Perceptual Criteria on Image Compression

Jesús Jaime Moreno Escobar

Computer Science Department

Universitat Autònoma de Barcelona

A dissertation submited to fullfill the degree of

Doctor en Informática (PhD on Informatics)

July 1st. 2011

Director	**Xavier Otazu Porter**
	Computer Science Department
	Universitat Autònoma de Barcelona, Spain.

Thesis Committee	**Jesús Malo López**
	Department of Optics
	School of Physics, Universitat de València, Spain.
	Michael W. Marcellin
	Department of Electrical and Computer Engineering
	University of Arizona, USA.
	Christine Fernandez-Maloigne
	Signal Image Communications Department
	Université de Poitiers, France.
	Jorge Núñez de Murga
	Astronomy and Meteorology Department
	Universitat de Barcelona, Spain.
	Joan Serra Sagristà
	Department of Information and Communication Engineering
	Universitat Autònoma de Barcelona, Spain.

A Erika y Jaimito (Osy)

...... y ...... a los que Dios nos dé.

Agradecimientos

Una frase de la cultura mexicana versa que *para bailar la bamba se necesita una poca de gracia y otra cosita*, es decir, el éxito de un proyecto no sólo depende del conocimiento del tema a tratar, sino también de saber sortear los factores externos que intervienen a favor o en contra.

Así, el terminar un trabajo doctoral no sólo significa agotar, en la medida de lo posible, los temas a tratar, sino también conlleva muchos sacrificios, esperando que estos valgan la pena. Para finalizar esta tesis se tuvieron que conjugar favorablemente una serie importante de circunstancias que sin la ayuda de mi gente, de mi *raza*, no se hubieran finalizado.

Es por ello que primeramente quisiera agradecer a mi esposa **Erika Aguilar** por respaldarme en esta aventura europea y ser soporte necesario para que esta y más aventuras lleguen a buen puerto. *Te amo mi vida.*

A mi pequeño hijo **Jaime III (Osy)** que con sus largas siestas y no despertar continuamente por la noche, hizo que papá se pudiese concentrar a fondo en su trabajo. *Esta va por tí, m'ijito.*

A mis padres, **Jaime Sr. y Marleny**, y a mi abuela, **Ana María**, que con sus enseñanzas y su temple, forjaron a la persona que hoy en día soy. *Que Dios los bendiga y los guarde por muchos años más.*

A mi amiga **Mary Cruz** que siempre estuvo muy atenta a que esta aventura no naufragara, me defendió sin algún interés y por ser la responsable al 100% del éxito

administrativo de esta misión. *No te fallé amiga.*

Pero le quiero agradecer sobre todas las cosas a mi **México**, que con todo y los problemas que ahora enfrentamos, me he dado cuenta que es no sólo un gran país sino una gran nación a la cual le puedo llamar mi único hogar. Que me formó y que a diario forma a personas que quieren hacer el bien, aunque no sean ni doctores, ni ingenieros, ni profesionales en alguna rama de las ciencias. *¡Viva México K!*

Al **Consejo Nacional de Ciencia y Tecnología** de México, por su apoyo y patrocinio para la realización de esta tesis, por otorgarme no sólo una beca sino una oportunidad de mejorar personal y profesionalmente. *Retribuiré profesionalmente cada centavo invertido en mî.*

También quiero agradecer al **Instituto Politécnico Nacional**, por darme la oportunidad de ausentarme por cuatro años de mi trabajo como profesor de Ingeniería en Comunicaciones y Electrónica para formarme profesionalmente. *Estoy seguro que no les voy a fallar.*

A mis profesores de la Maestría de Sistemas: Efraín Martínez, Ignacio Peón y Luis Manuel Hernández, por compartir conmigo sus invaluables conocimientos, que a seis años de distancia aún los conservo. *Gracias, me salvaron una vez más.*

A mi director de tesis el Dr. **Xavier Otazu**, que me rescató y encaminó de manera exitosa este proyecto cuando el panorama pintaba gris tendiendo a ser negro. *Gracias Xavi.*

A mis **amigos y familiares**, como Elena Acevedo y mi hermana Marle, que siempre tuvieron una palabra de apoyo y aliento para mí durante mis estudios. *Gracias por todo.*

<div align="center">

Jesús Jaime Moreno Escobar

Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Zacatenco

Instituto Politécnico Nacional, México, 2011

</div>

Abstract

Nowadays, digital images are used in many areas in everyday life, but they tend to be big. This increases amount of information leads us to the problem of image data storage. For example, it is common to have a representation a color pixel as a 24-bit number, where the channels red, green, and blue employ 8 bits each. In consequence, this kind of color pixel can specify one of $2^{24} \approx 16.78$ million colors. Therefore, an image at a resolution of 512×512 that allocates 24 bits per pixel, occupies 786,432 bytes. That is why image compression is important.

An important feature of image compression is that it can be lossy or lossless. A compressed image is acceptable provided these losses of image information are not perceived by the eye. It is possible to assume that a portion of this information is redundant. Lossless Image Compression is defined as to mathematically decode the same image which was encoded. In Lossy Image Compression needs to identify two features inside the image: the redundancy and the irrelevancy of information. Thus, lossy compression modifies the image data in such a way when they are encoded and decoded, the recovered image is similar enough to the original one. How similar is the recovered image in comparison to the original image is defined prior to the compression process, and it depends on the implementation to be performed.

In lossy compression, current image compression schemes remove information considered irrelevant by using mathematical criteria. One of the problems of these schemes is that although the numerical quality of the compressed image is low, it shows a high visual image quality, e.g. it does not show a lot of visible artifacts. It is because these mathematical criteria, used to remove information, do not take into account if the viewed information is perceived by the Human Visual System. Therefore, the aim of an image compression scheme designed to obtain images that do not show artifacts although their numerical quality can be low, is to eliminate the information that is not visible by the Human Visual System.

Hence, this Ph.D. thesis proposes to exploit the visual redundancy existing in an image by reducing those features that can be unperceivable for the Human Visual System.

First, we define an image quality assessment, which is highly correlated with the psychophysical experiments performed by human observers. The proposed $\mathcal{C}_w$PSNR metrics weights the

well-known PSNR by using a particular perceptual low level model of the Human Visual System, e.g. the Chromatic Induction Wavelet Model (CIWaM). Second, we propose an image compression algorithm (called Hi-SET), which exploits the high correlation and self-similarity of pixels in a given area or neighborhood by means of a fractal function. Hi-SET possesses the main features that modern image compressors have, that is, it is an embedded coder, which allows a progressive transmission. Third, we propose a perceptual quantizer (ρSQ), which is a modification of the uniform scalar quantizer. The ρSQ is applied to a pixel set in a certain Wavelet sub-band, that is, a global quantization. Unlike this, the proposed modification allows to perform a local pixel-by-pixel forward and inverse quantization, introducing into this process a perceptual distortion which depends on the surround spatial information of the pixel. Combining ρSQ method with the Hi-SET image compressor, we define a perceptual image compressor, called Φ_{SET}. Finally, a coding method for Region of Interest areas is presented, ρGBbBShift, which perceptually weights pixels into these areas and maintains only the more important perceivable features in the rest of the image.

Results presented in this report show that $\mathcal{C}_w$PSNR is the best-ranked image quality method when it is applied to the most common image compression distortions such as JPEG and JPEG2000. $\mathcal{C}_w$PSNR shows the best correlation with the judgement of human observers, which is based on the results of psychophysical experiments obtained for relevant image quality databases such as TID2008, LIVE, CSIQ and IVC. Furthermore, Hi-SET coder obtains better results both for compression ratios and perceptual image quality than the JPEG2000 coder and other coders that use a Hilbert Fractal for image compression. Hence, when the proposed perceptual quantization is introduced to Hi-SET coder, our compressor improves its numerical and perceptual efficiency. When ρGBbBShift method applied to Hi-SET is compared against MaxShift method applied to the JPEG2000 standard and Hi-SET, the images coded by our ROI method get the best results when the overall image quality is estimated. Both the proposed perceptual quantization and the ρGBbBShift method are generalized algorithms that can be applied to other Wavelet based image compression algorithms such as JPEG2000, SPIHT or SPECK.

Resumen

Hoy en día las imágenes digitales son usadas en muchas areas de nuestra vida cotidiana, pero estas tienden a ser cada vez más grandes. Este incremento de información nos lleva al problema del almacenamiento de las mismas. Por ejemplo, es común que la representación de un pixel a color ocupe 24 bits, donde los canales rojo, verde y azul se almacenen en 8 bits. Por lo que, este tipo de pixeles en color pueden representar uno de los $2^{24} \approx 16.78$ millones de colores. Así, una imagen de 512×512 que representa con 24 bits un pixel ocupa 786,432 bytes. Es por ello que la compresión es importante.

Una característica importante de la compresión de imágenes es que esta puede ser con perdidas o sin ellas. Una imagen es aceptable siempre y cuando dichas perdidas en la información de la imagen no sean percibidas por el ojo. Esto es posible al asumir que una porción de esta información es redundante. La compresión de imágenes sin pérdidas es definida como decodificar matemáticamente la misma imagen que fue codificada. En la compresión de imágenes con pérdidas se necesita identificar dos características: la redundancia y la irrelevancia de información. Así la compresión con pérdidas modifica los datos de la imagen de tal manera que cuando estos son codificados y decodificados, la imagen recuperada es lo suficientemente parecida a la original. Que tan parecida es la imagen recuperada en comparación con la original es definido previamente en proceso de codificación y depende de la implementación a ser desarrollada.

En cuanto a la compresión con pérdidas, los actuales esquemas de compresión de imágenes eliminan información irrelevante utilizando criterios matemáticos. Uno de los problemas de estos esquemas es que a pesar de la calidad numérica de la imagen comprimida es baja, esta muestra una alta calidad visual, dado que no muestra una gran cantidad de artefactos visuales. Esto es debido a que dichos criterios matemáticos no toman en cuenta la información visual percibida por el Sistema Visual Humano. Por lo tanto, el objetivo de un sistema de compresión de imágenes diseñado para obtener imágenes que no muestren artefactos, aunque su calidad numérica puede ser baja, es eliminar la información que no es visible por el Sistema Visual Humano.

Así, este trabajo de tesis doctoral propone explotar la redundancia visual existente en una imagen, reduciendo frecuencias imperceptibles para el sistema visual humano.

Por lo que primeramente, se define una métrica de calidad de imagen que está altamente correlacionada con opiniones de observadores. La métrica propuesta pondera el bien conocido PSNR por medio de una modelo de inducción cromática ($\mathcal{C}_w$PSNR). Después, se propone un algoritmo compresor de imágenes, llamado Hi-SET, el cual explota la alta correlación de un vecindario de pixeles por medio de una función Fractal. Hi-SET posee las mismas características que tiene un compresor de imágenes moderno, como ser una algoritmo *embedded* que permite la transmisión progresiva. También se propone un cuantificador perceptual(ρSQ), el cual es una modificación a la clásica cuantificación *Dead-zone*. ρSQes aplicado a un grupo entero de pixeles en una sub-banda Wavelet dada, es decir, se aplica una cuantificación global. A diferencia de lo anterior, la modificación propuesta permite hacer una cuantificación local tanto directa como inversa pixel-por-pixel introduciéndoles una distorsión perceptual que depende directamente de la información espacial del entorno del pixel. Combinando el método ρSQ con Hi-SET, se define un compresor perceptual de imágenes, llamado Φ_{SET}. Finalmente se presenta un método de codificación de areas de la Región de Interés, ρGBbBShift, la cual pondera perceptualmente los pixeles en dichas areas, en tanto que las areas que no pertenecen a la Región de Interés o el Fondo sólo contendrán aquellas que perceptualmente sean las más importantes.

Los resultados expuestos en esta tesis indican que $\mathcal{C}_w$PSNR es el mejor indicador de calidad de imagen en las distorsiones más comunes de compresión como son JPEG y JPEG2000, dado que $\mathcal{C}_w$PSNR posee la mejor correlación con la opinión de observadores, dicha opinión está sujeta a los experimentos psicofísicos de las más importantes bases de datos en este campo, como son la TID2008, LIVE, CSIQ y IVC. Además, el codificador de imágenes Hi-SET obtiene mejores resultados que los obtenidos por JPEG2000 u otros algoritmos que utilizan el fractal de Hilbert. Así cuando a Hi-SET se la aplica la cuantificación perceptual propuesta, Φ_{SET}, este incrementa su eficiencia tanto objetiva como subjetiva. Cuando el método ρGBbBShift es aplicado a Hi-SET y este es comparado contra el método MaxShift aplicado al estándar JPEG2000 y a Hi-SET, se obtienen mejores resultados perceptuales comparando la calidad subjetiva de toda la imagen de dichos métodos. Tanto la cuantificación perceptual propuesta ρSQ como el método ρGBbBShift son algoritmos generales, los cuales pueden ser aplicados a otros algoritmos de compresión de imágenes basados en Transformada Wavelet tales como el mismo JPEG2000, SPIHT o SPECK, por citar algunos ejemplos.

Contents

List of Figures

List of Tables

List of Acronyms

BbBShift	Bitplane-by-Bitplane Shift
bpp	Bits per Pixel
CIWaM	Chromatic Induction Wavelet Model
C$_\omega$PSNR	Peak Signal-to-Noise Ratio weighted by the Chromatic Induction Wavelet Model
DCT	Discrete Cosine Transform
EBCOT	Embedded Block Coding with Optimized Truncation
e-CSF	extended Contrast Sensitivity Function
F-ρSQ	Forward Perceptual Scalar Quantizer
GBbBShift	Generalized Bitplane-by-Bitplane Shift
H	Height of a 512×512 pixel image presented in an M_{size} LCD monitor with horizontal resolution of h_{res} pixels and v_{res} pixels of vertical resolution.
HVS	Human Visual System
ICT	Irreversible Component Transformation
IFC	Image Fidelity Criterion
I-ρSQ	Inverse Perceptual Scalar Quantizer
KROCC	Kendall Rank-Order Correlation Coefficient
LSP	List of Significant Pixels
MSE	Mean Square Error
MSSIM	Multiscale Structural Similarity Index

NQM	Noise Quality Measure
PCC	Pearson Correlation Coefficient
PSNR	Peak Signal-to-Noise Ratio
ρGBbBShift	Perceptual Generalized Bitplane-by-Bitplane Shift
RCT	Reversible Component Transformation
RD	Rate Distortion
ROI	Region of Interest
RSI	Remote Sensing Images
SNR	Signal-to-Noise Ratio
SQ	Dead-zone uniform scalar quantizer
SR	Strength of Relationship
SROCC	Spearman Rank-Order Correlation Coefficient
SSIM	Structural Similarity Index
UQI	Universal Quality Index
VFW	Visual Frequency Weighting
VIF	Visual Information Fidelity
VIFP	Pixel-based Visual Information Fidelity
VSNR	Visual Signal-to-Noise Ratio
WSNR	Weighted Signal-to-Noise Ratio

Chapter 1

Introduction

The main objective of this thesis is the introduction of perceptual criteria into the image compression process. On the one hand, a perceptual image quality assessment is defined, in order to evaluate the visual quality of a compressed image. On the other hand, to identify and to remove non-perceptual information of an image, maintaining as far as possible, the same entropy as the source image. Furthermore, we introduce these perceptual criteria into a proposed image compression system.

1.1 Problem Statement

One of the most amazing abilities of human beings is *Vision*, since it is considered the most important sense, but the most difficult to model. When a light ray enters into our eyes undergoes a highly complex process, which ends in the visual cortex of brain. Color researches try to model some of these features of the Human Visual System(HVS). If accurate models are developed, they can be easily incorporated into many image processing applications such as Quality Assessment instruments and image compression schemes.

Nowadays, Mean Squared Error (MSE) is still the most used quantitative performance metrics and several quality measures are based on it, Peak Signal-to-Noise Ratio (PSNR) is the best example of this usage. But some authors like Wang and Bovik in (57, 59) consider that MSE is a poor device to be used in quality assessment systems. Therefore, it is important to know what is the MSE and what is wrong with it, in order

to propose a new indicator that fulfills the properties of HVS and keeps the favorable features that the MSE has.

Digital image compression has been a research topic for many years and a number of image compression standards has been created for different applications. The JPEG2000 is intended to provide rate-distortion and subjective image quality performance superior to existing standards, as well as to supply another functionalities (10). However, JPEG2000 does not provide the most relevant characteristics of the human visual system, since for removing information in order to compress the image mainly information theory criteria are applied. This information removal introduces artifacts to the image that are visible at high compression rates, because of many pixels with high perceptual significance have been discarded.

Hence it is necessary an advanced model that removes information according to perceptual criteria, preserving the pixels with high perceptual relevance regardless of its numerical information. The Chromatic Induction Wavelet Model (CIWaM, proposed by Otazu et. al. in (32, 33)) presents some perceptual concepts that can be suitable for it. Both CIWaM and JPEG2000 use wavelet transform.

CIWaM uses the wavelet transform in order to generate an approximation to how every pixel is perceived from a certain distance taking into account the value of its neighboring pixels. CIWaM inhibits or attenuates the details that the human visual system is not able to perceive, enhances those that are perceptually relevant and producing an approximation of the image that the brain visual cortex perceives. By contrast, JPEG2000 applies a perceptual criteria for all coefficients in a certain spatial frequency independently of its surrounding values. In other words, JPEG2000 performs a global transformation of wavelet coefficients, while CIWaM performs a local one.

Therefore, this dissertation is centered in the definition of a perceptual image quality metrics, as well as the incorporation of CIWaM, in many parts of a image compression system.

1.2 Image Compression Systems

General System Theory defines $information = -entropy$, this is, entropy is the tendency that systems have when they wear down or disintegrate by themselves or by external factors(8). Thus, entropy means the loss of a given information. Then, a

compressed image should have almost the same total entropy as the original, but using fewer bits. That is, a compressed image has more entropy per bit than its original image. The main goal of modern image compression systems is to exploit redundancies of images, understanding some information as redundant. These redundancies can be either statistical or due to visual or application specific irrelevancies(51, Sec. 1.2).

In general, a system is composed by four subsystems: an input, a process, an output and a feedback (cybernetic model depicted in Figure 1.1). Hence, a system can be defined as a set of elements standing in interrelation among themselves and with environment.
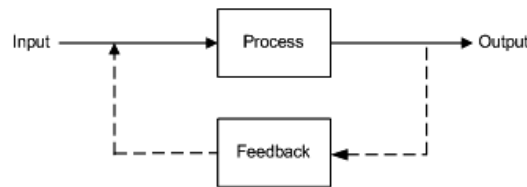
Figure 1.1: Description of *System* according to the General System Theory.

The subsystem *Process* is a black box for the subsystem *Feedback*, and vice versa. *Feedback* is employed in order to adjust some parameters or to assess the efficiency of the *Process*. Similarly, an image compression algorithm can be described as follows, Figure 1.2:

- *Input*: Original image considered with infinite quality $f(i, j)$;

- *Process*: Set of sub-processes, these are commonly: Forward Transformation (Section 3.3), Quantization (Section 3.4), Entropy Coding, Entropy Decoding, Inverse Quantization and Inverse Transformation. When a ROI algorithm is used, it is placed before Entropy Coding;

- *Output*: Reconstructed image $\hat{f}(i, j)$, whose quality has been presumably distorted;

- *Feedback*: Assessment of the posible distortion between original and reconstructed images, in order to measure the efficiency of the image compression system. MSE and PSNR are the most common image quality assessments. Advantages and drawbacks of these important measurements are described in Section 2.1.

Figure 1.2: General Block Diagram for an image compression system.

1.3 Proposed Perceptual Image Compression System

In this dissertation, we introduce perceptual criteria in specific sub-process of a general image compression system, Figure 1.1, such as Forward and Inverse Perceptual Quantization, Perceptual Region of Interest, a new Entropy Coder, besides a perceptual image quality assessment, green blocks in Figure 1.3.

