	3.6231	2.8663	2.6215
Gaussian Blur	9.7431	7.8512	7.8466	8.3783
FF	7.5123	5.6529	6.8956	6.8185

Table 3.8 Prediction accuracy (MAE) of PSNR, SSIM, Q_{DCT} and Q_{DWT} for various distortion types

Distortion	PSNR	SSIM	Q_{DCT}	Q_{DWT}
JPEG2000	5.5313	4.5145	4.7311	4.7986
JPEG	6.3658	4.5791	4.6235	4.9024
White Noise	1.9442	2.6881	2.1548	1.9866
Gaussian Blur	7.6642	5.6486	5.7831	6.1252
FF	5.7660	4.1517	5.0446	5.0484

Table 3.9 Prediction accuracy (CC) of PSNR, SSIM, Q_{DCT} and Q_{DWT} for various distortion types

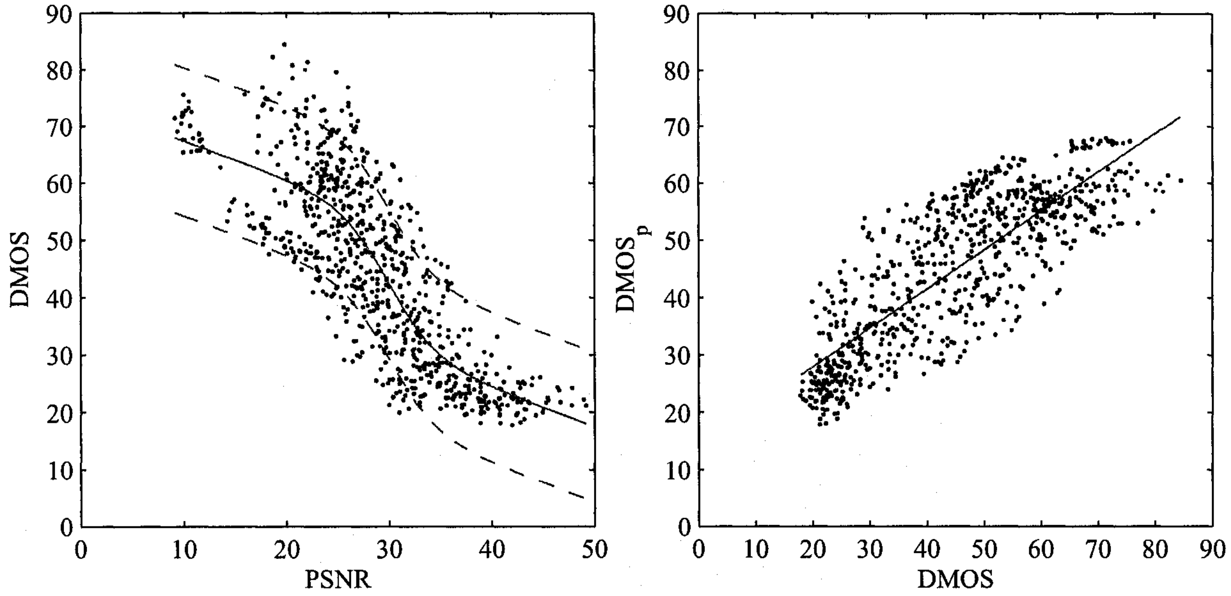
Distortion	PSNR	SSIM	Q_{DCT}	Q_{DWT}
JPEG2000	0.8964	0.9690	0.9665	0.9657
JPEG	0.8595	0.9671	0.9646	0.9601
White Noise	0.9880	0.9863	0.9915	0.9929
Gaussian Blur	0.7849	0.9326	0.9327	0.9228
FF	0.8896	0.9667	0.9500	0.9512

Table 3.10 Prediction monotonicity (SROCC) of PSNR, SSIM, Q_{DCT} and Q_{DWT} for various distortion types

Distortion	PSNR	SSIM	Q_{DCT}	Q_{DWT}
JPEG2000	0.8894	0.9710	0.9613	0.9614
JPEG	0.8413	0.9576	0.9462	0.9415
White Noise	0.9853	0.9817	0.9891	0.9907
Gaussian Blur	0.7816	0.9320	0.9134	0.8937
FF	0.8902	0.9639	0.9427	0.9454

Table 3.11 Prediction consistency (OR) of PSNR, SSIM, Q_{DCT} and Q_{DWT} for various distortion types

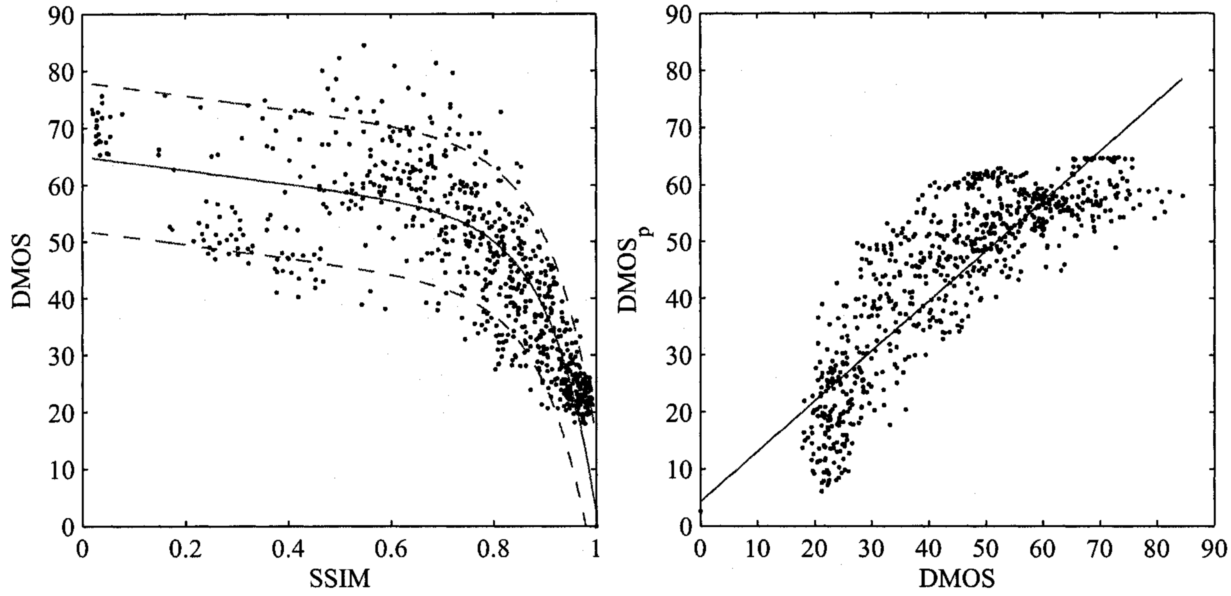
Distortion	PSNR	SSIM	Q_{DCT}	Q_{DWT}
JPEG2000	0.0711	0.0529	0.0396	0.0529
JPEG	0.1086	0.0429	0.0601	0.0730
White Noise	0	0	0	0
Gaussian Blur	0.1862	0.0977	0.1207	0.1437
FF	0.0965	0.0460	0.0805	0.0805