Therefore the parts, that our system includes,are:

- *Input*: Original image considered with infinite quality $f(i,j)$;

- *Process*: Set of sub-processes: Forward Wavelet Transformation (9/7 analysis Filter, Table 3.2), Forward Perceptual Quantization (using a Chromatic Induction Model, Section 2.2), Hi-SET Coding (Sec. 3.5), Hi-SET Decoding, Inverse Perceptual Quantization (Section 4.4) and Inverse Wavelet Transformation(9/7 synthesis Filter, Table 3.2). When it is important to encode and to decode an specific area of the image first, we propose a Region of Interest algorithm, ρGBbBShift method, described in Section 5.2.2);

- *Output*: Reconstructed image $\hat{f}_p(i,j)$, whose perceptually important frequencies have been enhanced of the rest of frequencies;

- *Feedback*: The proposed image compression system needs a perceptual metrics, which is why we propose a perceptual assessment, $\mathcal{C}_w$PSNR, based on the interpretation of perceptual energy degradation.
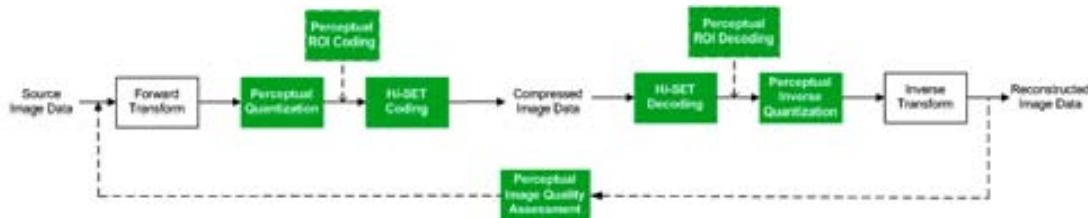
Figure 1.3: General Block Diagram for the proposed perceptual image compression system. Contribution of this thesis are the green blocks.

1.4 Thesis Outline

This dissertation consist of four chapters (2 to 5) that describe the contributions of this work.

In Chapter 2 we propose a quality assessment, which weights the mainstream PSNR by means of a chromatic induction model ($\mathcal{C}_w$PSNR). This is feasible referenced-measuring the rate of energy loss when an image is observed at different distances. $\mathcal{C}_w$PSNR is the best-image quality assessment, when an image is distorted by JPEG blocking or wavelet ringing, namely images compressed by any Discrete Cosine Transform (DCT) or wavelet based image coder, across databases TID2008, LIVE, CSIQ and IVC not only on an individual image database but also overall performance.

In Chapter 3 we present an effective and computationally simple coder for image compression based on Hilbert Scanning of Embedded quad Trees (Hi-SET). It allows to represent an image as an embedded bitstream along a fractal function, avoiding to store coordinate locations. Embedding is an important feature of modern image compression algorithms, in this way Salomon in (42, pg. 614) cites that another feature and perhaps a unique one is the fact of achieving the best quality for the number of bits input by the decoder at any point during the decoding. Hi-SET possesses also this latter feature. Furthermore, the Hi-SET coder is based on a quadtree partition strategy, which is naturally adapted to image transformation structures such as discrete cosine or wavelet transform. This last property allows to obtain an effective energy clustering both in frequency and space. The coding algorithm is composed of three general steps, using, unlike some state of the art algorithms, only one ordered list, the list of significant pixels.

The aim of Chapter 4 is to explain how to apply perceptual criteria in order to

define a perceptual forward and inverse quantizer. We present its application to the Hi-SET coder. Our approach consists in quantizing wavelet transform coefficients using some of the human visual system behavior properties. Taking in to account that noise is fatal to image compression performance, because it can be both annoying for the observer and consumes excessive bandwidth when the imagery is transmitted. Perceptual quantization reduces unperceivable details and thus improve both visual impression and transmission properties. The comparison between JPEG2000 coder and the combination of Hi-SET with the proposed perceptual quantizer (Φ_{SET}) shows that the latter is not favorable in PSNR than the former, but the recovered image is more compressed (less bit-rate) at the same or even better visual quality measured with well-know image quality metrics, such as MSSIM, UQI or VIF, for instance.

Chapter 5 describes a perceptual method (ρGBbBShift) for codding of Region of Interest (ROI) areas. It introduces perceptual criteria to the GBbBShift method when bitplanes of ROI and no-ROI background areas are shifted. This additional feature is intended for balancing perceptual importance of some coefficients regardless their numerical importance. Hence, there is no observing visual difference at ROI when the MaxShift method and the proposed method are compared, at the same time that perceptual quality of the entire image is improved.

Finally general conclusions are drawn, in addition to some recommendations for a continuation of this work are presented.

Chapter 2

Full-Reference Quality Assessment using a Chromatic Induction Model: JPEG and JPEG2000

2.1 Introduction

Nowadays, Mean Squared Error (MSE) is still the most used quantitative performance metrics and several image quality measures are based on it, being Peak Signal-to-Noise Ratio (PSNR) the best example. But some authors like Wang and Bovik in (57, 59) consider that MSE is a poor device to be used in quality assessment systems. Therefore it is important to know what is the MSE and what is wrong with it, in order to propose new metrics that fulfills the properties of human visual system and keeps the favorable features that the MSE has.

In this way, let $f(i,j)$ and $\hat{f}(i,j)$ represent two images being compared and the size of them is the number of intensity samples or pixels. Being $f(i,j)$ the original reference image, which has to be considered with perfect quality, and $\hat{f}(i,j)$ a distorted version of $f(i,j)$, whose quality is being evaluated. Then, the MSE and the PSNR are, respectively, defined as:

$$MSE = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left[f(i,j) - \hat{f}(i,j) \right]^2 \tag{2.1}$$

7

and

$$PSNR = 10 \log_{10} \left(\frac{\mathcal{G}_{max}^{2}}{MSE} \right) \qquad (2.2)$$

where $\mathcal{G}_{max}$ is the maximum possible intensity value in $f(i,j)$ ($M \times N$ size). Thus, for gray-scale images that allocate 8 bits per pixel (bpp) $\mathcal{G}_{max} = 2^{8} - 1 = 255$. For color images the PSNR is defined as in the Equation 2.2, whereas the color MSE is the mean among the individual MSE of each component.

An important task in image compression systems is to maximize the correlation among pixels, because the higher correlation at the preprocessing, the more efficient algorithm postprocessing. Thus, an efficient measure of image quality should take in to account the latter feature. In contrast to this, MSE does not need any positional information of the image, thus pixel arrangement is ordered as a one-dimensional vector.

Both MSE and PSNR are extensively employed in the image processing field, since these metrics have favorable properties, such as:

1. A convenient metrics for the purpose of algorithm optimization. For example in JPEG2000, MSE is used both in Optimal Rate Allocation (5, 51) and Region of interest (6, 51). Therefore MSE can find solutions for these kind of problems, when is combined with the instruments of linear algebra, since it is differentiable.

2. By definition MSE is the difference signal between the two images being compared, giving a clear meaning of the overall error signal energy.

However, the MSE has a poor correlation with perceived image quality. An example is shown in Fig. 2.1, where both (a) *Baboon* and (b) *Splash* Images are distorted by means of a JPEG2000 compression with PSNR=30 dB. These noisy images present dramatically different visual qualities. Thereby either MSE or PSNR do not reflect the way that human visual system (HVS) perceives images, since these measures represent an input image in a pixel domain.

In section 2.2 we outline the CIWaM chromatic induction model. It inhibits or enhances information according to perceptual criteria, preserving the pixels with high perceptual relevance and inhibiting those with low perceptual impact. This model is important for section 2.3, since $\mathcal{C}_{w}$PSNR makes use of it. The $\mathcal{C}_{w}$PSNR methodology is subdivided in five steps, which are also described in this section. Section 2.4 shows

(a) Image *Baboon*   (b) Image *Splash*

Figure 2.1: 256×256 patches (cropped for visibility) of Images *Baboon* and *Splash* distorted by means of JPEG2000 compression, although both images have the same objective quality (PSNR=30dB), their visual quality is very different. Original size 512×512 of both images are shown in Figures 2.10(b) and 2.10(c), respectively.

experimental results, comparing $\mathcal{C}_w$PSNR with twelve image quality metrics such as MSSIM (54), SSIM (45) and VIF (58), among others. In these tests, we use the perceptual image quality information supplied by four image databases TID2008 (38, 39), LIVE (46), CSIQ (22) and IVC (23).

2.2   Chromatic Induction Wavelet Model: Brief description.

The *C*hromatic *I*nduction *Wa*velet *M*odel (CIWaM) (32) is a low-level perceptual model of the HVS. It estimates the image perceived by an observer at a distance d just by modeling the perceptual chromatic induction processes of the HVS. That is, given an image $\mathfrak{I}$ and an observation distance d, CIWaM obtains an estimation of the perceptual image $\mathfrak{I}_\rho$ that the observer perceives when observing $\mathfrak{I}$ at distance d. CIWaM is based on just three important stimulus properties: spatial frequency, spatial orientation and surround contrast. This three properties allow to unify the chromatic assimilation and contrast phenomena, as well as some other perceptual processes such as saliency perceptual processes (29).

The CIWaM model takes an input image $\mathfrak{I}$ and decomposes it into a set of wavelet planes $\omega_{s,o}$ of different spatial scales s (i.e., spatial frequency ν) and spatial orientations

(a)   (b)
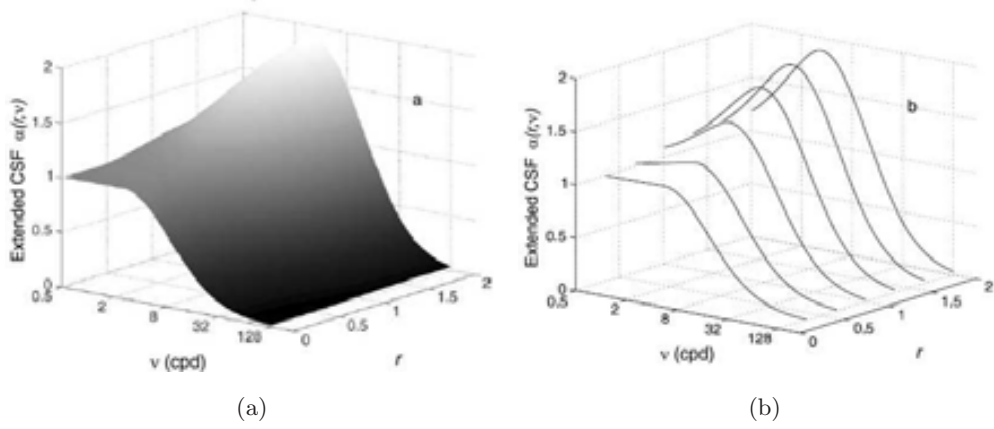
Figure 2.2: (a) Graphical representation of the e-CSF $(\alpha_{s,o,i}(r,\nu)))$ for the luminance channel. (b) Some profiles of the same surface along the Spatial Frequency (ν) axis for different centersurround contrast energy ratio values (r). The psychophysically measured CSF is a particular case of this family of curves (concretely for r = 1).

o. It is described as:

$$\mathfrak{I} = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \omega_{s,o} + c_n \, , \tag{2.3}$$

where n is the number of wavelet planes, c_n is the residual plane and o is the spatial orientation either *v*ertical, *h*orizontal or *d*ia*g*onal.

The perceptual image $\mathfrak{I}_\rho$ is recovered by weighting these $\omega_{s,o}$ wavelet coefficients using the *extended Contrast Sensitivity Function* (e-CSF, Fig. 2.2). The e-CSF is an extension of the psychophysical CSF (28) considering spatial surround information (denoted by r), visual frequency (denoted by ν, which is related to spatial frequency by observation distance) and observation distance (d). Perceptual image $\mathfrak{I}_\rho$ can be obtained by

$$\mathfrak{I}_\rho = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \alpha(\nu,r) \, \omega_{s,o} + c_n \, , \tag{2.4}$$

where $\alpha(\nu,r)$ is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term $\alpha(\nu,r) \, \omega_{s,o} \equiv \omega_{s,o;\rho,d}$ can be considered the *perceptual wavelet coefficients* of image $\mathfrak{I}$ when observed at distance d and is written as:

$$\alpha(\nu,r) = z_{ctr} \cdot C_d(\dot{s}) + C_{min}(\dot{s}) \, . \tag{2.5}$$

This function has a shape similar to the e-CSF and the three terms that describe it are defined as:

z_{ctr} Non-linear function and estimation of the central feature contrast relative to its surround contrast, oscillating from zero to one, defined by:

$$z_{ctr} = \frac{\left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2}{1 + \left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2} \tag{2.6}$$

being σ_{cen} and σ_{sur} the standard deviation of the wavelet coefficients in two concentric rings, which represent a center$-$surround interaction around each coefficient.

$C_d(\dot{s})$ Weighting function that approximates to the perceptual e-CSF, emulates some perceptual properties and is defined as a piecewise Gaussian function (27), such as:

$$C_d(\dot{s}) = \begin{cases} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ e^{-\frac{\dot{s}^2}{2\sigma_2^2}}, & \dot{s} = s - s_{thr} > 0. \end{cases} \tag{2.7}$$

$C_{min}(\dot{s})$ Term that avoids $\alpha(\nu, r)$ function to be zero and is defined by:

$$C_{min}(\dot{s}) = \begin{cases} \frac{1}{2} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ \frac{1}{2}, & \dot{s} = s - s_{thr} > 0. \end{cases} \tag{2.8}$$

taking $\sigma_1 = 2$ and $\sigma_2 = 2\sigma_1$. Both $C_{min}(\dot{s})$ and $C_d(\dot{s})$ depend on the factor s_{thr}, which is the scale associated to 4cpd when an image is observed from the distance d with a pixel size l_p and one visual degree, whose expression is defined by Equation 2.9. Where s_{thr} value is associated to the e-CSF maximum value.

$$s_{thr} = \log_2\left(\frac{d\tan(1°)}{4\,l_p}\right) \tag{2.9}$$

Fig. 2.3 shows three examples of CIWaM images of *Lenna*, calculated by Eq. 2.4 for a 19 inch monitor with 1280 pixels of horizontal resolution, at $d = \{30, 100, 200\}$ centimeters.

| (a) Original image | (b) d=30 cm. | (c) d=100 cm. | (d) d=200 cm. |
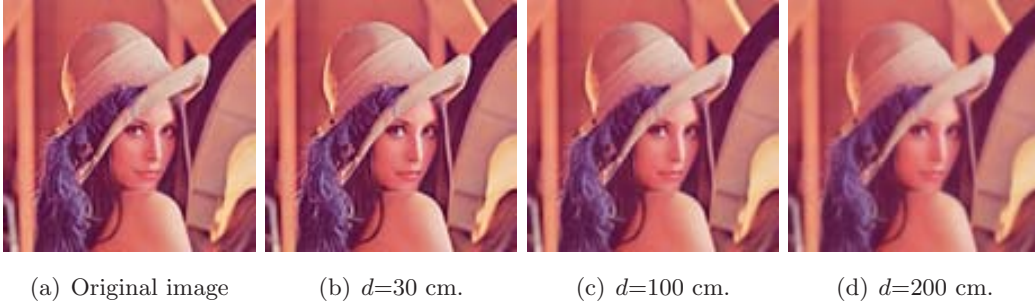
Figure 2.3: (a) Original color image *Lenna* . (b)-(d) Perceptual images obtained by CIWaM at different observation distances d.

2.3   CIWaM weighted Peak Signal-to-Noise Ratio

Figure 2.4: General block diagram for the proposed perceptual image compression system. $\mathcal{C}_w$PSNR is indicated by the green block.

In the referenced image quality issue, there is an original image $f(i,j)$ and a distorted version $\hat{f}(i,j) = \Lambda[f(i,j)]$ that is compared with $f(i,j)$, being Λ a distortion model. The difference between these two images depends on the features of the distortion model Λ. For example, blurring, contrast change, noise, JPEG blocking or wavelet ringing.

In Fig. 2.1, the images *Babbon* and *Splash* are compressed by means of JPEG2000. These two images have the same PSNR=30 dB when compared to their corresponding original image, that is, they have the same numerical degree of distortion (i.e. the same objective image quality PSNR). But, their subjective quality is clearly different, showing the image *Baboon* a better visual quality. Thus, for this example, PSNR and perceptual image quality has a small correlation. On the image *Baboon*, high spatial frequencies are dominant. A modification of these high spatial frequencies by Λ induces a high distortion, resulting a lower PSNR, even if the modification of these high frequencies are not perceived by the HVS. In contrast, on image *Splash*, mid and low frequencies are dominant. Modification of mid and low spatial frequencies also intro-

duces a high distortion, but they are less perceived by the HVS. Therefore, correlation of PSNR against the opinion of an observer is small. Fig. 2.5 shows the diagonal high spatial frequencies of these two images, where there are more high frequencies in image *Baboon*.
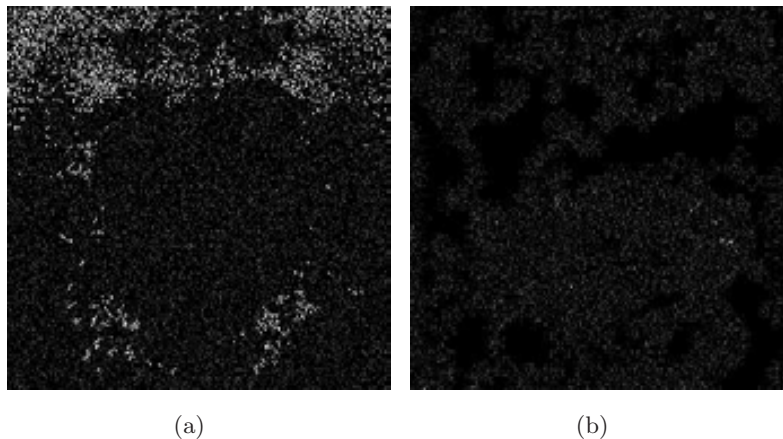
Figure 2.5: Diagonal spatial orientation of the first wavelet plane of Images (a) *Baboon* and (b)*Splash* distorted by JPEG2000 with PSNR=30dB.

If a set of distortions $\hat{f}_k(i,j) = \Lambda_k[f(i,j)]$ is generated and indexed by k (for example, let Λ be a blurring operator), the image quality of $\hat{f}_k(i,j)$ evolves while varying k, being k, for example, the degree of blurring. Hence, the evolution of $\hat{f}_k(i,j)$ depends on the characteristics of the original $f(i,j)$. Thus, when increasing k, if $f(i,j)$ contains many high spatial frequencies the PSNR rapidly decreases, but when low and mid frequencies predominated PSNR slowly decreases.

Similarly, the HVS is a system that induces a distortion on the observed image $f(i,j)$, whose model is predicted by CIWaM. Hence, CIWaM is considered a HSV particular distortion model $\Lambda \equiv$ CIWaM that generates a perceptual image $\hat{f}_\rho(i,j) \equiv \mathfrak{I}_\rho$ from an observed image $f(i,j) \equiv \mathfrak{I}$, i.e $\mathfrak{I}_\rho = CIWaM[\mathfrak{I}]$. Therefore, a set of distortions is defined as $\Lambda_k \equiv$ CIWaM$_d$, being d the observation distance. That is, a set of perceptual images is defined $\mathfrak{I}_{\rho,d} = $ CIWaM$_d[\mathfrak{I}]$ which is considered a set of perceptual distortions of image $\mathfrak{I}$.

When images $f(i,j)$ and $\hat{f}(i,j)$ are simultaneously observed at distance $\bar{d}$ and this distance is reduced, the differences between them are better perceived. In contrast, if $f(i,j)$ and $\hat{f}(i,j)$ are observed from a far distance human eyes cannot perceive their differences, in consequence, the perceptual image quality of the distorted image is

always high. The distance where the observer cannot distinguish any difference between these two images is $\bar{d} = \infty$. In practice, $\bar{d} = \mathcal{D}$ where differences are not perceived and range some centimeters from the position of the observer. Consequently, the less distorted $\hat{f}(i,j)$, the highest the image quality of $\hat{f}(i,j)$) and the shorter the distance D.

2.3.1 Methodology

Let $f(i,j)$ and $\hat{f}(i,j) = \Lambda[f(i,j)]$ be an original image and a distortion version of $f(i,j)$, respectively. $\mathcal{C}_w$PSNR methodology is based on finding a distance D, where there is no perpetual difference between the wavelet energies of the images $f(i,j)$ and $\hat{f}(i,j)$, when an observer observe them at d centimeters of observation distance. So measuring the PSNR of $\hat{f}(i,j)$ at D will yield a fairer perceptual evaluation of its image quality.

$\mathcal{C}_w$PSNR algorithm is divided in five steps, which is summarized by the Figure 2.6 and described as follows:
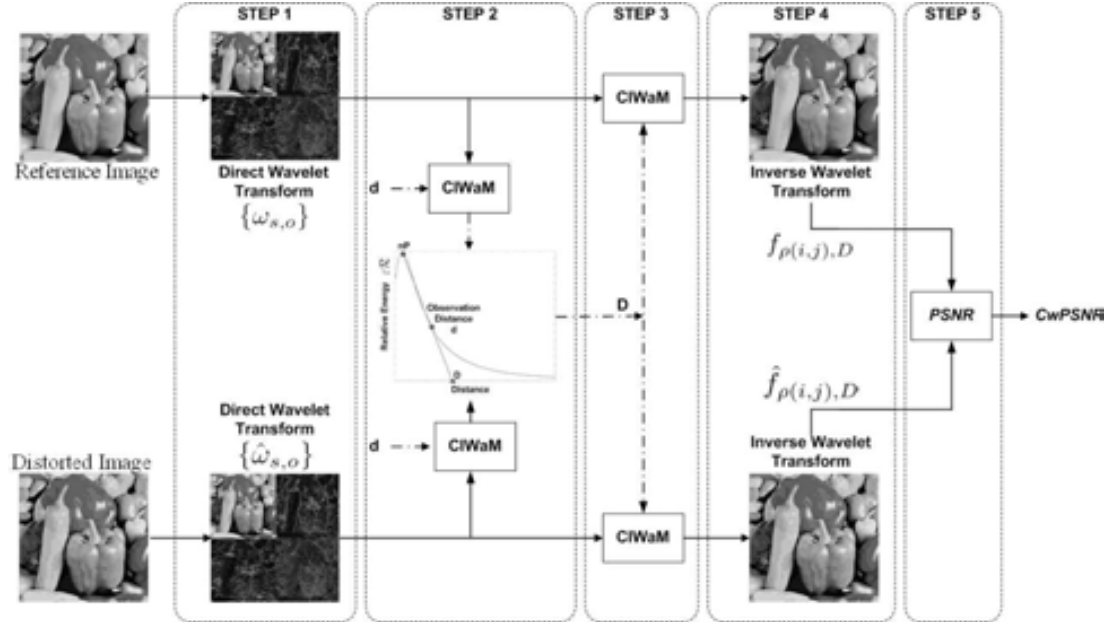
Figure 2.6: Methodology for PSNR weighting by means of CIWaM. Both Reference and Distorted images are wavelet transformed. The distance D where the energy of perceptual images obtained by CIWaM are equal is found. Then, PSNR of perceptual images at D is calculated, obtaining the $\mathcal{C}_w$PSNR metrics.

Step 1: Wavelet Transformation Forward wavelet transform of images $f(i,j)$ and $\hat{f}(i,j)$ is performed using Eq. 3.5, obtaining the sets $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$, respectively. The employed analysis filter is the Daubechies 9-tap/7-tap filter (Table 2.1).

Table 2.1: 9/7 Analysis Filter.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	0.6029490182363579	1.115087052456994
± 1	0.2668641184428723	-0.5912717631142470
± 2	-0.07822326652898785	-0.05754352622849957
± 3	-0.01686411844287495	0.09127176311424948
± 4	0.02674875741080976	

Step 2: Distance D The total energy measure or the *deviation signature*(53) $\bar{\varepsilon}$ is the absolute sum of the wavelet coefficient magnitudes, defined by (61)

$$\bar{\varepsilon} = \sum_{n=1}^{N} \sum_{m=1}^{M} |x(m,n)| \tag{2.10}$$

where $x(m,n)$ is the set of wavelet coefficients, whose energy is being calculated, being m and n the indexes of the coefficients. Basing on the traditional definition of a calorie, the units of $\bar{\varepsilon}$ are wavelet calories (wCal) and can also be defined by Eq. 2.10, since one wCal is the energy needed to increase the absolute magnitude of a wavelet coefficient by one scale.

From wavelet coefficients $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$ the corresponding perceptual wavelet coefficients $\left\{\omega_{s,o;\rho,\tilde{d}}\right\} = \alpha(\nu,r) \cdot \omega_{s,o}$ and $\left\{\hat{\omega}_{s,o;\rho,\tilde{d}}\right\} = \alpha(\nu,r) \cdot \hat{\omega}_{s,o}$ are obtained by applying CIWaM with an observation distance $\tilde{d}$. Therefore, Equation 2.11 expresses the relative wavelet energy ratio $\varepsilon\mathcal{R}\left(\tilde{d}\right)$, which compares how different are the energies of the reference and distorted CIWaM perceptual images, namely ε_ρ and $\hat{\varepsilon}_\rho$ respectively, when these images are watched from a given distance $\tilde{d}$.

$$\varepsilon\mathcal{R}\left(\tilde{d}\right) = 10 \cdot \left| \log_{10} \frac{\varepsilon_\rho\left(\tilde{d}\right)}{\hat{\varepsilon}_\rho\left(\tilde{d}\right)} \right| \tag{2.11}$$

Fig. 2.7(a) shows that distance D is composed by the sum of two distances, $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$. Thereby for the estimation of D, Eq. 2.12, it is necessary to know the observation distance d besides to figure out the $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$ distances. Furthermore Fig. 2.7(b) depicts a chart of $\varepsilon\mathcal{R}$, which sketches both the behavior of the relative energy when $\tilde{d}$ is varied from 0 to ∞ centimeters and the meaning of the distances D, $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$ inside an $\varepsilon\mathcal{R}$ chart.

$$\mathcal{D} = n\mathcal{P} + \varepsilon m\mathcal{L} \tag{2.12}$$
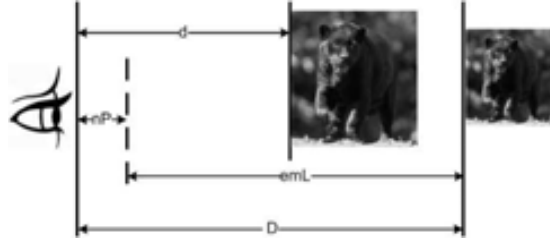
(a) Portrayal of distances employed by the $\mathcal{C}_w$PSNR algorithm.
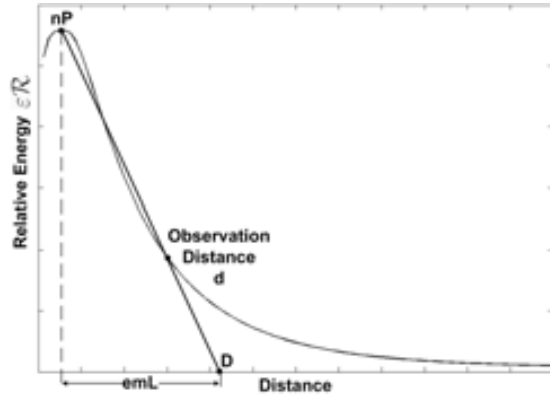
(b) $\varepsilon\mathcal{R}$ Chart.

Figure 2.7: D, $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$ depicted by (a) a graphical representation and (b) inside an $\varepsilon\mathcal{R}$ Chart.

The peak inside an $\varepsilon\mathcal{R}$ chart is $n\mathcal{P}$, which is the distance where the observer is able to better assess the difference between the images $f(i,j)$ and $\hat{f}(i,j)$. From this point $n\mathcal{P}$ the observer starts to perceive fewer the differences, until in ∞ these differences disappear, in practice, this point varies from 15 to 25 centimeters. Our metrics is based on finding an approximation of the distance D where the wavelet

energies are linearly the same, that is, $\varepsilon\mathcal{R}(D) \approx 0$. This is achieved by projecting the points $(n\mathcal{P}, \varepsilon\mathcal{R}(n\mathcal{P}))$ and $(d, \varepsilon\mathcal{R}(d))$ to $(D, 0)$.

Therefore, $\varepsilon m\mathcal{L}$ is the needed length to match the energies from the point where the observer has the best evaluation of the assessed images to D and it is described as follows:

$$\varepsilon m\mathcal{L} = \frac{\varepsilon\mathcal{R}(n\mathcal{P})}{d\varepsilon\mathcal{R} + \varsigma} \qquad (2.13)$$

where $\varepsilon\mathcal{R}(n\mathcal{P})$ is the relative energy at $n\mathcal{P}$ and $d\varepsilon\mathcal{R}$ is the energy loss rate (wCal/cm or wCal/visual degrees) between $(n\mathcal{P}, \varepsilon\mathcal{R}(n\mathcal{P}))$ and $(d, \varepsilon\mathcal{R}(d))$, namely, the negative slope of the line joining these points, expressed as:

$$d\varepsilon\mathcal{R} = \frac{\varepsilon\mathcal{R}(n\mathcal{P}) - \varepsilon\mathcal{R}(d)}{d - n\mathcal{P}} \qquad (2.14)$$

When a lossless compression is performed, consequently $f(i,j) = \hat{f}(i,j)$, hence $d\varepsilon\mathcal{R} = 0$ and $\varepsilon m\mathcal{L} \to \infty$. In order to numerically avoid it, parameter ς is introduced, which is small enough to not affect the estimation of $\varepsilon m\mathcal{L}$ when $d\varepsilon\mathcal{R} \neq 0$, in our MatLab implementation $\varsigma = realmin$.

Step 3: Perceptual Images Obtain the perceptual wavelet coefficients $\{\omega_{s,o;\rho,D}\} = \alpha(\nu, r) \cdot \omega_{s,o}$ and $\{\hat{\omega}_{s,o;\rho,D}\} = \alpha(\nu, r) \cdot \hat{\omega}_{s,o}$ at distance D, using Equation 2.4.

Step 4: Inverse Wavelet Transformation Perform the Inverse Wavelet Transform of $\{\omega_{s,o;\rho,D}\}$ and $\{\hat{\omega}_{s,o;\rho,D}\}$, obtaining the perceptual images $f_{\rho(i,j),D}$ and $\hat{f}_{\rho(i,j),D}$, respectively. The synthesis filter in Table 2.2 is an inverse Daubechies 9-tap/7-tap filter.

Table 2.2: 9/7 Synthesis Filter.

Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1.115087052456994	0.6029490182363579
±1	0.5912717631142470	-0.2668641184428723
±2	-0.05754352622849957	-0.07822326652898785
±3	-0.09127176311424948	0.01686411844287495
±4		0.02674875741080976

Step 5: PSNR **between perceptual images** Calculate the PSNR between perceptual images $f_{\rho(i,j),D}$ and $\hat{f}_{\rho(i,j),D}$ using Eq. 2.2 in order to obtain the CIWaM weighted PSNR i.e. the $\mathcal{C}_w$PSNR.

2.3.2 Discussion

In this section, we analyze the implications of three concepts of the $\mathcal{C}_w$PSNR algorithm; i.e. the $\varepsilon\mathcal{R}(n\mathcal{P})$ value, the distance D and the relation between these two points, and the observation distance d. In brief; first $\varepsilon\mathcal{R}(n\mathcal{P})$ gives a first assessment of the image quality, then, the shorter the distance D the better the predicted perceptual image quality, and finally when the HVS assesses the quality of an image, it depends on, among many parameters, the interaction of the points $n\mathcal{P}$ and d. Thereby the HVS evaluation of image quality is in a dynamic way, taking into account not only the observation distance but also the point where the observer can better perceive the distortions among images. We consider that $\mathcal{C}_w$PSNR is closer to the HVS, because our metrics employs the PSNR indicator for evaluating the images presumably are formed in our brain, that is, $f_p(i,j)$ and $\hat{f}_p(i,j)$ at distance D, maintaining its favorable properties. For visually illustrating some of these characteristics, some images from the *Miscellaneous volume* of the University of Southern California, Signal and Image Processing Institute image database (USC-SIPI image database, Figures A.5 and A.6) are used(2). All the distortions are implemented using JPEG2000 compression.

2.3.2.1   First Sub-indicator: $\varepsilon\mathcal{R}$

When two or more distorted versions of an original image are compared each other, the value of the $\varepsilon\mathcal{R}$ function, at any point, gives an approximation of perceived quality of the distorted image. Thus, when the $\varepsilon\mathcal{R}$ function tends to zero is because the perceived image quality tends to look like the original one, since there are less differences at any distance. Figs. 2.8(a) and 2.9(a), *Splash* and *Baboon* respectively, depict that 40dB images have a lower $\varepsilon\mathcal{R}(n\mathcal{P})$ than 30dB ones.

Thus, in the particular case where different distorted versions of the same original image are analyzed, the $\varepsilon\mathcal{R}(n\mathcal{P})$ value can be considered by itself a perceptual image quality metrics. For instance, in Figure 2.10 when the images *Baboon* and *Splash*, indexed by 1 and 2 respectively, are distorted 30dB and then compared them $\varepsilon\mathcal{R}(n\mathcal{P}_1) < \varepsilon\mathcal{R}(n\mathcal{P}_2)$. This clearly shows that the distorted image *Baboon*$_1$ has better perceptual
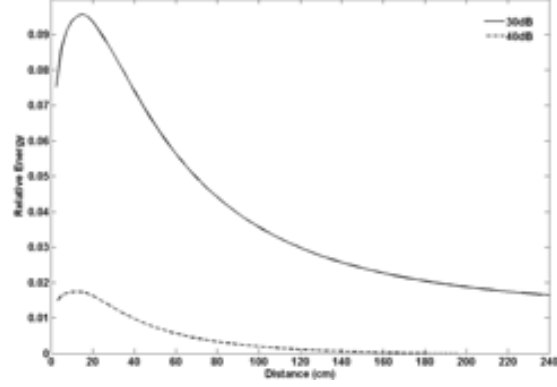
(a) Relative Energy Chart
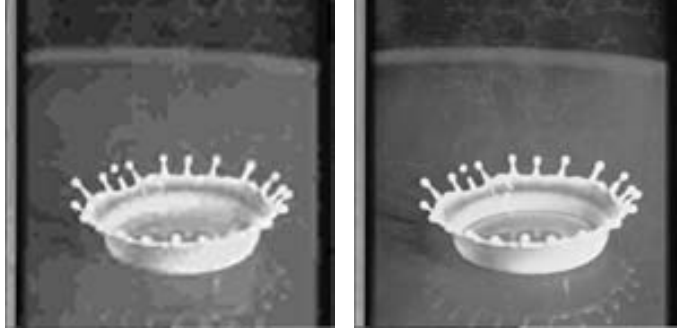
(b) PSNR=30dB · (c) PSNR=40dB

Figure 2.8: Relative Energy Chart of Image *Splash* (a), which is distorted by means of JPEG2000 (b) PSNR=30dB and (c) PSNR=40dB.

image quality than the one of $Splash_2$ and it would not be needful to know either their respective distances D_1 and D_2 or PSNR of perceptual images at those distances. But if D would be computed, $Splash_2$ would need of half of meter after the observation point in order to not perceive the differences between original image and distorted one, while only ten centimeters would be necessary for $Baboon_1$.

2.3.2.2   Second Sub-indicator: D

There are cases where $n\mathcal{P}$ does not give an accurate perceptual measurement. For example, in Fig. 2.11, Relative Energy Ratio of *Sailboat on Lake* image $\varepsilon\mathcal{R}\left(n\mathcal{P}_2\right)$ (index 2, Figure 2.11(c)) is twice $\varepsilon\mathcal{R}\left(n\mathcal{P}_1\right)$ of *Tiffany* image (index 1, Figure 2.11(b)), nevertheless *Sailboat on Lake* image has a better perceptual quality. However when $n\mathcal{P}_1$ is projected along together d_1 to D_1, $D_1 > D_2$, that is, the distorted version of $Tiffany_1$
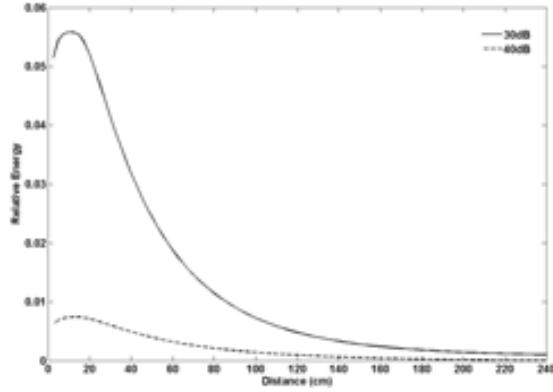
(a) Relative Energy Chart
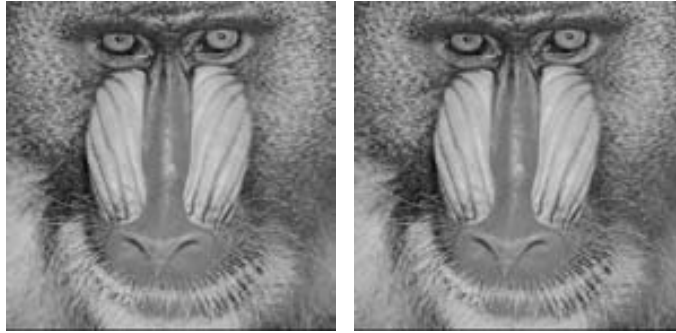
(b) PSNR=30dB   (c) PSNR=40dB

Figure 2.9: Relative Energy Chart of Image *Baboon* (a), which is distorted by means of JPEG2000 (b) PSNR=30dB and (c) PSNR=40dB.

needs 12cm more for matching the perceptual quality regarding its original pair than the *Sailboat on Lake$_2$* image. Moreover when $\mathcal{C}_w$PSNR algorithm is performed, with $d_1 = d_2 = 120cm$ as observation distance, the assessed image quality of *Sailboat on Lake$_2$* image is $36.77dB$ while in $Tiffany_1$ image is $34.82dB$, having approximately 2dB of perceptual difference despite these images were originally distorted $31dB$ by means of JPEG2000. Thus, distance D is a good approximation to an image quality estimator when the degree of distortion of the two images is the same, since the closer to d the better perceptual quality.

2.3.2.3 Third Sub-indicator: $\mathcal{C}_w$PSNR Metrics

However, D cannot be considered as a precise metrics, since in some cases, it predicts the same distance when the perceptual quality of compared images is evidently
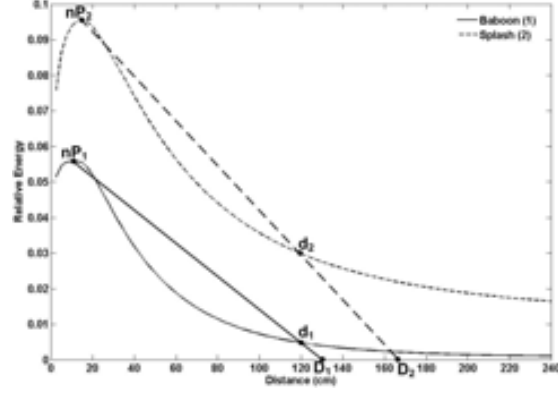
(a) Relative Energy Chart
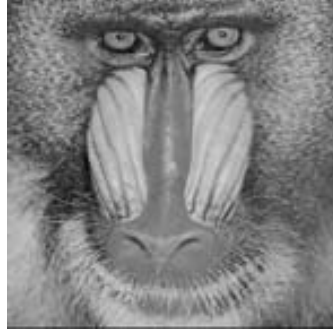
(b) D$_1$=130.36cm (c) D$_2$=167.46cm

Figure 2.10: (a) Relative Energy Chart of Images *Baboon* and *Splash*, both distorted by means of JPEG2000 with PSNR=30dB and Observation distance d=120cm. Perceptual quality $\mathcal{C}_w$PSNR is equal to 36.60dB for (b) and 32.21dB for (c).

different. For instance, in Figures 2.11(c) *Sailboat on Lake$_2$* and 2.12(b) *Splash$_1$* $D_2 = D_1 = 129cm$, but subjective quality of *Splash$_1$* is clearly better than the one of *Sailboat on Lake$_2$*. Thus, even when CIWaM versions of *Splash$_1$* and *Sailboat on Lake$_2$* are calculate at $129cm$, the resultant perceptual images have different objective quality. Hence, $\mathcal{C}_w$PSNR predicts that the error in Figure 2.12(b) is twice less ($\sim 3dB$) than in Figure 2.11(c).