(a) DMOS vs. PSNR

(b) $DMOS_p$ vs. DMOS

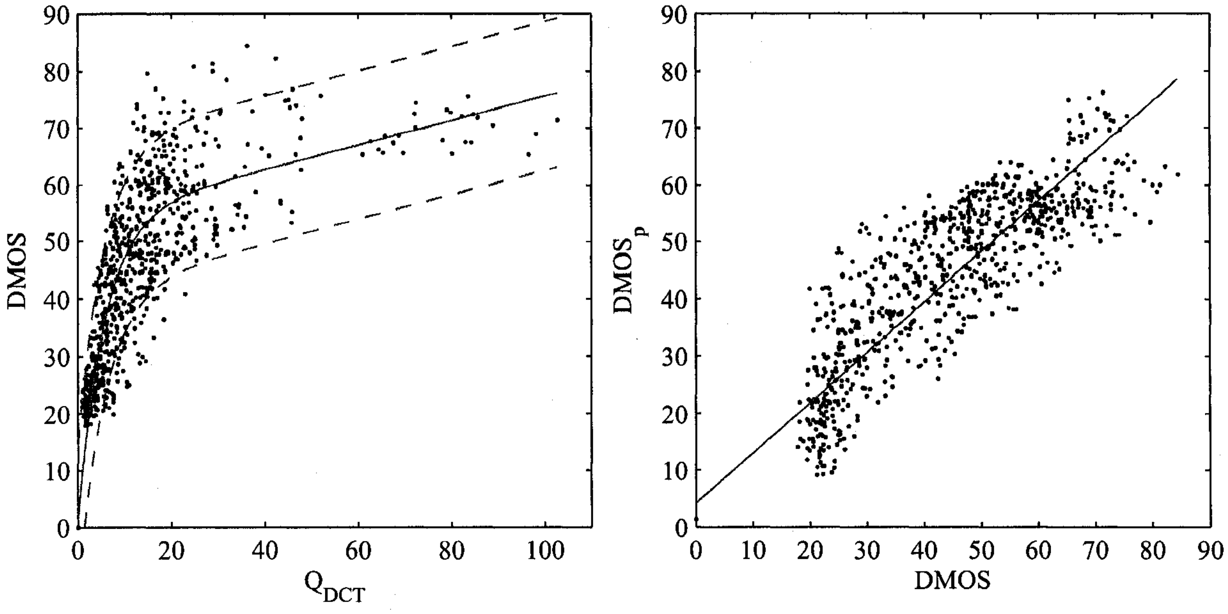
Fig. 3.9 Scatter plots for the quality prediction of LIVE images by PSNR



(a) DMOS vs. SSIM

(b) $DMOS_p$ vs. DMOS

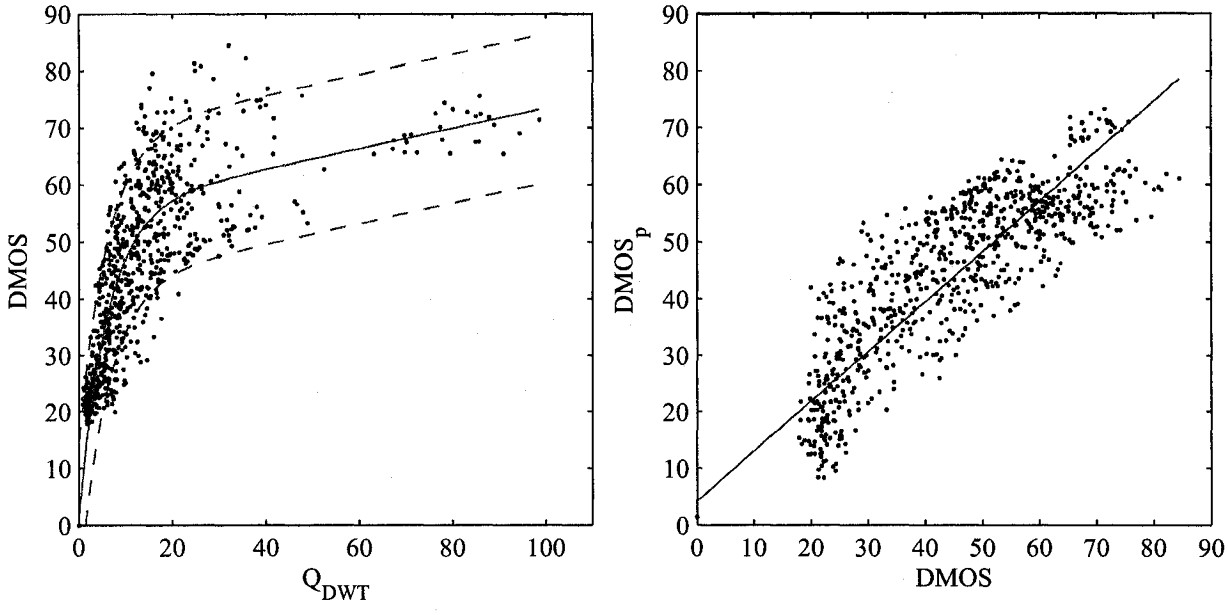
Fig. 3.10 Scatter plots for the quality prediction of LIVE images by SSIM



(a) DMOS vs. Q_{DCT}

(b) $DMOS_p$ vs. DMOS

Fig. 3.11 Scatter plots for the quality prediction of LIVE images by Q_{DCT}



(a) DMOS vs. Q_{DWT}

(b) $DMOS_p$ vs. DMOS

Fig. 3.12 Scatter plots for the quality prediction of LIVE images by Q_{DWT}

3.3 Use of Local Averaging to Improve Prediction Accuracy

Image distortions may vary across space. This is especially true in the case of block-based compression schemes such as JPEG and JPEG2000 which encode images using 8×8 blocks. Therefore it is safe to assume that quality estimation can be improved if local rather than global averaging procedures are used. Another substantial advantage of localized quality measurement is that it provides a varying quality map of each image.

Block-based versions of the $\mathrm{Q_{DCT}}$ and $\mathrm{Q_{DWT}}$ are incorporated as following:

1. The reference and distorted images are converted to a luminance/chrominance color space using Eq. 3.1.

2. Each luminance image is separated in $n \times n$ blocks.

3. Each block is transformed using the DCT or DWT.

4. Within each block, the MSE between the frequency coefficients of the reference and distorted images are computed for each quadrant (LL, HL, LH and HH). The errors in each quadrant are pooled using a weighted mean as shown in Eq. 3.16.

$$Q_k = (w_{\text{LL}}\ \mathrm{MSE_{LL}} + w_{\text{HL}}\ \mathrm{MSE_{HL}} + w_{\text{LH}}\ \mathrm{MSE_{LH}} + w_{\text{HH}}\ \mathrm{MSE_{HH}})^{1/2} \tag{3.16}$$

   where k denotes the block number; if the image dimensions are M x N, there is a total of $R_S = (M \times N)/n^2$ blocks. The weights are computed following the procedure described in Eq. 3.7 of Section 3.1

   The set of block quality values when displayed in a graph form a distortion map (which is of size M/n x N/n).

5. The overall quality value is defined as the average of the block quality values.

$$\mathrm{Q} = \frac{\sum_{k=1}^{R_S} Q_k}{R_S} \tag{3.17}$$

Fig. 3.13 shows the CC performance of $\mathrm{Q_{DCT}}$ and $\mathrm{Q_{DWT}}$ for various blocks sizes. Experiments on the complete LIVE data set indicate that the $\mathrm{Q_{DWT}}$ is relatively insensitive to variations in block sizes. When using the $\mathrm{Q}_{\mathrm{DWT}}$, the optimal block size is 8. However, using a block size of 8 is detrimental when evaluating images compressed using the DCT namely JPEG images. Baseline JPEG compression uses 8 x 8 blocks. This causes blocking artifacts which are visible at the edges of the blocks. If image quality is evaluated using 8 x 8 blocks, the distortion within each block is accurately assessed but there is no way to capture the distortion across block edges (blocking artifacts). This fact is reflected in the CC performance of the $\mathrm{Q}_{\mathrm{DCT}}$ where the CC drops for $n = 8$. One way to capture the blocking distortion in the case of JPEG images would be to change the type of window used: one could use an 8 x 8 moving window which moves pixel-by-pixel from the top-left corner to the bottom-right corner of the image (this however results in a distortion map that has the same size as the input images and this distortion map may in turn exhibits undesirable blocking artifacts); another type of window that can be used is a smooth window that is slightly larger than the JPEG block size (e.g. 11×11 Gaussian window). Another potential solution would be to combine the $\mathrm{Q}_{\mathrm{DWT}}$ with one of the several metric described in literature that evaluates blocking artifacts.

3.4 Addition of Chrominance Information to Improve Prediction Accuracy

The proposed metrics are extended to full color images by incorporating the two chrominance layers (Cb and Cr).

1. The reference and distorted images are converted to a luminance/chrominance color space using Eq. 3.1.

2. For each of the three layers (Y, Cb and Cr), a quality measure is computed using the procedure described in Eq. 3.6 of Section 3.1 thus yielding three quality measures Q_Y, Q_{Cb} and Q_{Cr}. The quadrant weighting factors (w_{LL}, w_{HL}, w_{LH} and w_{HH}) used to compute Q_Y are those given in Table 3.2. The weights used to compute Q_{Cb} and Q_{Cr} are given in Table 3.12: for the DCT-based metric, these weights are computed from the quantization factors of the standard DCT chrominance quantization matrix used for JPEG compression (Fig. 3.4(b), the quantization step for each quadrant (LL, HL, LH and HH) is set as the average of the quantizer steps in the respective DCT

region); the weights used for the DWT-based matric are computed from the wavelet quantization factors described in [28].

3. The quality measures for each channel are pooled to yield a single numerical value.

$$Q = w_Y\, Q_Y + w_{Cb}\, Q_{Cb} + w_{Cr}\, Q_{Cr} \tag{3.18}$$

Fig. 3.14 shows the CC performance of Q_{DCT} and Q_{DWT} for various weighting assignments (note that $w_{Cb} = w_{Cr}$ in all experiments). Generally speaking, adding the chrominance information shows no substantial improvement in the prediction accuracy of Q_{DCT} and Q_{DWT}. However, in the case of white noise the proposed metrics are most effective when only chrominance information is kept: this is due to the fact that luminance noise is simply a variation in the brightness of the pixels in the image while chrominance noise is a variation in the color and therefore is perceptually more detrimental than luminance noise. Note however that the improvement is of the order of 10^{-3}.

Table 3.12 Weighting assignments used in experiments (chrominance layer Cb)