That is why overall $\mathcal{C}_w$PSNR algorithm is the estimation of the objective quality taking into account the set of the interactions of parameters $n\mathcal{P}$, d and D. Figures 2.12 and 2.13 show examples when perceptual quality is the same and their respective points $(n\mathcal{P}, \varepsilon\mathcal{R}(n\mathcal{P}))$ do not correspond. In Figure 2.12, there is a difference of $6cm$ between D_1 and D_2, while in Figure 2.13, there is a small difference between distances
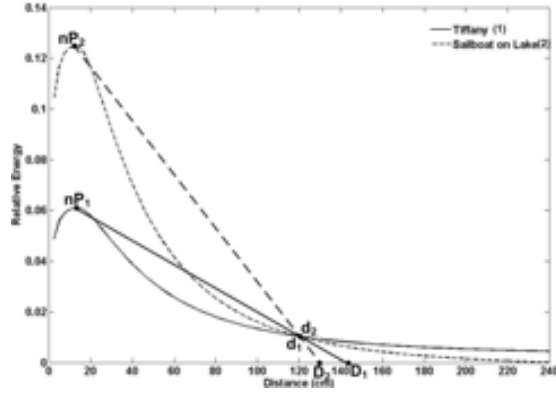
(a) Relative Energy Chart

(b) D_1=141.45cm   (c) D_2=129.67cm

Figure 2.11: (a) Relative Energy Chart of Images *Tiffany* and *Sailboat on Lake* both distorted by means of JPEG2000 with PSNR=31dB and Observation distance d=120cm. Perceptual quality $\mathcal{C}_w$PSNR is equal to 34.82dB for (b) and 36.77dB for (c).

D_1 and D_2.

2.4 Experimental Results

In this section, $\mathcal{C}_w$PSNR performance is assessed by comparing the statistical significance with the psychophysical results obtained by human observers when judging the visual quality of an specific image. These results are expressed in Mean Opinion Scores either differential (DMOS) or not (MOS) of well-known image databases. In this way, perceived image quality predicted by $\mathcal{C}_w$PSNR is tested only for JPEG and JPEG2000 distortions across four image databases:

1. Tampere Image Database (TID2008) of the Tampere University of Technology,
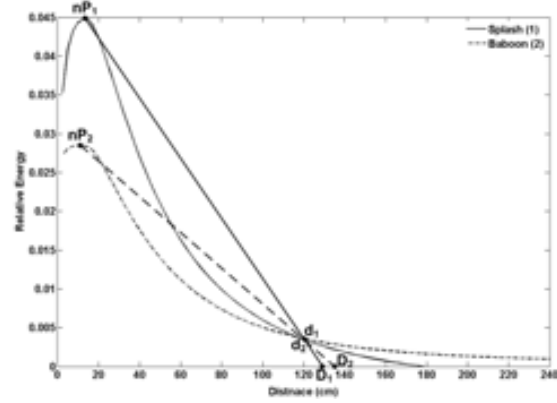
(a) Relative Energy Chart

(b) D_1=129.10cm

(c) D_2=135.89cm

Figure 2.12: (a) Relative Energy Chart of Images *Splash* and *Baboon* both distorted by means of JPEG2000 with $\mathcal{C}_w$PSNR=39.69dB and Observation distance d=120cm. Objective quality PSNR is equal to 35.88dB for (b) and 31.74dB for (c).

presented by Ponomarenko et.al. in (38, 39).

2. Image Database of the Laboratory for Image and Video Engineering (LIVE) of University of Texas at Austin, presented by Sheikh et.al. in (46).

3. Categorical Subjective Image Quality Image Database (CSIQ) of the Oklahoma State University, presented by Larson and Chandler in (22).

4. Image and Video-Communication image Database (IVC) of the Université de Nantes, presented by le Callet and Autrusseau in (23).

TID2008 Database contains 25 original images (Figure A.2), which are distorted by 17 different types of distortions, each distortion has 4 degrees of intensity, that is, 68
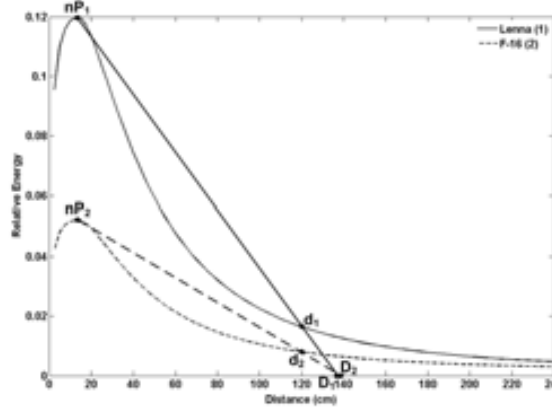
(a) Relative Energy Chart

(b) D_1=137.12cm   (c) D_2=139.49cm

Figure 2.13: a) Relative Energy Chart of Images *Lenna* and *F-16* both distorted by means of JPEG2000 with $\mathcal{C}_w$PSNR=34.75dB and Observation distance d=120cm. Objective quality PSNR is equal to 31.00dB for (b) and 30.87dB for (c).

versions of each source image. TID2008 also supplies subjective ratings by comparing original and distorted images by 654 observers from Italy, Finland and Ukraine. Thus, for JPEG and JPEG2000 compression distortions, there are 200 (25 images $\times$ 2 distortions $\times$ 4 distortion degrees) images in the database. MOS is presented as the global rating.

LIVE Database contains 29 original images (Figure A.3), with 26 to 29 altered versions or each original image. In addition, rating of perceptual quality for each distorted image is given in DMOS values. LIVE uses 5 distortions, including 234 distorted images for JPEG compression degradation and 228 for JPEG2000 one.

CSIQ Database includes 30 original images (Figure A.4), which are distorted by six different types of distortions at 4 or 5 grades. In this way, for JPEG and JPEG2000

compression distortions, CSIQ Database contains 150 distorted versions of these two degradations of the original images. CSIQ Database also has 5000 perceptual evaluations of 25 observers and its assessments are reported in DMOS values.

IVC Database includes 10 original images (Fig. A.1) with 4 different distortions (JPEG, JPEG2000, LAR coding and Blurring) and 5 distortion degrees, that is, there are 50 degraded images by distortion. Perceptual ratings are reported by DMOS.

2.4.1 Performance Measures

Strength of Relationship (SR) is measured by a correlation coefficient. SR means how strong is the tendency of two variables to move in the same (opposite) direction. Pearson Correlation Coefficient (PCC) is the most common measure for predicting SR, when parametric data are used. But in the case of the correlation of non-parametric data the most common indicator is Spearman Rank-Order Correlation Coefficient (SROCC). Results of image quality metrics have no lineal relationship, which is why, it is not convenient to employ PCC, since even PSNR and MSE are the same metrics, PCC calculates different values.

Hence SROCC is a better choice for measuring SR between the opinion of observers and the results of a given metrics. However SROCC is appropriate for testing a null hypothesis, but when this null hypothesis is rejected is difficult to interpret(17). In the other hand, Kendall Rank-Order Correlation Coefficient (KROCC) corrects this problem by reflecting SR between compared variables. Furthermore KROCC estimates how similar are two rank-sets against a same object set. Thus, KROCC is interpreted as the probability to rank in the same order taking into account the number of inversions of pairs of objects for transforming one rank into the other(1). Which is why, $\mathcal{C}_w$PSNR and the rest of metrics are evaluated using KROCC. One of Limitation of KROCC is located in complexity of the algorithm, which takes more computing time than PCC and SROCC, but KROCC can show us an accurate Strength of Relationship between a metric and the opinion of an human observer.

MSE(18), PSNR(18), SSIM(45), MSSIM(54), VSNR(12), VIF(58), VIFP(45), UQI(55), IFC(47), NQM(14), WSNR(25) and SNR are compared against the performance of $\mathcal{C}_w$PSNR for JPEG and JPEG2000 compression distortions. We chose for evaluating these assessments the implementation provided in (21), since it is based on the parameters proposed by the author of each indicator.

$\mathcal{C}_w$PSNR is implemented assuming the following features:

- Observation Distance, d=8H, where H is the height of a 512×512 image.

- 19" LCD monitor with horizontal resolution of 1280 pixels and 1024 pixels of vertical resolution.

- Gamma correction, $\gamma = 2.2$

- Wavelet Transform, set of wavelet planes ω with $n = 3$, Eq. 3.5.

2.4.2 Overall Performance

Table 2.3 shows the performance of $\mathcal{C}_w$PSNR and the other twelve image quality assessments across the set of images from TID2008, LIVE, CSIQ and IVC image databases employing KROCC for testing the distortion produced by a JPEG compression.

Table 2.3: KROCC of $\mathcal{C}_w$PSNR and other quality assessment algorithms on multiple image databases using JPEG distortion. The higher the KROCC the more accurate image assessment. Bold and italicized entries represent the best and the second-best performers in the database, respectively. The last column shows the KROCC average of all image databases.

Metrics	Image Database				
	TID2008	LIVE	CSIQ	IVC	All
Images	100	234	150	50	534
MSE	0.7308	0.7816	0.6961	0.5187	0.6818
PSNR	0.7308	0.7816	0.6961	0.5187	0.6818
SSIM	0.7334	0.8287	0.7529	0.6303	0.7363
MSSIM	*0.7580*	*0.8435*	0.8097	0.7797	*0.7977*
VSNR	0.7344	0.8149	0.7117	0.5827	0.7109
VIF	0.7195	0.8268	*0.8287*	0.7911	0.7915
VIFP	0.7004	0.8140	0.8188	0.6763	0.7524
UQI	0.5445	0.7718	0.6990	0.6254	0.6602
IFC	0.5909	0.7767	0.7644	*0.8158*	0.7369
NQM	0.7142	0.8269	0.7907	0.6664	0.7495
WSNR	0.7300	0.8181	0.8020	0.6959	0.7615
SNR	0.6035	0.7735	0.6942	0.4481	0.6298
$\mathcal{C}_w$PSNR	**0.7616**	**0.8457**	**0.8473**	**0.8335**	**0.8220**

Table 2.3 also shows an average performances for the 534 images of the cited image databases. Bold and Italicized represent the best and the second best performance

assessment, respectively. It is appropriate to say that $\mathcal{C}_w$PSNR is the best performer both in each image database and average of them. MSSIM is the second best-ranked metrics not only in all databases but also on the average, except for the CSIQ database, where VIF has this place. $\mathcal{C}_w$PSNR is better 0.0243 than MSSIM and improves the performance of PSNR or MSE by 0.1402 for JPEG compression degradation.

While Table 2.4 shows the performance of $\mathcal{C}_w$PSNR for JPEG2000 compression distortion across all image databases comparing the same twelve metrics presented in Table 2.3.

Table 2.4: KROCC of $\mathcal{C}_w$PSNR and other quality assessment algorithms on multiple image databases using JPEG2000 distortion. The higher the KROCC the more accurate image assessment. Bold and italicized entries represent the best and the second-best performers in the database, respectively. The last column shows the KROCC average of all image databases.

Metrics	Image Database				
	TID2008	LIVE	CSIQ	IVC	All
Images	100	228	150	50	528
MSE	0.6382	0.8249	0.7708	0.7262	0.7400
PSNR	0.6382	0.8249	0.7708	0.7262	0.7400
SSIM	0.8573	0.8597	0.7592	0.6916	0.7919
MSSIM	*0.8656*	*0.8818*	0.8335	*0.7821*	*0.8408*
VSNR	0.8042	0.8472	0.7117	0.6949	0.7645
VIF	0.8515	0.8590	0.8301	0.7903	0.8327
VIFP	0.8215	0.8547	*0.8447*	0.7229	0.8110
UQI	0.7415	0.7893	0.6995	0.6061	0.6602
IFC	0.7905	0.7936	0.7667	0.7788	0.7824
NQM	0.8034	0.8574	0.8242	0.6801	0.7913
WSNR	0.8152	0.8402	0.8362	0.7656	0.8143
SNR	0.5767	0.8055	0.7665	0.6538	0.7006
$\mathcal{C}_w$PSNR	**0.8718**	**0.8837**	**0.8682**	**0.7981**	**0.8555**

Thus, for JPEG2000 compression distortion, $\mathcal{C}_w$PSNR is also the best metrics for each database. $\mathcal{C}_w$PSNR gets its better results when correlation is 0.8837 for a corpus of 228 images of the LIVE database. On the average, our algorithm is also the best performing metrics with a SR, using KROCC, of 0.8555. For this distortion, MSSIM is also the second best indicator for TID2008, LIVE and IVC image databases in addition to the average. For CSIQ image database VIFP occupies this place. In this way the

results of MSSIM correlates with the opinion of observers 0.0143 less than the ones of $\mathcal{C}_w$PSNR. Furthermore $\mathcal{C}_w$PSNR improves 0.1155 the perceptual functioning of PSNR when this metrics compares perceptual images in a dynamic way.

In summary, $\mathcal{C}_w$PSNR is the best performing algorithm for JPEG and JPEG2000 compression distortions, that is, for image compression algorithms, which use either Discrete Cosine Transform or Wavelet Transform as method of pixel transformation in samples for the quantization process(51, pg. 14).

2.5   Conclusions

$\mathcal{C}_w$PSNR is a new metric for full-reference image quality based on perceptual weighting of PSNR by using a perceptual low-level model of the Human Visual System (CIWaM model). The proposed $\mathcal{C}_w$PSNR metrics is based on three concepts.

The $\mathcal{C}_w$PSNR assessment was tested in four well-known image databases such as TID2008, LIVE, CSIQ and IVC. It is the best-ranked image quality method in these databases for JPEG and JPEG2000 distortions when compared to several state-of-the-art metrics. Concretely, it is 2.5% and 1.5% better that MSSIM (the second best performing method) for JPEG and JPEG2000 distortions, respectively. $\mathcal{C}_w$PSNR significantly improves the correlation of PSNR with perceived image quality. On average, when $\mathcal{C}_w$PSNR is applied on the same distortion, it improves the results obtained by PSNR and MSE by 14% and 11.5%,respectively.

Chapter 3

Image Coder Based on Hilbert Scanning of Embedded quadTrees

3.1 Introduction

One of the biggest challenges of image compressors is the massive storage and ordering of data coordinates. In some algorithms, like EZW (44), SPIHT (41) and SPECK (34, 35, 36), the execution path defines the correct order of the coefficients by comparison of its branching points (52). Our coder makes use of a Hilbert Scanning, which exploits the self-similarity of pixels. Since the space-filling path of Hilbert's fractal is known *a priori*, it implicitly defines the coefficient coordinates. Hence, the decoder only needs the coefficient magnitudes in order to recover them. Furthermore, applying a Hilbert Scanning to Wavelet Transform coefficients takes the advantage of the self-similarity of neighbor pixels, helping to exploit their redundancy and to develop an optimal progressive transmission coder. In this way, at any step of the decoding process the quality of the recovered image is the best that can be achieved for the number of bits processed by the decoder up to that moment.

Figure 3.1 shows the block diagram of image compressor based on *Hi*lbert *S*canning of *E*mbedded quad*T*rees (H*i*-SET) for the encoding and decoding processes. The green blocks in 3.2 indicate the position of these latter processes inside the proposed perceptual compression system. The source image data may contain one or more components (up to 2^3 in the case of H*i*-SET). Each component is decomposed by a discrete wavelet transform into a set of wavelet planes of different spatial frequencies and orientations.

Figure 3.1: General block diagram of a generic compressor that uses Hi-SET for encoding and decoding.

Figure 3.2: General block diagram for the proposed perceptual image compression system. The Hi-SET compression algorithm is indicated by the green blocks.

Wavelet plane coefficients are quantized with a dead-zone uniform scalar quantizer (SQ) for reducing the precision of data in order to make them more compressible. This Quantization block introduces distortion and it is only employed for lossy compression. In the following step, Hi-SET algorithm encodes the entropy among quantized coefficients, obtaining an output bitstream. The decompression process is the inverse of the compression one: the bitstream is entropy decoded by Hi-SET, dequantized by SQ and an inverse discrete wavelet transform is performed, getting as a result the reconstructed image data.

3.2 Component Transformations

Image compression algorithms are usually used in color images. These images can be numerically represented in several color spaces, such as RGB, YC_bC_r, YCM, and HSB, being RGB the most commonly used.

In this way, an RGB color image is decomposed into three components, namely Red, Green, and Blue color components. Figure 3.3 depicts that when Hi-SET performs a
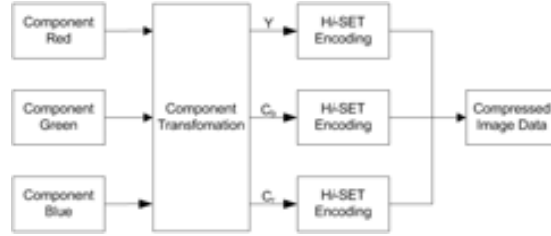
Figure 3.3: Hi-SET multiple component encoder.

color compression, a complete encoding is developed at each color layer. R, G and B color components are statistically more dependent than Y, C_r and C_b, thus the chrominance channels can be processed independently at lower resolution than luminance one in order to achieve better compression rates (59).

Hi-SET supports both Reversible Component Transformation (RCT) and Irreversible Component Transformation(ICT) (10, Annex G). For lossy coding is employed an ICT, which makes use of the the 9/7 irreversible wavelet transform, forward and inverse are calculated by the Equation 3.1 and 3.2, respectively (48, 51).

$$
\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.16875 & -0.33126 & 0.5 \\ 0.5 & -0.41869 & -0.08131 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.1}
$$

$$
\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.0 & 0 & 0.114 \\ 1.0 & -0.34413 & -0.71414 \\ 1.0 & 1.772 & 0 \end{bmatrix} \begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix}. \tag{3.2}
$$

RCT is used for lossy and lossless codding, together with the 5/3 reversible wavelet transform. The forward RCT transformations is achieved by means of the Equation 3.3 while the inverse by the Equation 3.4.

$$
\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} \left\lfloor \frac{R+2G+B}{4} \right\rfloor \\ R - G \\ B - G \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.3}
$$

$$
\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} Y - \left\lfloor \frac{C_r+C_b}{4} \right\rfloor \\ C_b + G \\ C_r + G \end{bmatrix} \begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} \tag{3.4}
$$

3.3 Wavelet Transform

The input image $\mathcal{I}$ used by Hi-SET is separated into different spatial frequencies and orientation using a multiresolution discrete wavelet decomposition (DWT) either re-

versible or irreversible (3, 49), by each component. Thus $\mathfrak{I}$ is decomposed into a set of wavelet planes ω of different spatial frequencies, where each wavelet plane contains details at different spatial resolutions and it is described by:

$$DWT\{\mathfrak{I}\} = \sum_{s=1}^{n} \sum_{o=v,h,d} \omega_s^o + c_n \qquad (3.5)$$

where $s = 1, \ldots, n$, n the number of wavelet planes and c_n the residual plane. $o = v, h, d$ represents the spatial orientation either vertical, horizontal or diagonal, respectively.

The DWT is performed in order to filter each row and column of $\mathfrak{I}$ with a high-pass and low-pass filter. Since this procedure derives in double the number of samples, the output from each filter is downsampled by 2, thus the sample rate remains constant. It is not important if the rows or the columns of the component matrix are filtered first, because the resulting DWT is the same. The reversible transformation is implemented by means of 5/3 filter. The analysis and the respective synthesis filter of coefficients are described by the Table 3.1. The irreversible transform is implemented by means of the 9/7 filter and Table 3.2 illustrates its analysis and synthesis filters.

Table 3.1: 5/3 Analysis and Synthesis Filter.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	6/8	1
±1	2/8	-1/2
±2	-1/8	
Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1	6/8
±1	1/2	-2/8
±2		-1/8

The number of filtering stages, i.e. the number n of wavelet planes, depends on its implementation. Nevertheless, taking into account the trade-off between image quality and compression ratio, some authors report that the best results are obtained with $n = 3$ (41).

Figure 3.4 depicts the DWT generation of the Y component the image *Peppers* with $n = 3$.