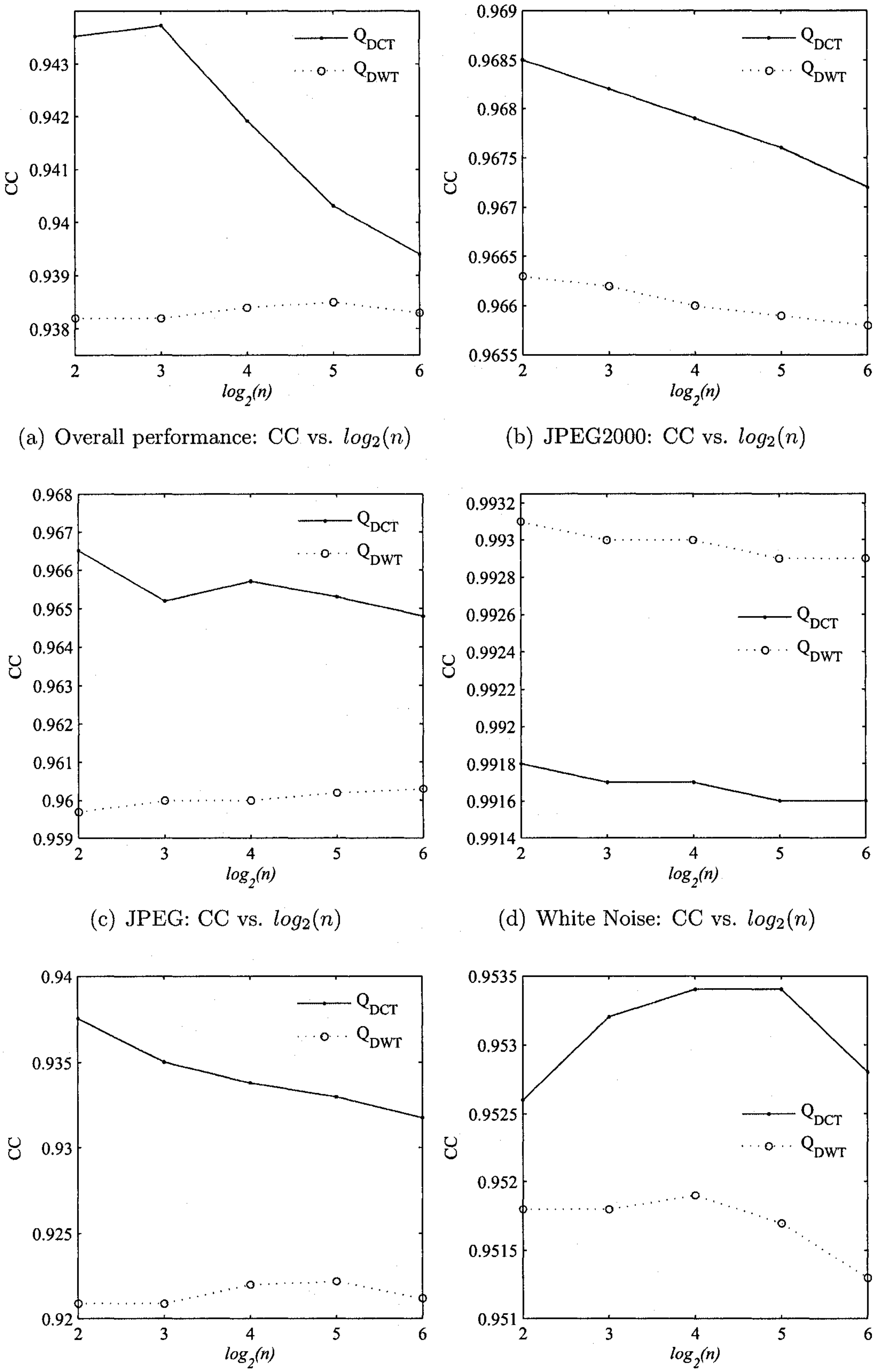
DCT	q_{ij}	w_{ij}
LL	11.766	0.7372
HL	99	0.0876
LH	99	0.0876
HH	99	0.0876

DWT	q_{ij}	w_{ij}
LL	55.249	0.3954
HL	86.789	0.2517
LH	86.789	0.2517
HH	215.84	0.1012

Table 3.13 Weighting assignments used in experiments (chrominance layer Cr)

DCT	q_{ij}	w_{ij}
LL	11.766	0.7372
HL	99	0.0876
LH	99	0.0876
HH	99	0.0876

DWT	q_{ij}	w_{ij}
LL	25.044	0.5076
HL	60.019	0.2118
LH	60.019	0.2118
HH	184.64	0.0688