Table 3.2: 9/7 Analysis and Synthesis Filter.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	0.6029490182363579	1.115087052456994
±1	0.2668641184428723	-0.5912717631142470
±2	-0.07822326652898785	-0.05754352622849957
±3	-0.01686411844287495	0.09127176311424948
±4	0.02674875741080976	
Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1.115087052456994	0.6029490182363579
±1	0.5912717631142470	-0.2668641184428723
±2	-0.05754352622849957	-0.07822326652898785
±3	-0.09127176311424948	0.01686411844287495
±4		0.02674875741080976
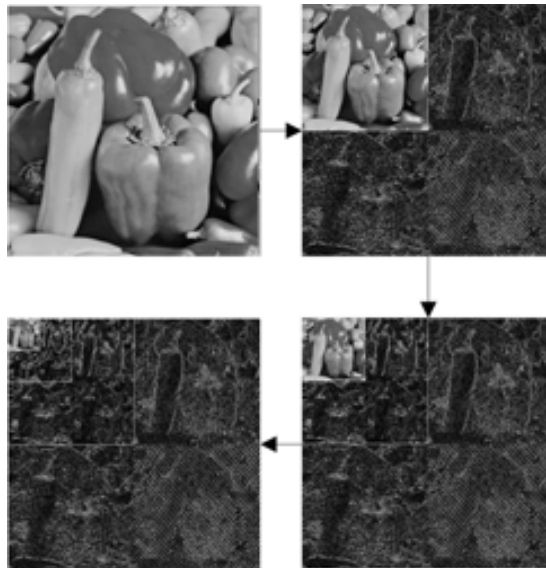
Figure 3.4: Three-level wavelet decomposition of the *Peppers* image.

3.4 Dead-zone Uniform Scalar Quantizer

Marcellin et.al. summarize in (24), among other, the uniform scalar quantizer. This quantizer is described as a function that maps each element of a subset of the real numbers into a particular value, which ensures that more zeros result. This way, quantization values are uniformly spaced by step size Δ except for the interval containing the zero value, which is called the dead-zone, that extends from $-\Delta$ to $+\Delta$. Thus, a

dead-zone means that the quantization range around 0 is 2Δ.

Taking a given wavelet plane ω_s^o , a particular quantizer step size Δ_s^o is used to quantize all the coefficients in that spatial frequency s and orientation o. Hence a particular quantized index is defined as:

$$q = sign(y) \left\lfloor \frac{|y|}{\Delta_s^o} \right\rfloor \tag{3.6}$$

where y is the input to the quantizer (i.e., the original wavelet coefficient value), $sign(y)$ denotes the sign of y and q is the resulting quantized index. Figure 3.5 illustrates such a quantizer with step size Δ, here vertical lines indicate the endpoints of the quantization intervals and heavy dots represent reconstruction values.

The inverse quantizer or the reconstructed $\hat{y}$ is given by

$$\hat{y} = \begin{cases} (q + \delta)\Delta_s^o, & q > 0 \\ (q - \delta)\Delta_s^o & q < 0 \\ 0, & q = 0 \end{cases} \tag{3.7}$$

where δ is a parameter often set to place the reconstruction value at the centroid of the quantization interval and varies form 0 to 1.
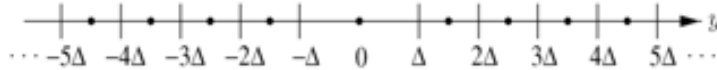
Figure 3.5: Dead-zone uniform scalar quantizer with step size Δ: vertical lines indicate the endpoints of the quantization intervals and heavy dots represent reconstruction values.

The International Organization for Standardization recommends to adopt the midpoint reconstruction value, setting $\delta = 0.5$ (10). Experience indicates that some small improvements can be obtained by selecting a slightly smaller value. Pearlman and Said in (34) suggest $\delta = 0.375$, especially for higher frequency subbands (e.g. high frequency wavelet planes). It is important to realize that when $-\Delta < y < \Delta$, the quantizer level and reconstruction value are both 0. Since it is known that many coefficients in a wavelet transform are close to zero (usually those of higher frequencies), it implies that they are on the dead-zone, thus, the quantizer sets them to q=0.

Once a wavelet plane ω_s^o is quantized, it is further losslessly encoded, since the image compression degradations are only induced by the Quantization process.

3.5 The H*i*-SET Algorithm

3.5.1 Startup Considerations

3.5.1.1 Hilbert space-filling Curve

The Hilbert curve is an iterated function that is represented by a parallel rewriting system, concretely a L-system. In general, a L-system structure is a tuple of four elements:

1. *Alphabet*: the variables or symbols to be replaced.

2. *Constants*: set of symbols that remain fixed.

3. *Axiom* or *initiator*: the initial state of the system.

4. *Production rules*: how variables are replaced.

In order to describe the Hilbert curve alphabet let us denote the upper left, lower left, lower right, and upper right quadrants as $\mathcal{W}$, $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively, and the variables as $\mathcal{U}$ (*up*, $\mathcal{W} \to \mathcal{X} \to \mathcal{Y} \to \mathcal{Z}$), $\mathcal{L}$ (*left*, $\mathcal{W} \to \mathcal{Z} \to \mathcal{Y} \to \mathcal{X}$), $\mathcal{R}$ (*right*, $\mathcal{Z} \to \mathcal{W} \to \mathcal{X} \to \mathcal{Y}$), and $\mathcal{D}$ (*down*, $\mathcal{X} \to \mathcal{W} \to \mathcal{Z} \to \mathcal{Y}$). Where $\to$ indicates a movement from a certain quadrant to another. Each variable represents not only a trajectory followed through the quadrants, but also a set of 4^m transformed pixels in m level.

The structure of our Hilbert Curve representation does not need fixed symbols, since it is just a linear indexing of pixels.

(a)   (b)
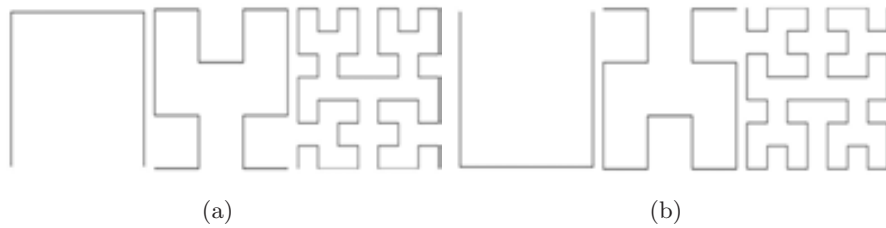
Figure 3.6: First three levels of a Hilbert Fractal Curve. (a) Axiom = $\mathcal{D}$ proposed by David Hilbert in (16). (b) Axiom = $\mathcal{U}$ employed for this work.

The original work by David Hilbert(16) proposes an axiom with a $\mathcal{D}$ trajectory (Figure 3.6(a)), while we propose to start with an $\mathcal{U}$ trajectory (Figure 3.6(b)). Our proposal is based on the most of the image energy is concentrated where the higher

subbands with lower frequencies are, namely at the upper-left quadrant. The first three levels are portrayed in left-to-right order by Figure 3.6.

The production rules of the Hilbert Curve are defined by

- $\mathcal{U}$ is changed by the string $\mathcal{LUUR}$

- $\mathcal{L}$ by $\mathcal{ULLD}$

- $\mathcal{R}$ by $\mathcal{DRRU}$

- $\mathcal{D}$ by $\mathcal{RDDL}$.

In this way high order curves are recursively generated replacing each former level curve with the four later level curves.

The Hilbert Curve has the property of remaining in an area as long as possible before moving to a neighboring spatial region. Hence, correlation between neighbor pixels is maximized, which is an important property in image compression processes. The higher the correlation at the preprocessing, the more efficient the data compression.

3.5.1.2 Linear Indexing

A linear indexing is developed in order to store the coefficient matrix into a vector. Let us define the Wavelet Transform coefficient matrix as $\mathcal{H}$ and the interleaved resultant vector as $\overrightarrow{\mathcal{H}}$, being $2^\gamma \times 2^\gamma$ be the size of $\mathcal{H}$ and 4^γ the size of $\overrightarrow{\mathcal{H}}$, where γ is the Hilbert curve level. Algorithm 1 generates a Hilbert mapping matrix θ with level γ, expressing each curve as four consecutive indexes. The level γ of θ is acquired concatenating four different θ transformations in the previous level $\gamma - 1$. Algorithm 1 generates the Hilbert mapping matrix θ, where $\overrightarrow{\beta}$ refers a 180 degree rotation of β and β^T is the linear algebraic transpose of β. Figure 3.7(b) shows an example of the mapping matrix θ at level $\gamma = 3$. Thus, each wavelet coefficient at $\mathcal{H}_{(i,j)}$ is stored and ordered at $\overrightarrow{\mathcal{H}}_{\theta_{(i,j)}}$, being $\theta_{(i,j)}$ the location index of it into $\overrightarrow{\mathcal{H}}$.

3.5.1.3 Significance Test

A significance test is defined as the trial of whether one coefficient from a set of coefficients achieves a predefined significance criterion. A coefficient that fulfills the criterion is considered *significant*, otherwise it is considered *insignificant*. The significance test

---

Algorithm 1: Function to generate Hilbert mapping matrix θ of size $2^\gamma \times 2^\gamma$.

Input: γ

Output: θ

1 **if** $\gamma = 1$ **then**

2 $\quad$ $\theta = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$

3 **else**

4 $\quad$ $\beta = $ **Algorithm 1** $(\gamma - 1)$

5 $\quad$ $\theta = \begin{bmatrix} \beta^T & (\widetilde{\beta})^T + (3 \times 4^{\gamma-1}) \\ \beta + 4^{\gamma-1} & \beta + (2 \times 4^{\gamma-1}) \end{bmatrix}$

---

also defines how these subsets are formed and what coefficients are considered significant.

With the aim of recovering the original image at different qualities and compression ratios, it is not needed to sort and store all the coefficients $\overrightarrow{\mathcal{H}}$ but just a subset of them: the subset of significant coefficients. Those coefficients $\overrightarrow{\mathcal{H}}_i$ such that $2^{thr} \leq |\overrightarrow{\mathcal{H}}_i|$ are called *significant* otherwise they are called *insignificant*. The smaller the *thr*, the better the final image quality and the lower the compression ratio.

Let us define a bit-plane as the subset of coefficients $\mathcal{S}_o$ such that $2^{thr} \leq |\mathcal{S}_o| < 2^{thr+1}$. The significance of a given subset $\mathcal{S}_o$ amongst a particular bit-plane is store at $\widehat{\mathcal{H}_{sig}}$ and is defined as:

$$\widehat{\mathcal{H}_{sig}} = \begin{cases} 1, & 2^{thr} \leq |\mathcal{S}_o| < 2^{thr+1} \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

Algorithm 2 shows how a set $\mathcal{S}_o$ is divided into four equal parts (line 6) and how the significance test (lines 7-12) is performed, resulting in four subsets ($\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$) with their respective significance stored at the end of $\widehat{\mathcal{H}_{sig}}$. The subsets $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ are 2×1 cell arrays. The fist cell of each array contains one of the four subsets extracted from $\mathcal{S}_o$, $\mathcal{S}_i(1)$ and the second one stores its respective significance test result, $\mathcal{S}_i(2)$.

3.5.2 Coding Algorithm

Similarly to SPIHT and SPECK (34, 35), Hi-SET considers three coding passes: Initialization, Sorting and Refinement, which are described in the next subsections. SPIHT uses three ordered lists, namely the *list of significant sets* (*LIS*), the *list of insignificant*

Algorithm 2: Subset Significance Test.

Data: $\mathcal{S}_o$, thr
Result: $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$ and $\widehat{\mathcal{H}_{sig}}$

1 $\gamma = \log_4(length\ of\ \mathcal{S}_o)$
2 The cell 1 of the subsets $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$ and $\mathcal{S}_4$ is declared with $4^{\gamma-1}$ elements, while the cell 2 with just one element.
3 $i = 1$
4 $\widehat{\mathcal{H}_{sig}}$ is emptied.
5 **for** *j=1* **to** 4^γ **do**
6 $\quad$ Store $\mathcal{S}_o\left[from\ j\ to\ \left(i \times 4^{\gamma-1}\right)\right]$ into $\mathcal{S}_i(1)$.
7 $\quad$ **if** $2^{thr} \leq \max|\mathcal{S}_i(1)| < 2^{thr+1}$ **then**
8 $\quad\quad$ $\mathcal{S}_i(2) = 1$
9 $\quad\quad$ Add **1** at the end of the $\widehat{\mathcal{H}_{sig}}$.
10 $\quad$ **else**
11 $\quad\quad$ $\mathcal{S}_i(2) = 0$
12 $\quad\quad$ Add **0** at the end of the $\widehat{\mathcal{H}_{sig}}$.
13 $\quad$ i and j are incremented by 1 and $4^{\gamma-1}$, respectively.

---

pixels (LIP), and the *list of significant pixels* (LSP). The latter represents just the individual coefficients, which are considered the most important ones. SPECK employs two of these lists, the LIS and the LSP. In contrast, Hi-SET makes use of only one ordered list, the LSP.

Using a single LSP place extra load on the memory requirements of the coder, because the total number of significant pixels remains the same even if the coding process is working in insignificant branches. That is why we employ spare lists, storing significant pixels in several sub-lists. This smaller lists have the same length than significant coefficients found in the processed branch. With the purpose of speeding up the coding process, Hi-SET uses not only spare lists, but also spare cell arrays, both are denoted by an apostrophe, LSP', $\widehat{\mathcal{H}}'$ or $\mathcal{S}_1'$, for instance.

3.5.2.1 Initialization Pass

The first step in this stage is to define threshold thr as

$$thr = \left\lfloor \log_2\left(\max\left\{\overrightarrow{\mathcal{H}}\right\}\right) \right\rfloor , \tag{3.9}$$

that is, thr is the maximum integer power of two not exceeding the maximum value of $\overrightarrow{\mathcal{H}}$.

The second step is to apply Algorithm 2 with *thr* and $\overrightarrow{\mathcal{H}}$ as input data, which divides $\overrightarrow{\mathcal{H}}$ into four subsets of $4^{\gamma-1}$ coefficients and adds their significance bits at the end of $\widehat{\mathcal{H}}$.

3.5.2.2 Sorting Pass

Algorithm 3 shows a simplified version of the classification or sorting step of the H*i*-SET Coder. The H*i*-SET sorting pass exploits the recursion of fractals. If a quadtree branch is *significant* it moves forward until finding an individual pixel, otherwise the algorithm stops and codes the entire branch as *insignificant*.

Algorithm 3: Sorting Pass

    Data: $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, $\mathcal{S}_4$, *thr* and γ
    Result: *LSP* and $\widehat{\mathcal{H}}$

1   *LSP* and $\widehat{\mathcal{H}}$ are emptied.
2   **if** $\gamma = 0$ **then**
3   **for** $i = 4$ **to** 1 **do**
4   **if** $\mathcal{S}_i(2)$ *is significant* **then**
5   Add $\mathcal{S}_i(1)$ at the beginning of the *LSP*.
6   **if** $\mathcal{S}_i(1)$ *is positive* **then**
7   Add **0** at the beginning of the $\widehat{\mathcal{H}}$.
8   **else**
9   Add **1** at the beginning of the $\widehat{\mathcal{H}}$.

10   **else**
11   **for** $i=1$ **to** 4 **do**
12   **if** $\mathcal{S}_i(2)$ *is significant* **then**
13   Call **Algorithm 2** with $\mathcal{S}_i(1)$ and *thr* as input data and Store the results into $\mathcal{S}'_1$, $\mathcal{S}'_2$, $\mathcal{S}'_3$, $\mathcal{S}'_4$ and $\widehat{\mathcal{H}}'$.
14   Add $\widehat{\mathcal{H}}'$ at the end of the $\widehat{\mathcal{H}}$.
15   Call **Algorithm 3** with $\mathcal{S}'_1$, $\mathcal{S}'_2$, $\mathcal{S}'_3$, $\mathcal{S}'_4$, *thr* and $\gamma - 1$ as input data and Store the results into $\widehat{\mathcal{H}}'$ and *LSP'*.
16   Add $\widehat{\mathcal{H}}'$ at the end of the $\widehat{\mathcal{H}}$.
17   Add *LSP'* at the end of the *LSP*.

Algorithm 3 is divided into two parts: Sign Coding (lines 2 to 9) and Branch Significance Coding (lines 11 to 16). The algorithm performs *the Sign Coding* by decomposing a given quadtree branch up to level $\gamma = 0$, i.e. the branch is represented by only 4 coefficients with at least one of them being *significant*. The initial value of γ is $\log_4(length\ of\ \overrightarrow{\mathcal{H}}) - 1$. Only the sign of the significant coefficients is coded, 0 for positives and 1 for negatives. Also each significant coefficient is added into a spare *LSP* or *LSP'*.

The Branch Significance Coding calls the Algorithm 2 in order to quarter a branch in addition to call recursively an entire sorting pass at level $\gamma - 1$ up to reach the elemental level when $\gamma = 0$. The Significance Test results of a current branch (obtained by the Algorithm 2) and the ones of next branches (acquired by Algorithm 3, denoted as $\widehat{\mathcal{H}}'$) are added at the end of $\widehat{\mathcal{H}}$. Also, all the significant coefficients found in previous branches (all the lists LSP') are added at the end of the LSP. This processes is repeated for all four subsets of $\overrightarrow{\mathcal{H}}$.

3.5.2.3 Refinement Pass

At the end of $\widehat{\mathcal{H}}$, the $(thr - 1)$-th most significant bits of each ordered entry of the LSP, including those entries added in the last sorting pass, are added. Then, thr is decremented and another Sorting Pass is performed. The Sorting and Refinement steps are repeated up to $thr = 1$.

The decoder employs the same mechanism as the encoder, since it knows the fractal applied to the original image. When the bitstream $\widehat{\mathcal{H}}$ is received, by itself describes the significance of every variable of the fractal. Then with these bits, the decoder is able to reconstruct both partially and completely, the same fractal structure of the original image, refining the pixels progressively as the algorithm proceeds.

3.5.3 A Simple Example

In order to highlight the operations employed by Hi-SET, a simple example is shown.

The wavelet transform coefficient matrix $\mathcal{H}$ of an 8×8 pixels image is depicted in Figure 3.7(a), which is a three scale ($n = 3$) transformation, which implies $\gamma = 3$. The indexed vector $\overrightarrow{\mathcal{H}}$ (Figure 3.7(c)) is acquired interleaving $\mathcal{H}$ with a three-level matrix θ (Figure 3.7(b)).

Table 3.3 shows the entire process up to the first bit-plane. The eleven steps in Table 3.3 represent the three passes of the scheme. *Initialization Pass* is described by steps 1 and 2, *Sorting Pass* by steps 3-10, while step 11 illustrates *Refinement Pass*. Figure 3.8 depicts the fractal partitioning diagram of the first bit-plane encoding.

The following remarks refer to steps of the Table 3.3:

Step 1 The largest coefficient magnitude inside $\overrightarrow{\mathcal{H}}$ is 63, thus the initial threshold, defined by the Equation 3.9, is $thr = 5$ (i.e. $2^5 = 32$). It implies that the first

bit-plane is placed at $(-64, -32]$ and $[32, 64)$. Both LSP and $\widehat{\mathcal{H}}$ are emptied and level $\gamma = 3$ is adopted by the axiom $(3\mathcal{U})$.

Step 2 Using the production rules, a $3\mathcal{U}$ curve changes to $2\mathcal{L}\mathcal{U}\mathcal{U}\mathcal{R}$. At the first bit-plane, the $2\mathcal{L}$ and $2\mathcal{U}$ curves are subsets of 4^2 pixels, where at least one coefficient is significant, in this case 63, -34 and 49 for $2\mathcal{L}$ (e.g. upper left quadrant) and 47 for $2\mathcal{U}$ (lower left quadrant). The other two curves, $2\mathcal{U}$ and $2\mathcal{R}$, have only insignificant coefficients. Therefore the significance of these curves is 1100, which is placed at $\widehat{\mathcal{H}}$.

Step 3 Using the production rules, a $2\mathcal{L}$ curve changes to $1\mathcal{U}\mathcal{L}\mathcal{L}\mathcal{D}$. At the first bit-plane, the $1\mathcal{U}$ and $1\mathcal{L}$ curves are subsets where at least one pixel is significant, in this case 63 and -34 for $1\mathcal{U}$ and 49 for $1\mathcal{L}$. The other two curves, $1\mathcal{L}$ and $1\mathcal{D}$,

(a) Matrix $\mathcal{H}$ (b) Matrix θ
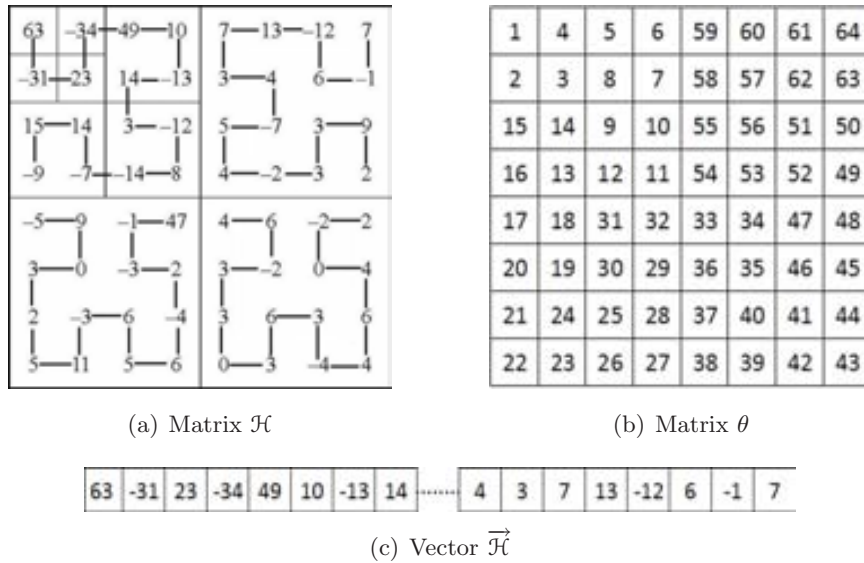
(c) Vector $\overrightarrow{\mathcal{H}}$

Figure 3.7: Example of Hilbert indexing of an 8×8 pixels image. (a) Three-scale wavelet transform matrix $\mathcal{H}$ with its Hilbert path. (b) Hilbert Indexing matrix θ when $\gamma = 3$. (c) Interleaved resultant vector $\overrightarrow{\mathcal{H}}$.
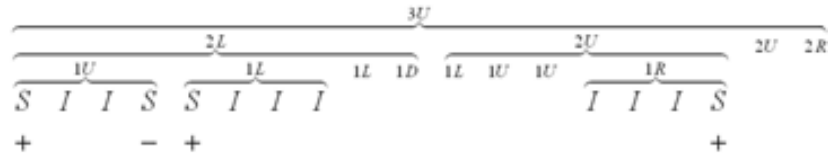
Figure 3.8: Fractal partitioning diagram of the first bit-plane encoding, using H*i*-SET scheme.

Table 3.3: The First bit-plane encoding using Hi-SET scheme. $\mathcal{H}$, θ and $\overrightarrow{\mathcal{H}}$ are taken from Figure 3.7, with initial threshold $thr = 5$.

Step	Former Curve	Current Curve(s)	Bitstream $\widehat{\mathcal{H}}$	Decoded LSP
1		$3\mathcal{U}$		
2	$3\mathcal{U}$	$2\mathcal{L}\mathcal{U}\mathcal{U}\mathcal{R}$	1100	
3	$2\mathcal{L}$	$1\mathcal{U}\mathcal{L}\mathcal{L}\mathcal{D}$	1100	
4	$1\mathcal{U}$	$\mathcal{S}\mathcal{I}\mathcal{I}\mathcal{S}$	1001	
5	$sign$	$+-$	01	$+32$ -32
6	$1\mathcal{L}$	$\mathcal{S}\mathcal{I}\mathcal{I}\mathcal{I}$	1000	
7	$sign$	$+$	0	$+32$ -32 $+32$
8	$2\mathcal{U}$	$1\mathcal{L}\mathcal{U}\mathcal{U}\mathcal{R}$	0001	
9	$1\mathcal{R}$	$\mathcal{I}\mathcal{I}\mathcal{I}\mathcal{S}$	0001	
10	$sign$	$+$	0	$+32$ -32 $+32$ $+32$
11	$ref.$		1010	$+48$ -32 $+48$ $+32$

have only insignificant coefficients. Therefore the significance of these curves is 1100, which is placed at $\widehat{\mathcal{H}}$.

Step 4 The $1\mathcal{U}$ curve represents 4^1 pixels, e.g. 63, -31, 23 and -34, which are significant ($\mathcal{S}$), insignificant ($\mathcal{I}$), insignificant and significant coefficients, respectively. Thereby, the significance of this curve is 1001, which is placed at $\widehat{\mathcal{H}}$.

Step 5 At $1\mathcal{U}$ only the signs of 63 and -34 are coded. Thus, sign bits for these pixels are 01, which are placed at $\widehat{\mathcal{H}}$. Furthermore, 63 and -34 are laid into the LSP.

Step 6 From Step 3, the $1\mathcal{L}$ curve represents 4^1 pixels, e.g. 49 ($\mathcal{S}$), 10 ($\mathcal{I}$), -13 ($\mathcal{I}$) and 14 ($\mathcal{I}$). Thus, the significance bits in this curve are 1000, which are placed at $\widehat{\mathcal{H}}$.

Step 7 At $1\mathcal{L}$ only the sign of 49 is coded. Thus, sign bit for this pixel is 0, which is placed at $\widehat{\mathcal{H}}$. Furthermore, 49 is laid into the LSP.

Step 8 From Step 2, using the production rules, a $2\mathcal{U}$ curve changes to $1\mathcal{L}\mathcal{U}\mathcal{U}\mathcal{R}$. At the first bit-plane, the first three curves $1\mathcal{L}$, $1\mathcal{U}$ and $1\mathcal{U}$ are subsets with insignificant coefficients, while the last one $1\mathcal{R}$ has at least one significant pixel, in this case only 47. Therefore the significance of these curves is 0001, which is placed at $\widehat{\mathcal{H}}$.

Step 9 The $1\mathcal{R}$ curve represents 4^1 pixels, e.g. 2 ($\mathcal{I}$), -3 ($\mathcal{I}$), -1 ($\mathcal{I}$) and 47 ($\mathcal{S}$). Thus, the significance bits in this curve are 0001, which are placed at $\widehat{\mathcal{H}}$.

Step 10 At $1\mathcal{R}$ only the sign of 47 is coded. Thus, sign bit for this pixel is 0, which is placed at $\widehat{\mathcal{H}}$. Furthermore, 47 is laid into the *LSP*.

Step 11 The encoded *LSP* contains four ordered entries: 63(1**1**1111), −34(1**0**0010), 49(1**1**0001) and 47(1**0**1111). At the end of $\widehat{\mathcal{H}}$ is added the *second* most significant bits of each entry of the encoded LSP, i.e. 1010. Therefore, when the bitstream $\widehat{\mathcal{H}}$ is received by the decoder, it recovers a *LSP* with the following values: +48(110000), −32(100000), +48(110000) and +32(100000). Binary magnitudes in parentheses are in absolute value beacuse the sign bits are encoded (or decoded) previously.

3.6 H*i*-SET Codestream Syntax

The H*i*-SET Codestream Syntax is a compressed representation of image data that contains all parameters used in the encoding process and it is also a lineal stream of bits. This bitstream is mainly divided into two consecutive groups: Headers and the $\widehat{\mathcal{H}}$ obtained in the coding process (Figure 3.9).
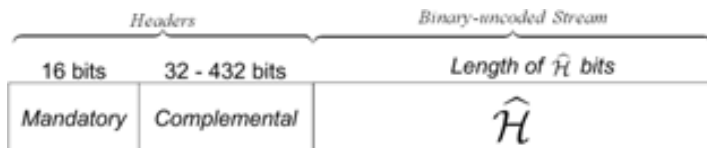
Figure 3.9: H*i*-SET Codestream Syntax.

Headers are subdivided in groups of Markers. We consider two types: Mandatory and Complemental Headers. Figure 3.10(a) shows the structure of the Mandatory Header, that is a 16 bit fixed size substream. This Header is fractionated in six Markers, namely $Image_{size}$, thr_{max}, w_{lev}, $Channels$, w_{filter} and Q_{step} , described as:

$Image_{size}$ (4 bits). If this marker is different to zero means that the processed image is squared with both height and width equal to $2^{Image_{size}+1}$. Thus the overall size of a square image varies from 4^2 to 4^{16} pixels. Otherwise when $Image_{size} = 0000$ the markers $Image_{height}$ and $Image_{width}$ of the Complemental Header are used for establishing the image size.

thr_{max} (4 bits). It stores the maximum threshold $thr - 1$ defined in eq (3.9), hence, its value varies from 1 to 16. Thus, Hi-SET can process an image up to 16 bit-planes.

w_{lev} (3 bits). This marker contains the number of spatial frequencies minus one performed by the wavelet transform, thus its value varies from 1 to 8 wavelet spatial frequencies.

$Channels$ (3 bits). The number of image (color) components minus one is stored in this marker, thus managing up to eight components.

w_{filter} (1 bit). If it is one, a 9/7 wavelet filter is used, otherwise a 5/3 filter is employed.

Q_{step} (1 bit). It indicates whether the coefficients are quantized or not. If they are quantized, the size of Quantization steps Δ_s^o are placed in a marker at the end of the Complemental Header.

(a) Mandatory Header.
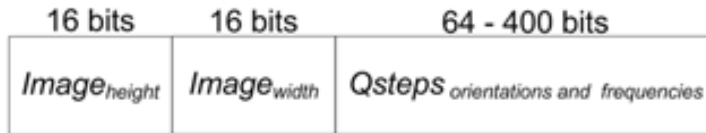
(b) Complemental Header.

Figure 3.10: Hi-SET Headers with their Markers.

Figure 3.10(b) shows the Complemental Header, which is formed by three consecutive Markers: two for storing the size of a non-squared image and the other one for the quantization steps.

$Image_{height}$ (16 bits). It contains the height of a non-squared image. Hence, an image up to 65535 pixel height can be supported.

$Image_{width}$ (16 bits). It contains the width of a non-squared image. Hence, an image up to 65535 pixel width can be supported.

$Qsteps_{orientation\ and\ frequencies}$ (64-400 bits). This marker is a collection of several sub-markers. H*i*-SET can use a quantization step Δ_s^o for every spatial frequency (indexed by s) and spatial orientation (indexed by o) for a wavelet plane $\omega_{s,o}$, in addition to another one for the residual plane $c_{w_{lev}+1}$.

Since the Codestream of H*i*-SET supports up to $w_{lev} + 1$ spatial frequencies and three spatial orientations, there are $3 \times w_{lev} + 4$ quantization steps.

Each quantization step is represented by a two-byte long sub-marker, which is divided in three parts: Sign, Exponent ε_s^o and Mantissa μ_s^o (Figure 3.11).

The most significant bit of the sub-marker is the sign of Δ_s^o, whether 0 for positive or 1 for negative. The ten least significant bits are employed for the allocation of μ_s^o, which is defined by (10) as:

$$\mu_s^o = \left\lfloor 2^{10} \left(\frac{\Delta_s^o}{2^{R_s^o - \varepsilon_s^o}} - 1 \right) + \frac{1}{2} \right\rfloor \tag{3.10}$$

Equation (3.11) expresses how ε_s^o is obtained, which is stored at the 5 remaining bits of the Δ_s^o sub-marker

$$\varepsilon_s^o = R_s^o - \lceil \log_2 |\Delta_s^o| \rceil \tag{3.11}$$

where R_s^o is the number of bits used to represent the peak coefficient inside ω_s^o, defined as

$$R_s^o = \lceil \log_2 [\max \{\omega_s^o\}] \rceil \ . \tag{3.12}$$

Figure 3.11: Structure of the Δ_s^o Sub-marker.

45

3.7   Experiments and Numerical Results

The aim of this section is to show how much error is introduced by H*i*-SET during the compression process. The quality of the recovered image is obtained by comparing it against the original image.

3.7.1   Comparison with Hilbert Curve based algorithms

H*i*-SET has some resemblances with other image compression algorithms, concretely we are interested in those developed by Kim and Li (20) and Biswas (9). Similarly to them, H*i*-SET maximizes the correlation between pixels using a Hilbert scanning. The differences between H*i*-SET and these methods are that H*i*-SET is an embedded algorithm and also proposes a coding scheme, while the Kim and Li and Biswas methods are not embedded because the entropy is encoded by a Huffman coder.

Figure 3.12 shows the comparison between these two algorithms and H*i*-SET. This comparison has been performed only for the case of the image *Lenna* because it is the only result reported by these authors.

Figure 3.12(a) shows the PSNR difference between H*i*-SET and Kim and Li algorithm as a function of the bit-rate (bits per pixel, bpp). On the upper horizontal axis, we show the PSNR obtained at the bpp shown on the lower horizontal axis. On average, H*i*-SET reduces PSNR in 4.75 dB (i.e. reduces the Mean Square Error around 63.07 %).

Similarly, Figure 3.12(b) shows the difference between H*i*-SET and Biswas algorithm. On average, H*i*-SET diminishes the MSE in 84.66% (8.15 dB). For example, the quality of a H*i*-SET compressed image stored at 22.4 KB (0.70 bpp) is 36.37 dB, while the Biswas algorithm obtains 28.73 dB, that is, 7.65 dB less.

Thus, on average our method improves the image quality of these two Hilbert fractal based methods in approximately 6.20 dB.

3.7.2   Comparing H*i*-SET and JPEG2000 coders

Two tests are performed in order to compare H*i*-SET and JPEG2000 coders. The first test is to apply to the coders the same parameters and the second one is to employ the same subset of wavelet coefficients.
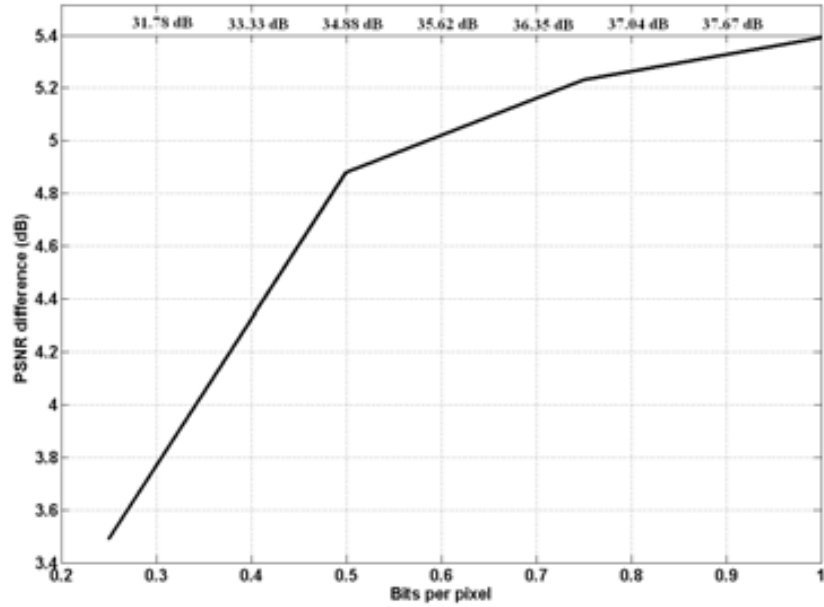
(a) Kim and Li algorithm
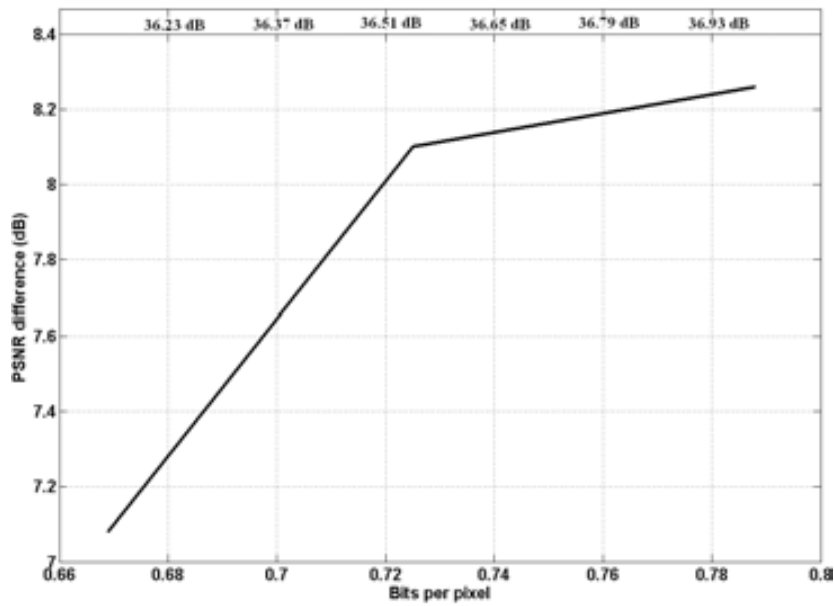
(b) Biswas algorithm

Figure 3.12: Performance comparison (PSNR difference) between H*i*-SET and the algorithms proposed by Kim and Li and Biswas, for a gray-scale image *Lenna*. On the upper part of the figures we show the obtained PSNR at the bpp shown on the lower part.

3.7.2.1 With the same parameters

In this section we compare Hi-SET and JPEG2000 with the same parameters. The comparison of Rate Distortion (RD) performance for JPEG2000 is taken from (43, Sec. 1.5), where the parameters used are the following:

- Single Tile.

- 3 levels of wavelet decomposition 9/7 Filter (Table 3.2).

- Size code blocks 64 × 64.

- Keeping the step size the same (default $\Delta = 1/128$).

Therefore, the only way to achieve a given bit-rate is the truncation of the compressed code-block bit-stream, which forms a single layer. The Embedded Block Coding with Optimized Truncation (EBCOT) (51) postcompression RD optimization procedure is used to determine these truncation points.
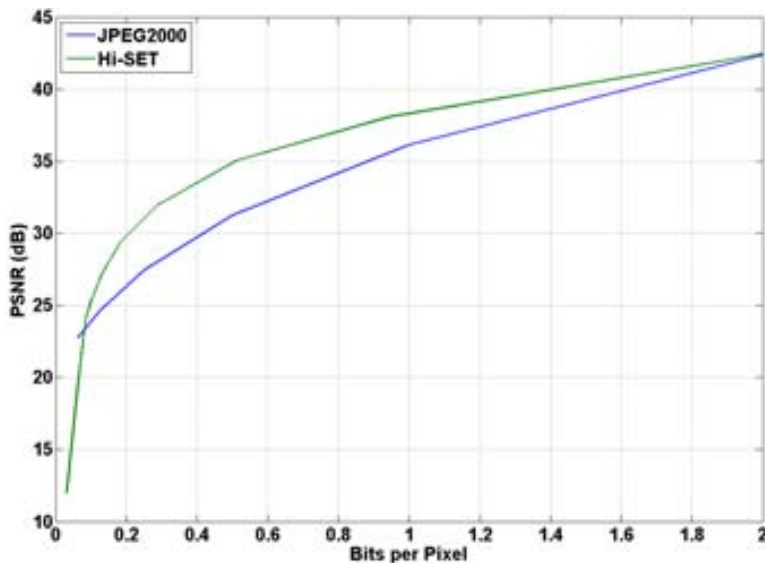
Figure 3.13: Comparison of RD performance of JPEG2000 and Hi-SET for the image *Lenna*. The JPEG2000 results are taken from (43, Sec. 1.5)).

Figure 3.13 shows the comparison of RD performance of JPEG2000 and Hi-SET for the image *Lenna*(Fig. 2.3(a)) Channel Y. The Hi-SET coder obtains, on the average, in a higher PSNR=1.38 dB than the JPEG2000 standard. The reported results are obtained for lossy coding for bit-rates of 0.0625, 0.125, 0.25, 0.5, 1.0 and 2.0.

3.7.2.2 With the same subset of wavelet coefficients

An image compression system is a set of processes with the aim of representing the image with a string of bits, keeping the length as small as possible. These processes are mainly Transformation, Quantization and Entropy Coding. For the sake of comparing the performance between the JPEG2000 standard entropy coder (51) and Hi-SET entropy coder, the entropy coding is isolated from the rest of the subprocess of the compression system. This way, a subset of wavelet coefficients are selected from the original source image data $\mathfrak{I}_{org}$ such that $\mathfrak{I}_{org} \geq 2^{thr-bpl+1}$, being bpl the desired bit-plane and thr the maximum threshold

$$thr = \left\lfloor \log_2 \left(\max_{(i,j)} \left\{ \left| \mathfrak{I}_{org(i,j)} \right| \right\} \right) \right\rfloor . \tag{3.13}$$

These selected coefficients are inverse wavelet transformed in order to create a new source of image data, i.e. $\mathfrak{I}'_{org}$, which are near-losslessly compressed, that is until the last bit-plane, by each coder. Figure 3.14 depicts this process. The software used to perform JPEG2000 compression is *Kakadu* (50) and *JJ2000* (40). The irreversible component transformation (ICT, YC_bC_r) is used in addition to the 9/7 irreversible wavelet transform.