(a) Overall performance: CC vs. $log_2(n)$

(b) JPEG2000: CC vs. $log_2(n)$

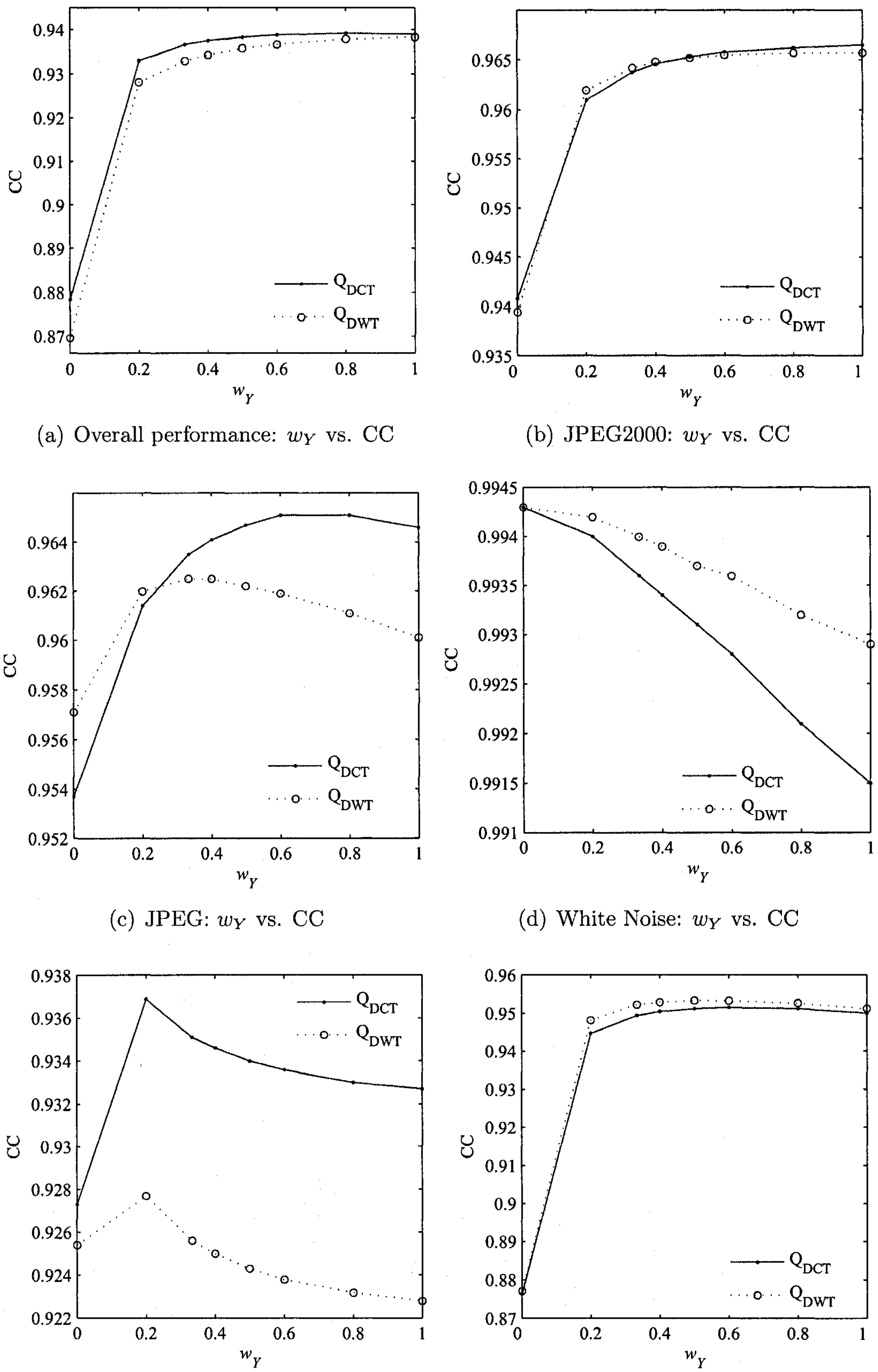
(c) JPEG: CC vs. $log_2(n)$

(d) White Noise: CC vs. $log_2(n)$

(e) Gaussian Blur: CC vs. $log_2(n)$

(f) JPEG2000 on FF channel: CC vs. $log_2(n)$

Fig. 3.13 Prediction accuracy (CC) of Q_{DCT} and Q_{DWT} as a function of



(a) Overall performance: w_Y vs. CC

(b) JPEG2000: w_Y vs. CC

(c) JPEG: w_Y vs. CC

(d) White Noise: w_Y vs. CC

(e) Gaussian Blur: w_Y vs. CC

(f) JPEG2000 on FF channel: w_Y vs. CC

Fig. 3.14 Prediction accuracy (CC) of Q_{DCT} and Q_{DWT} as a function of w_Y

3.5 Internal signals generated by the proposed method

In this section, we take a brief look at some of the internal signals generated by the proposed method namely the MSE in each quadrant. The average MSE for each quadrant (after transformation to the frequency domain via the DCT or DWT) is given in Table 3.14. These numbers were obtained by averaging the values obtained for each of the 982 test images in the LIVE database.

Table 3.14 MSE in each quadrant for the luminance and chrominance images

DCT	*Y*	*Cb*	*Cr*
$\mathrm{MSE}_{\mathrm{LL}}$	475.19	163.90	176.91
$\mathrm{MSE}_{\mathrm{HL}}$	177.64	139.56	157.65
$\mathrm{MSE}_{\mathrm{LH}}$	172.35	139.30	157.13
$\mathrm{MSE}_{\mathrm{HH}}$	153.45	140.60	158.51

DWT	*Y*	*Cb*	*Cr*
$\mathrm{MSE}_{\mathrm{LL}}$	575.78	200.94	214.68
$\mathrm{MSE}_{\mathrm{HL}}$	191.37	144.33	162.70
$\mathrm{MSE}_{\mathrm{LH}}$	196.45	144.31	161.98
$\mathrm{MSE}_{\mathrm{HH}}$	153.90	136.37	153.67

As expected, in the luminance layer the largest errors are in the low-frequency quadrant (LL) and the smallest ones in the high-frequency (HH) quadrant. This is further justification of our weighting choices in Table 3.2 (biggest weight for LL quadrant, smallest weight for the HH quadrant). Another interesting observation is the fact that the magnitude of the MSE in chrominance layers is in the same range as the MSE in the higher frequency (LH, HL and HH) quadrants of the luminance layer: since these errors are small (compared to the MSE in the LL quadrant of the luminance layer), we can expect them to have less impact on visual quality and this fact was confirmed by our experimental results in Section 3.4 which showed that addition of chrominance information has no substantial effect on the prediction accuracy of the proposed metrics.

3.6 Summary

This chapter presented two image quality metrics based on the DCT and DWT. Their development is driven by pre-determined applications, namely visual quality assessment of compressed images. The metrics are suitable for direct integration into image compression schemes as shown in Fig. 3.15 since they use the same linear transforms as modern compression schemes (e.g. JPEG and JPEG2000). Loss in quality is directly related to coefficient quantization errors. The metrics themselves turn out to be much more general and prove reliable over a wide range of image distortions making them suitable to use for other image processing applications. In the next chapter, we seek to adapt these metrics to perform reliably in reduced reference frameworks

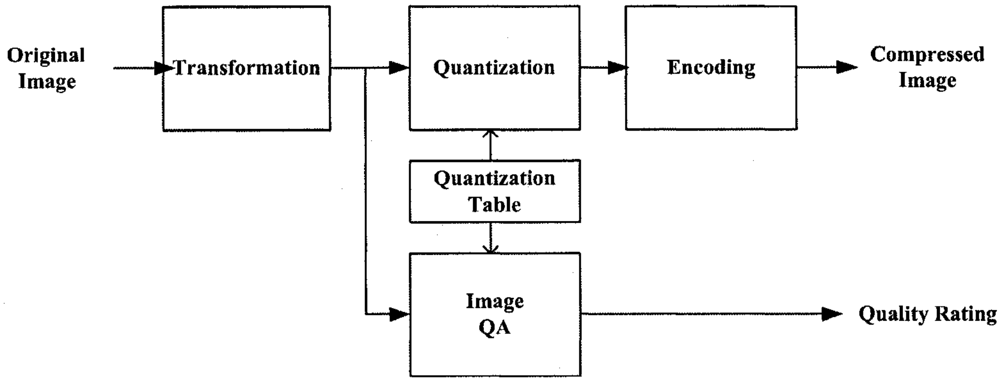


Fig. 3.15 Integration of the proposed metrics in compression systems

With regards to the "bottom-up" and "top-down" classification discussed in Section 2.5, the proposed algorithm contains elements from both categories. While it does not explicitly model each stage of the HVS, it implicitly accounts for important HVS properties. For instance the first lines of the quantization tables vary like the inverse of CSF function; therefore, the sensitivity of the human eye to spatial frequencies is implicitly taken into account.