Hi-SET is tested on the 24 bit color images of Tampere Image Database (TID2008)(39), which contains 24 images (Figure A.2). All images in the database are 512×384 pixels. The fixed size of all images is obtained by cropping selected fragments of this size from the original images.

The compression algorithms are evaluated in five experiments: low resolution gray-scale images, medium resolution gray-scale images, low resolution color images, medium resolution color images and high resolution gray-scale images. For these experiments the JPEG2000 compression is performed by *JJ2000* implementation(40).

Experiment 1. Low resolution gray-scale images. In order to test the image coders in the worst possible conditions, the image database is transformed and resized into gray-scale images (Y component) of 128×96 pixels. The less pixels an image contains, the less redundancies can be exploited on it. Figure 3.15 shows the quality of the recovered images as a function of their compression rate. On the average, an image with 30 dB is compressed by JPEG2000 coder
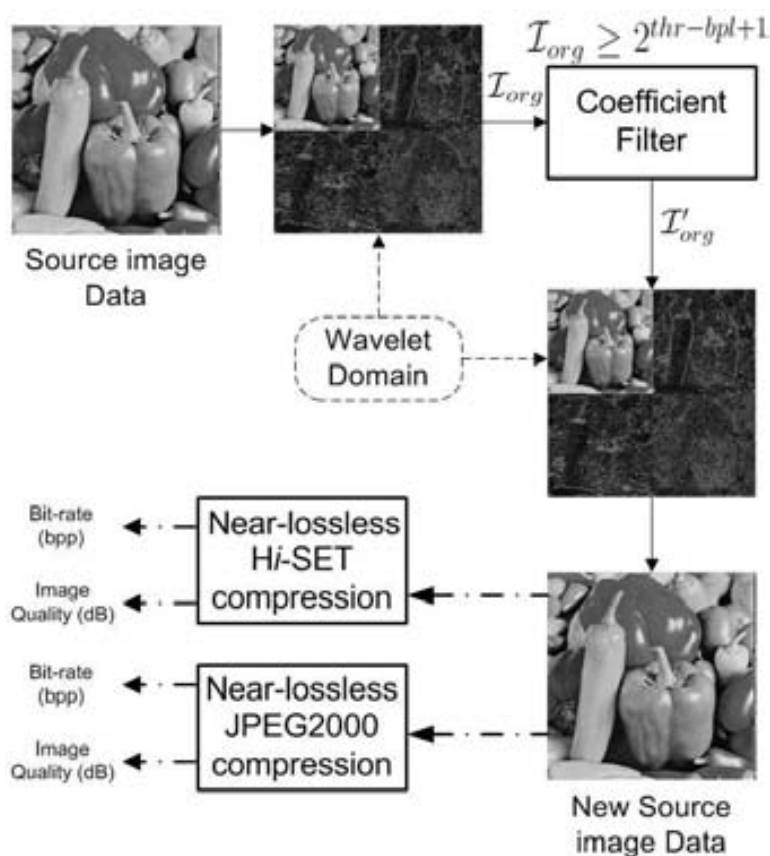
Figure 3.14: Bit-plane selection. Some coefficients are selected provided that they fulfil the current threshold.

(dashed function) at 1.59 bpp (1:5.04 compression ratio) in 2.38 KBytes and by Hi-SET (continuous function) at 1.10 bpp (1:7.3 ratio) in 1.64 KBytes. Figure 3.16 shows this differences when the image *kodim18* is compressed at 0.8 bpp by JPEG2000 and Hi-SET, being the latter 2.36 dB better. In general, for 128×96 gray-scale images the JPEG2000 coder compresses either 0.551 bpp less or stores 847 Bytes more than Hi-SET with the same objective visual quality. At the same compression rate Hi-SET is 1.84 dB better.

Experiment 2 Medium resolution gray-scale images. In this experiment, the source image data both for the JPEG2000 standard coder and Hi-SET algorithms are the selected images from the TID2008 (Figure A.2) transformed into gray-scale images (Y component). Figure 3.17 shows the average quality of the recovered images as a function of compression rate, for both JPEG2000 (dashed

Figure 3.15: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 1: Compression rate vs image quality of the 128×96 gray-scale image database.
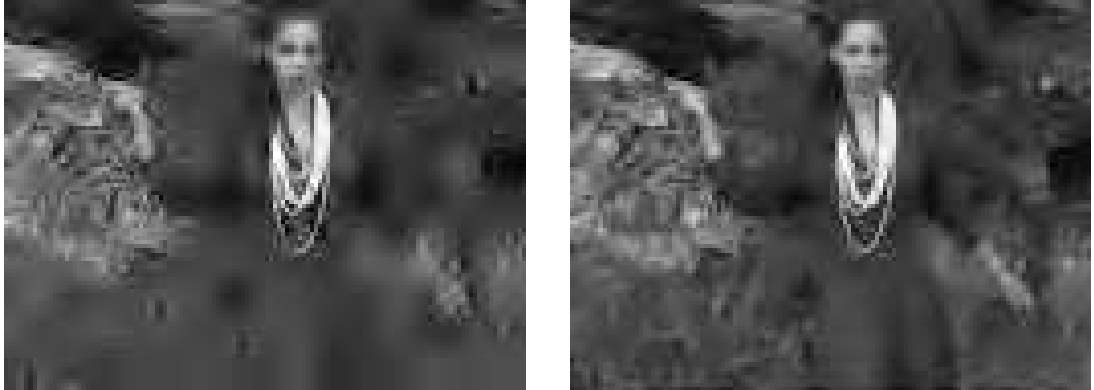
(a) JPEG2000 PSNR=23.99 dB

(b) H*i*-SET PSNR=26.35 dB

Figure 3.16: Experiment 1. Example of 128×96 reconstructed image *kodim18* compressed at 0.8 bpp (Y Component).

function) and H*i*-SET (continuous function). H*i*-SET improves the image quality in approximately 0.427 dB with the same compression rate, or the bit-rate in approximately 0.174 bpp with the same image quality. It implies saving around 4.18 KBytes for 512×384 pixels gray-scale images. On average, a 512×384 image compressed by JPEG2000 with 30 dB needs 19.8 KBytes at 0.827 bpp, while H*i*-SET needs 5.75 KBytes less at 0.587 bpp. In Figure 3.18, the difference

in visual quality between JPEG2000 and H*i*-SET when the image *kodim23* is compressed at 0.2 bpp are seen . The image quality of the recovered image coded by (a) JPEG2000 is 2.74 dB lower than the one obtained by (b) H*i*-SET.

Figure 3.17: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 2: Compression rate vs image quality of the original image database in gray-scale.

(a) JPEG2000 PSNR=27.05 dB (b) H*i*-SET PSNR=29.79 dB

Figure 3.18: Experiment 2. Example of 512×384 recovered image *kodim23* compressed at 0.2 bpp (Y Component).

Experiment 3. Low resolution color images. As previously explained, the image database is resized (performing a cropping process) to 128×96 pixels images. They are transformed into the YC_bC_r color space (the one used by JPEG2000).

Figure 3.19 shows the PSNR of recovered images as a function of compression rate. On the average, an image compressed by Hi-SET(continuous function) with 34 dB is stored in 4.87 KBytes at 3.25 bpp, while using JPEG2000 (dashed function) it is stored in 6.76 KBytes at 4.51 bpp. In Figure 3.20 we can see these differences when image *kodim06* is compressed at 1.4 bpp by JPEG2000 standard (a) and Hi-SET(b). Thus, at the same compression rate, Hi-SET obtains a better image quality (up to 2.26 dB better) than JPEG2000 coder. On avergage, Hi-SET compresses either 0.925 bpp or saves 1.39 KBytes more than the JPEG2000 coder with the same statistical error induced by the coding process or 1.43 dB with the same compression rate.
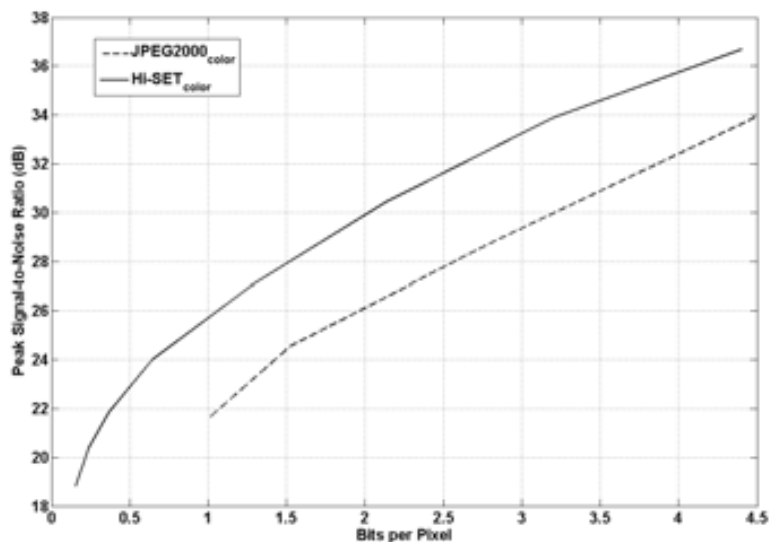
Figure 3.19: Comparison between Hi-SET and JPEG2000 image coders. Experiment 3: Compression rate vs image quality of the 128×96 color image data base.

Experiment 4. Medium resolution color images. In this fourth experiment, tests are made on the selected images of the Kodak test set transformed into YC_bC_r color space (it is the color space used by JPEG2000). Figure 3.21 shows the relation between compression rate and average quality. On average, a 512×384 image compressed by Hi-SET(continuous function) with 35 dB is stored in 46.8 KBytes at 1.95 bpp, while JPEG2000 (dashed function) stores it in 53.2 KBytes at 2.22 bpp. In Figure 3.22 we can see the difference when the image *kodim04* is compressed at 0.4 bpp by JPEG2000 (a) and Hi-SET(b). At the same compres-

(a) JPEG2000 PSNR=25.99 dB   (b) H*i*-SET PSNR=28.25 dB

Figure 3.20: Experiment 3. Example of 128×96 recovered image *kodim06* compressed at 1.4 bpp (Y, C_b and C_r Components).

sion ratio, H*i*-SET improves image quality by 1.83 dB. On average H*i*-SET either compresses 0.33 bpp more with the same image quality or reduces in 1.06 dB the error with the same bit-rate. Thus, H*i*-SET saves 7.9 KBytes more than the JPEG2000 standard for 512×384 color images.

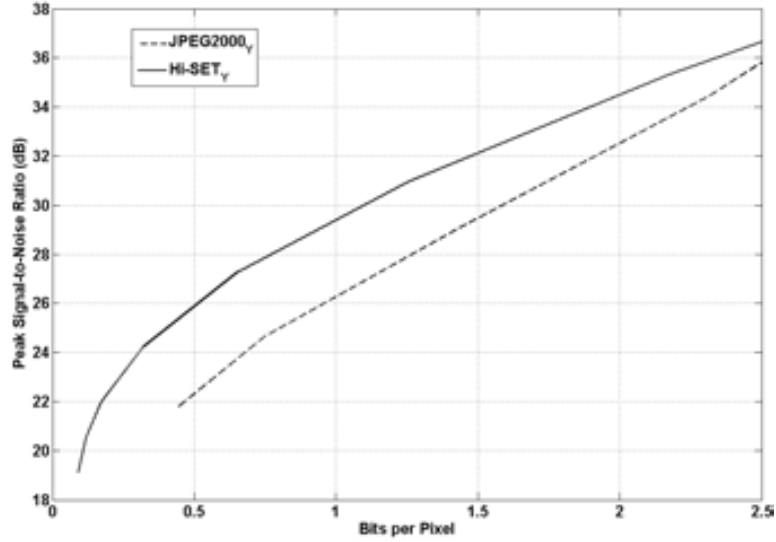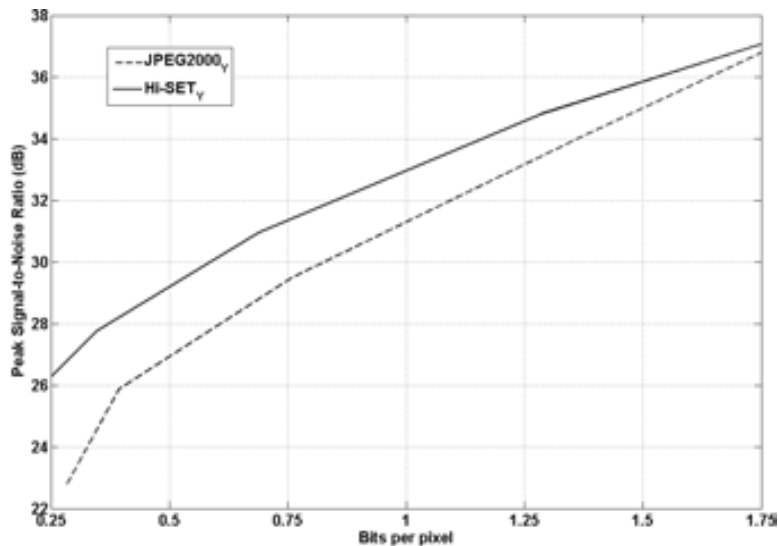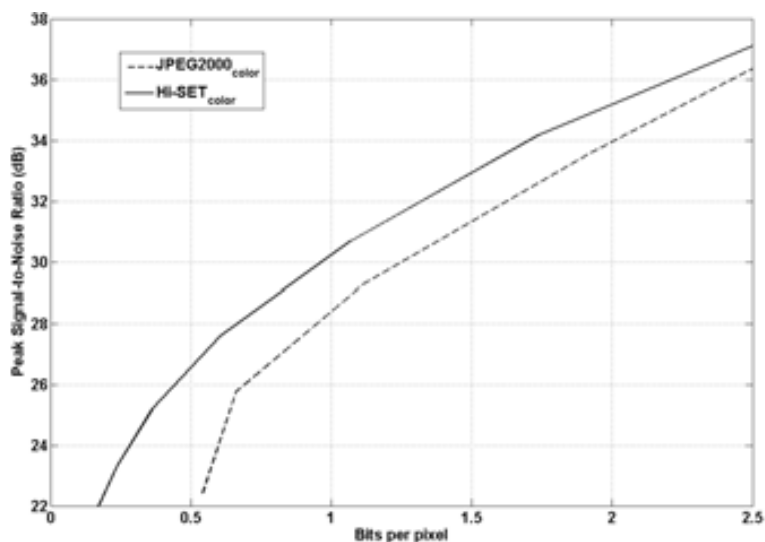**Figure 3.21:** Comparison between H*i*-SET and JPEG2000 image coders. Experiment 4: Compression rate vs image quality of the original color image data base.

Experiment 5. High resolution gray-scale images. This experiment is performed in order to test the H*i*-SET compression performance with high resolution images.
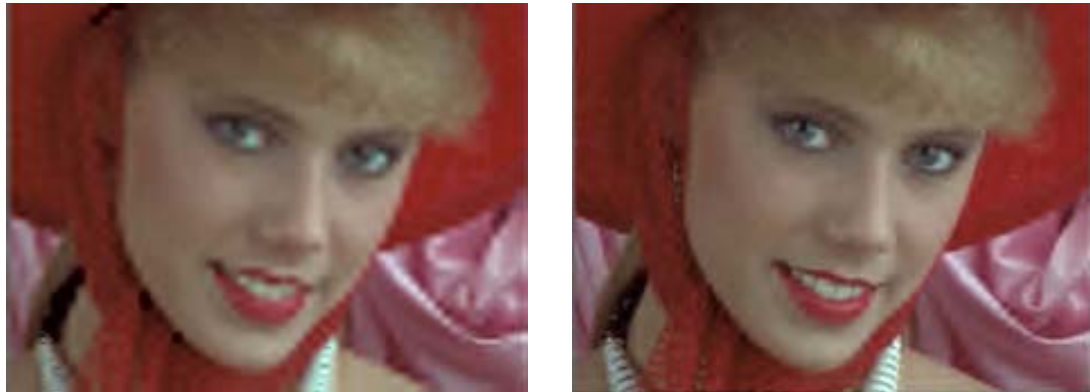
(a) JPEG2000 PSNR=28.53 dB

(b) H*i*-SET PSNR=30.36 dB

Figure 3.22: Experiment 4. Example of 512×384 recovered image *kodim04* compressed at 0.4 bpp (Y, C_b and C_r Components).

We use th e Y component of image *Bicycle* (19). Table 3.4 shows the PSNR obtained by JPEG2000 and H*i*-SET at 0.25, 0.50 and 0.75 bpp. On average, images recovered by JPEG2000 are 3.16 dB lower than the ones decoded by H*i*-SET. Figure 3.23 shows image *Bicycle* compressed both by JPEG2000 (a) and H*i*-SET(b) at 0.38 bpp (e.g. 1:21.05), which is stored in 243 KBytes. The right column of Figure 3.23 shows bottom left squared sections of 512×512 pixels. These regions are cropped to ease the visual inspection of the differences between algorithms. On the other hand, left column displays recovered images in their original size. This Figure shows that the image processed by H*i*-SET has a better visual quality (it reduces the mean squared error in 80.41 percent in comparison to JPEG2000).

Table 3.4: Comparison of lossy encoding by JPEG2000 standard and H*i*-SET for the image *Bicycle*.

bpp (rate)	JPEG2000 PSNR in dB's	H*i*-SET PSNR in dB's
0.25 (32:1)	19.08	23.82
0.50 (16:1)	24.91	28.00
0.75 (10.67:1)	29.65	31.30

(a) JPEG2000 PSNR=19.48 dB

(b) Hi-SET PSNR=26.56 dB

Figure 3.23: Experiment 5. Examples of 2048×2560 recovered image *Bicycle* compressed at 0.38 bpp (Y Component).

3.7.2.3 Perceptual Image Quality Analysis

Although H*i*-SET is not developed taking into account perceptual criteria, we compare the perceptual image quality of JPEG2000 and H*i*-SET. H*i*-SET and JPEG2000 are compared with some state of the art numerical image quality estimators. Concretely, H*i*-SET and JPEG2000 performances are compared using MSE(18), PSNR(18), SSIM(45), MSSIM(54), VSNR(12), VIF(58), VIFP(45), UQI(55), IFC(47), NQM(14), WSNR(25) SNR and $\mathcal{C}_w$PSNR(Section 2.3.1). This comparison is made across the image databases: CMU (Sec. A.5, using *Kakadu* implementation for JPEG2000 compression (50)), CSIQ(Sec. A.4, using *JJ200* implementation for JPEG2000 compression), IVC(Sec. A.1, using *JJ2000* implementation for JPEG2000 compression), LIVE(Sec. A.3, using *Kakadu* implementation for JPEG2000 compression) and TID2008(Sec. A.2, using *JJ2000* implementation for JPEG2000 compression), for color and gray-scale (Y Channel) compression. For the sake of simplicity, in this Section, only $\mathcal{C}_w$PSNR results are exposed (Fig. 3.24) and the rest of metrics are shown in Annex B. Thus, Fig. 3.24 shows that H*i*-SET significantly improves the results of JPEG2000 coder. That is, the images obtained by H*i*-SET are perceptually better than the ones obtained by JPEG2000, regardless the JPEG2000 implementation.

3.8 Conclusions

The H*i*-SET coder is based on Hilbert scanning of embedded quadTrees. It has low computational complexity and some important properties of modern image coders such as embedding and progressive transmission. This is achieved by using the principle of partial sorting by magnitude when a sequence of thresholds decreases. The desired compression rate can be controlled just by chunking the stream at the desired file length. When compared to other algorithms that use Hilbert scanning for pixel ordering, H*i*-SET improves image quality by around 6.20 dB. H*i*-SET achieves higher compression rates than JPEG2000 coder not only for high and medium resolution images but also for low resolution ones where it is difficult to find redundancies among spatial frequencies.