Chapter 4

Towards Reduced Reference Image Quality Metrics

4.1 General Philosophy

So far, the proposed scheme assumes that the reference image is available in its entirety (FR framework). However, in many applications (e.g. multimedia communication networks), the reference image data is not available in its entirety (e.g. receiving end of a transmission). Metrics are needed that rely only on a very limited amount of information about the reference image. In such RR frameworks (shown in Fig. 4.1), low bandwidth features extracted from the reference image are transmitted to the receiver, where they are used in conjunction with the receiver data to assess the quality of the received image.

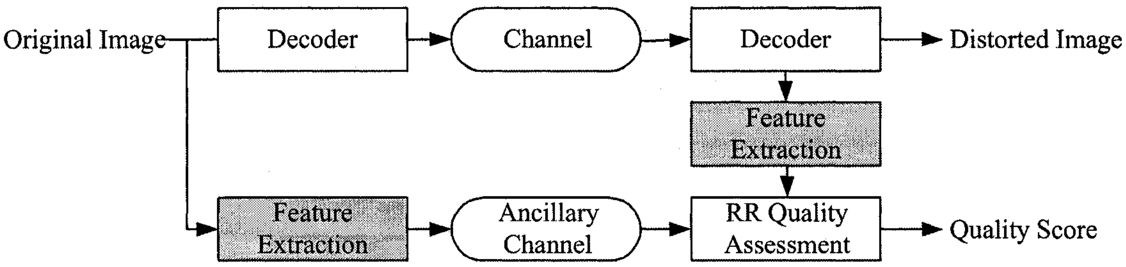


Fig. 4.1 Framework of RR quality assessment systems

The bandwidth available for transmitting the side information determines the amount of RR features that: if a high bandwidth is available then a large amount of information about the reference image can be included (if this bandwidth is high enough, FR QA is applicable). Conversely, smaller bandwidths reduce the amount of information about the reference image available at the receiver. The relationship between the bandwidth available for RR features and the accuracy of image quality assessment can be modeled as a monotonically increasing function as illustrated in Fig. 4.2: generally speaking, higher bandwidths enable more accurate assessment of image quality. The biggest challenge in RR QA resides in finding efficient features to optimize image quality prediction accuracy under the constraints of the available bandwidth.

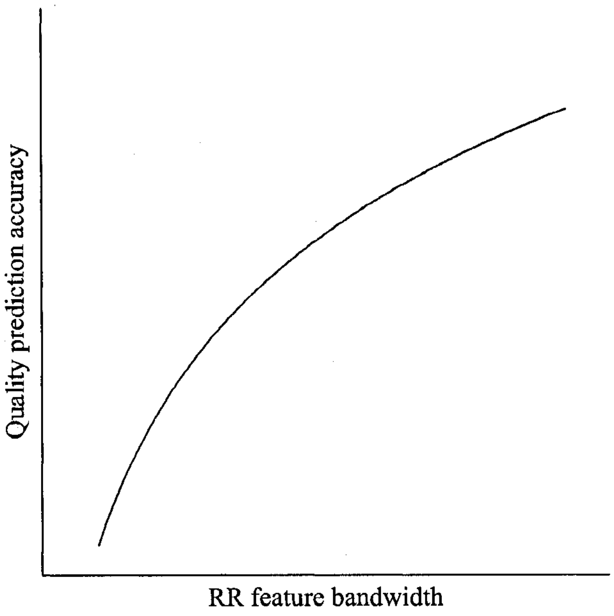


Fig. 4.2 Tradeoff between RR feature bandwidth and quality prediction accuracy

4.2 Reduced Reference Image Quality Assessment Using the DWT

The multiresolution nature of the DWT makes it an ideal feature candidate. The orthogonal matrices U and V Eq. 3.2 used to consist of two parts. The upper half-parts of U and V correspond to low-pass filters with coefficients $\{h_n\}$, and the lower half-parts are high-pass filters with coefficients $\{g_n\}$. U acts on the columns of the image and V^T acts on the rows of the image. The DWT is typically implemented through by filter banks which divide the image into four parts as follows:

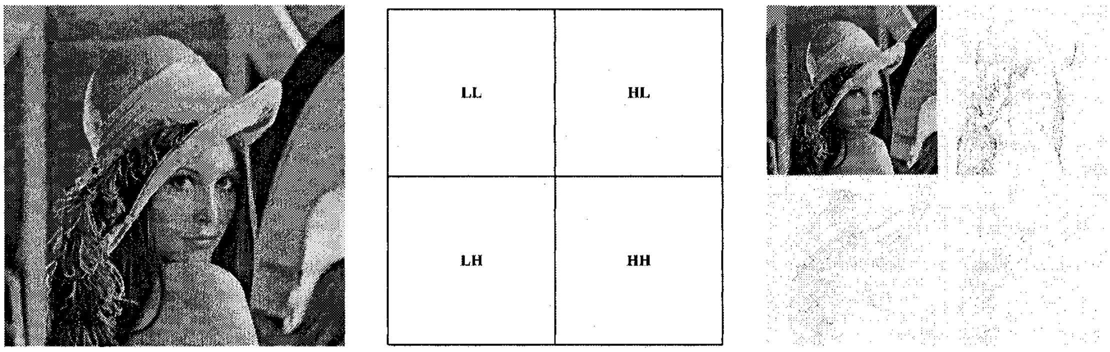
1. The top left part (LL) is produced by the two-dimensional scaling function $\varphi(x)\varphi(y)$ and is in fact an approximation (low-pass filtered and downsampled) version of the original image.

2. The top right part (HL) is produced by the vertical wavelet function $\psi(x)\varphi(y)$ and contains information on vertical details.

3. The bottom left part (LH) is produced by the horizontal wavelet function $\varphi(x)\psi(y)$ and contains information on horizontal details.

4. The bottom right (HH) part is is produced by the diagonal wavelet function $\psi(x)\psi(y)$ nd contains information on diagonal details.

This concept is illustrated in Fig. 4.3. The LL quadrant is smooth and has large values. The other three parts typically have small absolute values except for the edges.



(a) Original Lena image

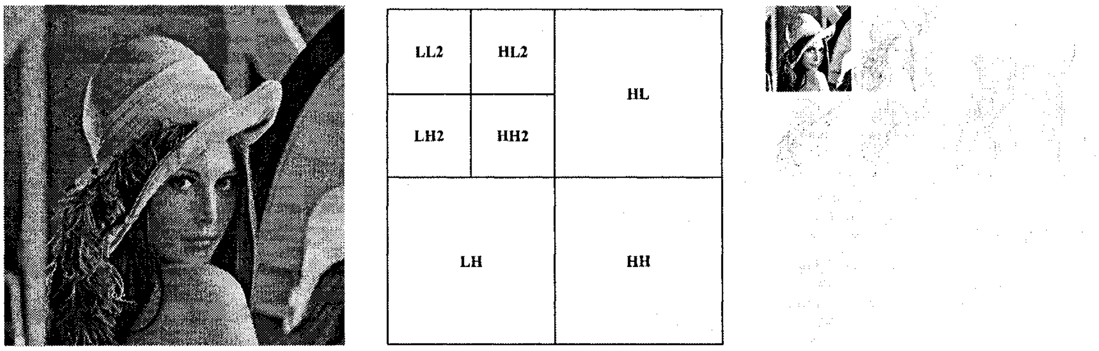
(b) DWT decomposition

(c) DWT decomposition of Lena

Fig. 4.3 DWT decomposition

The DWT can be applied recursively the LL subband. This process yields a n-level pyramid structure with $3n + 1$ different frequency bands including a single LL frequency band denoted LL_n which is a coarse approximation of the original signal. An example is given in Fig. 4.4 with $n = 2$.

We propose the following procedure based on the DWT to evaluate image quality in RR frameworks:



(a) Original Lena image

(b) 2-level DWT decomposition

(c) 2-level DWT decomposition of Lena

Fig. 4.4 2-level DWT decomposition

1. The reference and distorted images are converted to a luminance/chrominance color space using Eq. 3.1.
2. An n-level DWT is applied to the luminance images.
3. The RMSE between the LL_n DWT coefficients of the original and distorted images is be used to compute a RR image quality index Q_{RR} given by Eq. 4.1 below:

$$\mathrm{Q}_{\mathrm{LL}_n} = \sqrt{\mathrm{MSE}_{\mathrm{LL}_n}} \tag{4.1}$$

Practically speaking, for $M \times N$ images, this means that the size of the reference data needed for the comparison is $(M \times N)/2^{2n}$.

4.3 Experimental Results

Table 4.1 and Fig. 4.5 shows the CC performance of $\mathrm{Q}_{\mathrm{LL}_n}$ for different decompositions levels (n). The results show that as a general purpose metric, $\mathrm{Q}_{\mathrm{LL}_n}$ works well for 1- and 2-level decompositions and provides results that are comparable to the $\mathrm{Q}_{\mathrm{DWT}}$. However, for higher decomposition levels (LL_3 and above) the representation of the original image becomes too coarse and many perceptually important features are lost thus leading to a decrease in the performance of the metric. For distortions which discards a big portion of high-frequency components (JPEG2000 and JPEG) or those that affect mainly low-frequency components

(Gaussian blur) considerable reference data rate reductions are possible: in all of these cases the optimal value for n is 3.

Table 4.1 Prediction accuracy (CC) of Q_{LL_n} for n-level DWT decomposition

Distortion	Q_{DWT}	Q_{LL_1}	Q_{LL_2}	Q_{LL_3}	Q_{LL_4}	Q_{LL_5}	Q_{LL_6}	Q_{LL_7}
All images	0.9383	0.9407	0.9345	0.9206	0.9062	0.8792	0.8736	0.8671
JPEG2000	0.9657	0.9708	0.9767	0.9787	0.9784	0.9754	0.9711	0.954
JPEG	0.9601	0.9671	0.9725	0.9738	0.9737	0.972	0.9658	0.9541
White Noise	0.9929	0.9911	0.9868	0.9809	0.9725	0.963	0.9534	0.9408
Gaussian Blur	0.9228	0.9429	0.9679	0.9691	0.9674	0.9664	0.9584	0.963
FF	0.9512	0.9428	0.9008	0.8695	0.8634	0.7923	0.7692	0.7624

4.4 Summary

This chapter extended the scope of the frequency domain image QA method to RR frameworks. A reduced reference image quality metric based on the DWT was developed. Experiments confirmed the effectiveness of the RR metric making it suitable for real-time applications where only a limited amount of bandwidth is available for transmission of information about the reference image.

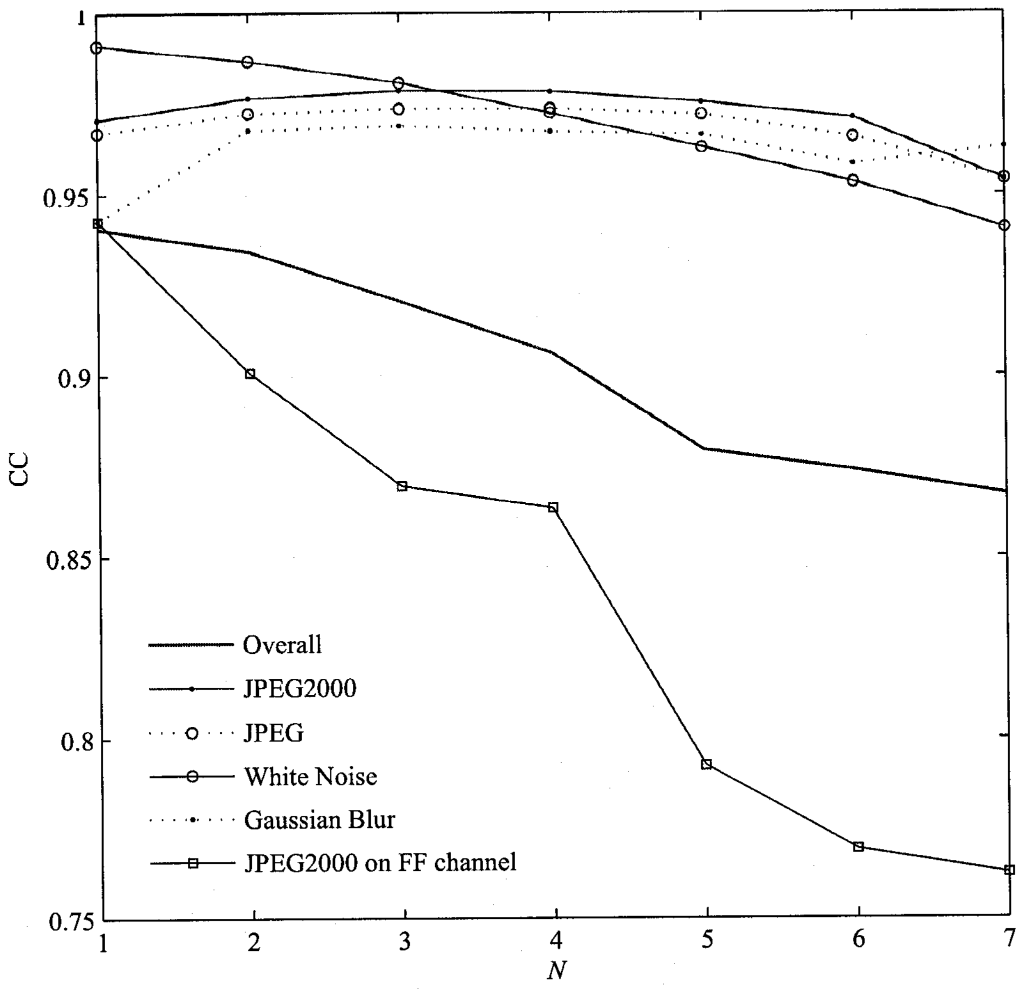


Fig. 4.5 CC performance of Q_{LL_n} as a function of decomposition levels n

Chapter 5

Conclusions

5.1 Thesis Summary

This thesis presents an quality assessment method that is based on frequency domains transforms, namely the DCT and the DWT. This approach is driven by pre-determined applications, namely visual quality assessment of compressed digital images. Two FR metrics for image quality assessment are developed. In simulations, these metrics outperform state-of-the-art metrics and prove to be useful over a wide range of image distortions. The method is successfully extended to RR frameworks where low bandwidth features are extracted from the reference image and transmitted to a RR algorithm running at the output of an image transmission system where these RR features are used to assess image quality. The metrics presented in this thesis are easy to implement, computationally efficient and do not require any extensive calibration.

5.2 Future Research Work

The metrics presented in this thesis could be extended to video quality assessment. One obvious and simple way to implement video quality metrics would be to apply the image quality assessment metrics developed in this thesis on a frame-by-frame basis and average the results to give a global video quality ratings. A more sophisticated approach would model the temporal dimension in the design of the metrics. Several implementation issues would need to be considered. One important factor affecting the feasibility of a video quality metric is its computational complexity. An extension of the FR image quality metrics developed in this thesis to FR video quality assessment would require tremendous computational resources. A RR video quality metrics based on the RR image quality metric developed in Section 4.2 would be the most practical solution. The challenge would reside in frames finding a suitable temporal information feature.

References

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

[2] ITU, "Methodology for the subjective assessment of quality for television pictures," Tech. Rep. ITU-R Recommendation BT.500-10, International Telecommunication Union, Geneva, Switzerland, Mar. 2000.

[3] V.-M. Liu, J.-Y. Lin, and K.-G. C.-N. Wang, "Objective image quality measure for block-based DCT coding," *IEEE Transactions on Consumer Electronics*, vol. 43, pp. 511–516, August 1997.

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

[5] A. C. Bovik and S. Liu, "DCT-domain blind measuement of blocking artifacts in DCT-coded images," in *Proc. IEEE Int. Conf. Image Acoust., Speech and Signal Proc.*, vol. 3, pp. 1725–1728, May 2001.

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

[7] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: Application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, pp. 163–172, February 2004.

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

[9] I. P. Gunawan and M. Ghanbari, "Reduced-reference picture quality estimation by using local harmonic amplitude information," in *Proc. London Communications Symposium*, pp. 137–140, 2003.

[10] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Proc. SPIE Conf. on Human Vision and Electronic Imaging X*, vol. 5666, pp. 149–159, January 2005.

[11] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Communications*, vol. 43, pp. 2959–2965, December 1995.

[12] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2."

[13] H. R. Sheikh, *Image Quality Assessment Using Natural Scene Statistics.* Ph.D. thesis, University of Texas at Austin, May 2004.

[14] B. A. Wandell, *Foundations of Vision.* Sunderland, MA: Sinauer Associates, 1995.

[15] D. H. Hubel, *Eye, Brain and Vision.* New York: W.H. Freeman Company, 1988.

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

[17] A. M. Eskicioglu and P. S. Fisher, "A survey of quality measures for gray scale image compression," in *Proc. 1993 Space and Earth Science Data Compression Workshop*, pp. 49–61, Apr. 1993.

[18] T. Pappas and R. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing* (A. Bovik, ed.), pp. 669–684, San Diego, CA: Academic Press, 2000.

[19] J. L. Manos and D. J. Sakrinson, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Information Theory*, vol. 20, no. 4, pp. 525–536, 2004.

[20] S. Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision* (A. B. Watson, ed.), pp. 163–178, Cambridge, MA: MIT Press, 1993.

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

[22] J. Lubin, "A visual discrimination model for image sstem design and evaluation," in *Visual Models for Target Detection and Recognition* (E. Peli, ed.), pp. 207–220, Singapore: World Scientific Publisher, 1995.

[23] P. Teo and D. Heeger, "Perceptual image distortion," in *Proc. IEEE Int. Conf. on Image Processing*, vol. 2, (Austin, TX), pp. 982–986, Nov. 1994.

[24] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE: Human Vision, Visual Processing and Digital Display IV*, vol. 1913, pp. 202–216, Sept. 1993.

[25] H. Peterson, A. Ahumada, and A. Watson, "The visibility of dct quantization noise," in *Proceedings of the Society For Information Display*, vol. 24, pp. 942–945, 1993.

[26] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Processing*, vol. 8, pp. 717–730, May 1999.

[27] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, pp. 600–612, Apr. 2004.

[28] A. Watson, G. Yang, J. Soloman, and J. Villasenor, "Visual thresholds for wavelet quantization error," in *Proceedings of the SPIE*, vol. 2657, p. 382392, 1996.

[29] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," tech. rep., Video Quality Expets Group, Aug. 2003.