(a) CMU gray-scale

(b) CMU color
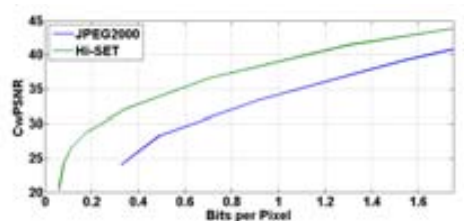
(c) CSIQ gray-scale
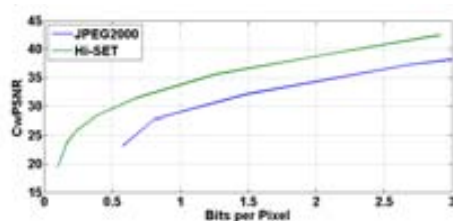
(d) CSIQ color
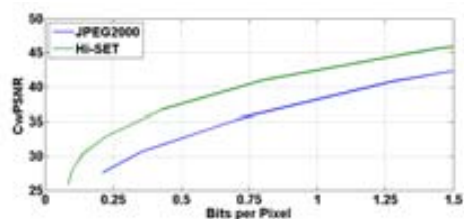
(e) IVC gray-scale
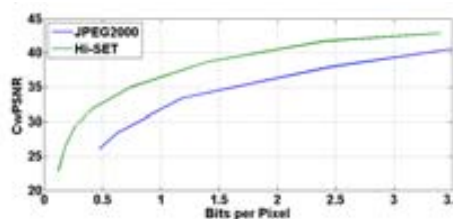
(f) IVC color
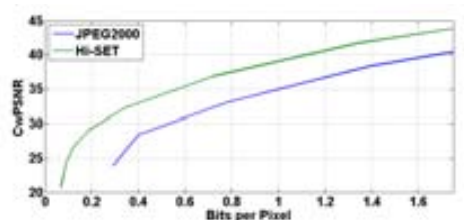
(g) LIVE gray-scale
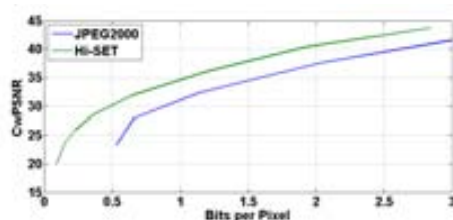
(h) LIVE color
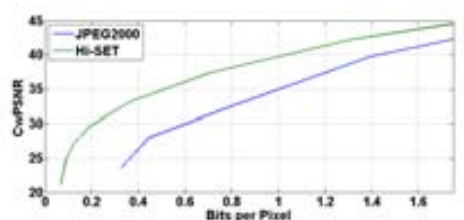
(i) TID2008 gray-scale
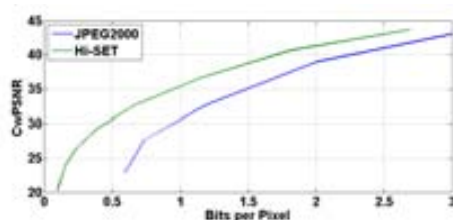
(j) TID2008 color

Figure 3.24: Comparison between JPEG2000 vs Hi-SET image coders. Compression rate vs perceptual image quality, performed by $\mathcal{C}_w$PSNR, of the CMU (a-b), CSIQ (c-d), CMU (e-f), LIVE (g-h) and TID2008 (i-j) image databases. In left column is shown the gray-scale compression of all image databases, while the right one color compression is depicted.

Chapter 4

Perceptual Quantization

4.1 Introduction

Digital image compression has been a research topic for many years and a number of image compression standards has been created for different applications. The JPEG2000 is intended to provide rate-distortion and subjective image quality performance superior to existing standards, as well as to supply functionality (10). However, JPEG2000 does not provide the most relevant characteristics of the human visual system, since for removing information in order to compress the image mainly information theory criteria are applied. This information removal introduces artifacts to the image that are visible at high compression rates, because of many pixels with high perceptual significance have been discarded.

Hence, it is necessary an advanced model that removes information according to perceptual criteria, preserving the pixels with high perceptual relevance regardless of the numerical information. The Chromatic Induction Wavelet Model presents some perceptual concepts that can be suitable for it. Both CIWaM and JPEG2000 use wavelet transform. CIWaM uses it in order to generate an approximation to how every pixel is perceived from a certain distance taking into account the value of its neighboring pixels. By contrast, JPEG2000 applies a perceptual criteria for all coefficients in a certain spatial frequency independently of the values of its surrounding ones. In other words, JPEG2000 performs a global transformation of wavelet coefficients, while CIWaM performs a local one.

CIWaM attenuates the details that the human visual system is not able to perceive,

enhances those that are perceptually relevant and produces an approximation of the image that the brain visual cortex perceives. At long distances, as Figure 2.3(d) depicts, the lack of information does not produce the well-known compression artifacts, rather it is presented as a softened version, where the details with high perceptual value remain (for example, some edges).

4.2 JPEG2000 Global Visual Frequency Weighting

In JPEG2000, only one set of weights is chosen and applied to wavelet coefficients according to a particular viewing condition (100, 200 or 400 dpi's) with fixed visual weighting(10, Annex J.8). This viewing condition may be truncated depending on the stages of embedding, in other words at low bit rates, the quality of the compressed image is poor and the detailed features of the image are not available since at a relatively large distance the low frequencies are perceptually more important.

The table 4.1 specifies a set of weights which was designed for the luminance component based on the CSF value at the mid-frequency of each spatial frequency. The viewing distance is supposed to be 4000 pixels, corresponding to 10 inches for 400 dpi print or display. The weight for LL is not included in the table, because it is always 1. Levels 1, 2, ... , 5 denote the spatial frequency levels in low to high frequency order with three spatial orientations, *horizontal, vertical* and *diagonal*.

Table 4.1: Recommended JPEG2000 frequency (s) weighting for 400 dpi's ($s = 1$ is the lowest frequency wavelet plane).

s	*horizontal*	*vertical*	*diagonal*
1	1	1	1
2	1	1	0.731 668
3	0.564 344	0.564 344	0.285 968
4	0.179 609	0.179 609	0.043 903
5	0.014 774	0.014 774	0.000 573

4.3 Perceptual Forward Quantization

4.3.1 Methodology

Quantization is the only cause that introduces distortion into a compression process. Since each transform sample at the perceptual image $\mathfrak{I}_\rho$ (from Eq. 2.4) is mapped independently to a corresponding step size either Δ_s or Δ_n, thus $\mathfrak{I}_\rho$ is associated with a specific interval on the real line. Then, the perceptually quantized coefficients $\mathfrak{Q}$, from a known viewing distance d, are calculated as follows:

$$\mathfrak{Q} = \sum_{s=1}^{n} \sum_{o=v,h,d} sign(\omega_{s,o}) \left\lfloor \frac{|\alpha(\nu,r) \cdot \omega_{s,o}|}{\Delta_s} \right\rfloor + \left\lfloor \frac{c_n}{\Delta_n} \right\rfloor \qquad (4.1)$$

Unlike the classical techniques of Visual Frequency Weighting (VFW) on JPEG2000, which apply one CSF weight per sub-band (10, Annex J.8), Perceptual Quantization using CIWaM (ρSQ) applies one CSF weight per coefficient over all wavelet planes $\omega_{s,o}$. In this section we only explain Forward Perceptual Quantization using CIWaM (F-ρSQ). Thus, Equation 4.1 introduces the perceptual criteria of Equation 2.4 (Perceptual Images) to each quantized coefficient of Equation 3.6(Dead-zone Scalar Quantizer). A normalized quantization step size $\Delta = 1/128$ is used, namely the range between the minimal and maximal values at $\mathfrak{I}_\rho$ is divided into 128 intervals. Finally, the perceptually quantized coefficients are entropy coded, before forming the output code stream or bitstream. Figure 2.3 shows three CIWaM images of *Lena*, which are calculated by Equation 4.1 ($\Delta_s = 1$ and $\Delta_n = 1$) for a 19 inch screen with 1280 pixels of horizontal resolution, at 30, 100 and 200 centimeters of distance. In this specific case, Eq. 2.4 = Eq. 4.1.

4.3.2 Experimental Results applied to JPEG2000

The Perceptual quantizer F-ρSQ in JPEG2000 is tested on all the color images of the *Miscellaneous volume* of the University of Southern California Image Data Base(2). The data sets are eight 256×256 pixel images (Fig. A.5) and eight 512×512 pixel images (Fig. A.6), but only visual results of the well-known images *Lena*, *F-16* and *Baboon* are depicted, which are 24-bit color images and 512×512 of resolution. The CIWaM model is performed for a 19 inch monitor with 1280 pixels of horizontal resolution at 50

centimeters of viewing distance. The software used to obtain a JPEG2000 compression for the experiment is *JJ2000*(40).

Figure 4.1 shows the assessment results of the average performance of color image compression for each bit-plane using a Dead-zone Uniform Scalar Quantizer (SQ, Section 3.4, function with heavy dots), and it also depicts the results obtained when applying F-ρSQ(function with heavy stars).
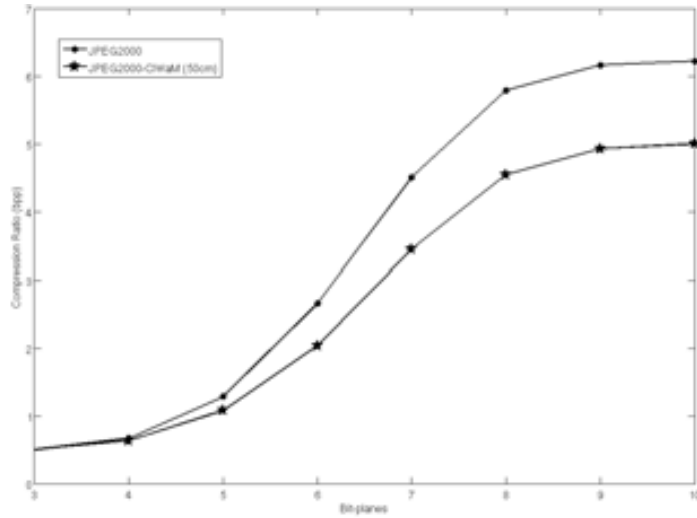
Figure 4.1: JPEG2000 Compression ratio (bpp) as a function of Bit-plane. Function with heavy dots shows JPEG2000 only quantized by the dead-zone uniform scalar manner. While function with heavy stars shows JPEG2000 perceptually pre-quantized by F-ρSQ.

Using CIWaM as a method of forward quantization, achieves better compression ratios than SQ with the same threshold, obtaining better results at the highest bit-planes, since CIWaM reduces unperceivable features. Figure 4.2 shows the contribution of F-ρSQ in the JPEG2000 compression ratio, for example, at the eighth bit-plane, CIWaM reduces 1.2423 bits per pixel than the bit rate obtained by SQ, namely in a 512×512 pixel color image, CIWaM estimates that 39.75KB of information is perceptually irrelevant at 50 centimeters.

Both Figure 4.3 and 4.4 depict examples of recovered images compressed at 0.9 and 0.4 bits per pixel, respectively, by means of JPEG2000 (a) without and (b) with F-ρSQ. Also these figures show that the perceptual quality of images forward quantized by ρSQ is better than the objective one.

Figure 4.5 shows examples of recovered images of *Baboon* compressed at 0.59, 0.54
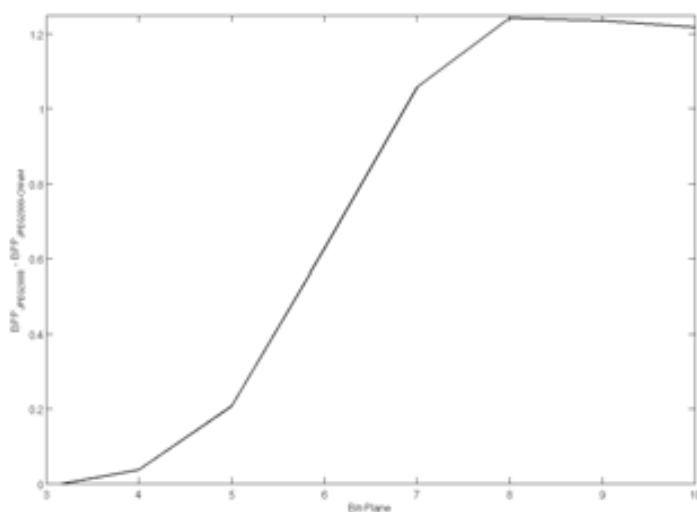
Figure 4.2: The bit-rate decrease by each Bit-plane after applying F-ρSQ on the JPEG2000 compression.
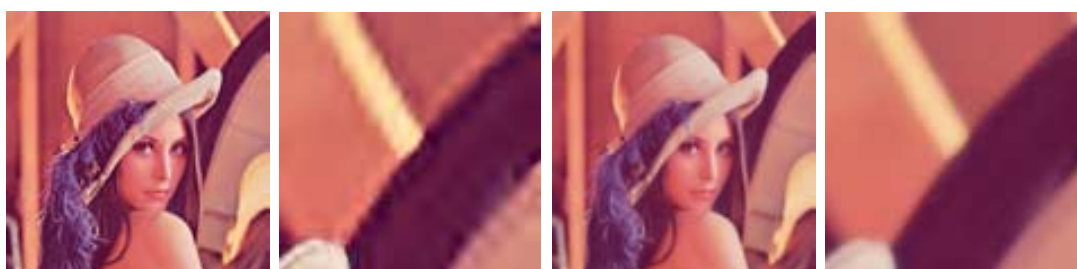
(a) JPEG2000 PSNR=31.19 dB.   (b) JPEG2000-F-ρSQ PSNR=27.57 dB.

Figure 4.3: Examples of recovered images of Lenna compressed at 0.9 bpp.

(a) JPEG2000 PSNR=25.12 dB.   (b) JPEG2000-F-ρSQ PSNR=24.57 dB.

Figure 4.4: Examples of recovered images of F-16 compressed at 0.4 bpp.

and 0.45 bits per pixel by means of JPEG2000 (a) without and (b and c) with F-ρSQ. In Fig. 4.5(a) PSNR=26.18 dB and in Fig. 4.5(b) PSNR=26.15 dB but a perceptual

metrics like WSNR (25), for example, assesses that it is equal to 34.08 dB. Therefore, the recovered image Forward quantized by ρSQ is perceptually better than the one only quantized by a SQ. Since the latter produces more compression artifacts, the ρSQ result at 0.45 bpp (Fig. 4.5(c)) contains less artifacts than SQ at 0.59 bpp. For example the *Baboon*'s eye is softer and better defined using F-ρSQ and it additionally saves 4.48 KB of information.
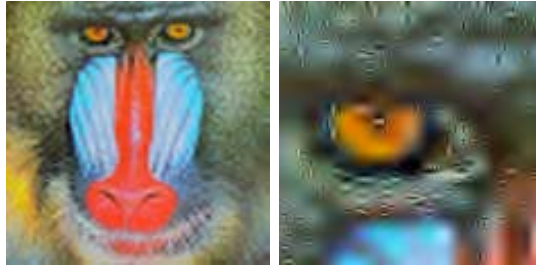
(a) JPEG2000 compressed at 0.59 bpp.
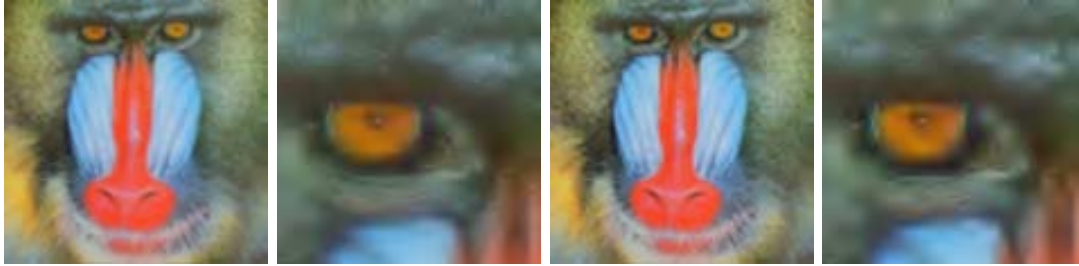
(b) JPEG2000-F-ρSQ compressed at 0.54 bpp. (c) JPEG2000-F-ρSQ compressed at 0.45 bpp.

Figure 4.5: Examples of recovered images of Baboon.

4.4 Perceptual Inverse Quantization

The proposed Perceptual Quantization is a generalized method, which can be applied to wavelet-transform-based image compression algorithms such as EZW, SPIHT, SPECK or JPEG2000. In this work, we introduce both forward (F-ρSQ) and inverse perceptual quantization (I-ρSQ) into the Hi-SET coder. This process is shown in the green blocks of Fig. 4.6. An advantage of introducing ρSQ is to maintain the embedded features not only of Hi-SET algorithm but also of any wavelet-based image coder. Thus, we call Perceptual Quantization + Hi-SET = PHi-SET or Φ_{SET}.

Figure 4.6: The Φ_{SET} image compression algorithm. Green blocks are the F-ρSQ and I-ρSQ procedures.

Both JPEG2000 and Φ_{SET} choose their VFWs according to a final viewing condition. When JPEG2000 modifies the quantization step size with a certain visual weight, it needs to explicitly specify the quantizer, which is not very suitable for embedded coding. While Φ_{SET} neither needs to store the visual weights nor to necessarily specify a quantizer in order to keep its embedded coding properties.

The main challenge underlies in to recover not only a good approximation of coefficients $\mathcal{Q}$ but also the visual weight $\alpha(\nu, r)$(Eq. 4.1) that weighted them. A recovered approximation $\widehat{\mathcal{Q}}$ with a certain distortion Λ is decoded from the bitstream by the entropy decoding process. The VFWs were not encoded during the entropy encoding process, since it would increase the amount of stored data. A possible solution is to embed these weights $\alpha(\nu, r)$ into $\widehat{\mathcal{Q}}$. Thus, our goal is to recover the $\alpha(\nu, r)$ weights only using the information from the bitstream, namely from the Forward quantized coefficients $\widehat{\mathcal{Q}}$.

Therefore, our hypothesis is that an approximation $\widehat{\alpha}(\nu, r)$ of $\alpha(\nu, r)$ can be recovered applying CIWaM to $\widehat{\mathcal{Q}}$, with the same viewing conditions used in $\mathcal{I}$. That is, $\widehat{\alpha}(\nu, r)$ is the recovered e-CSF. Thus, the perceptual inverse quantizer or the recovered $\widehat{\alpha}(\nu, r)$ introduces perceptual criteria to 3.7(Inverse Scalar Quantizer) and is given by:

$$\widehat{\mathcal{I}} = \begin{cases} \sum_{s=1}^{n} \sum_{o=v,h,d} sign(\widehat{\omega_{s,o}}) \dfrac{\Delta_s \cdot (|\widehat{\omega_{s,o}}| + \delta)}{\widehat{\alpha}(\nu, r)} + (\widehat{c}_n + \delta) \cdot \Delta_n & |\widehat{\omega_{s,o}}| > 0 \\ 0, & \widehat{\omega_{s,o}} = 0 \end{cases} \tag{4.2}$$

For the sake of showing that the encoded VFWs are approximately equal to the decoded ones, that is $\alpha(\nu, r) \approx \widehat{\alpha}(\nu, r)$, we perform two experiments.

Experiment 1: Histogram of $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$. The process of this short experiment is shown by Figure 4.7. Figure 4.7(a) depicts the process for obtaining losslessy both Encoded and Decoded visual weights for the 512×512 *Lena* image, channel Y at 10 meters. While Figures 4.7(b) and 4.7(c) shows the frequency histograms of $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$, respectively. In both graphs, the horizontal

axis represents the sort of VFW variations, whereas the vertical axis represents the number of repetitions in that particular VFW. The distribution in both histograms is similar and they have the same shape.
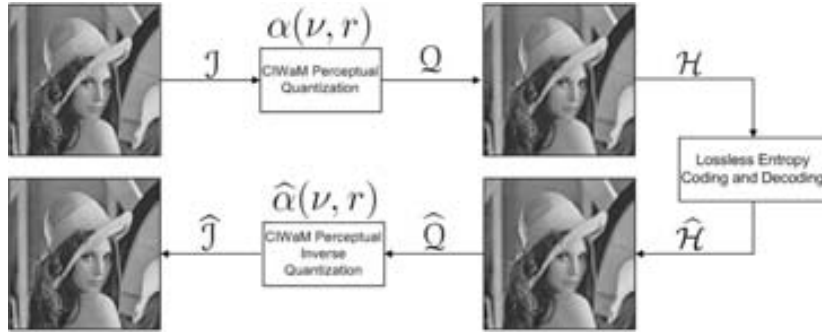
(a)
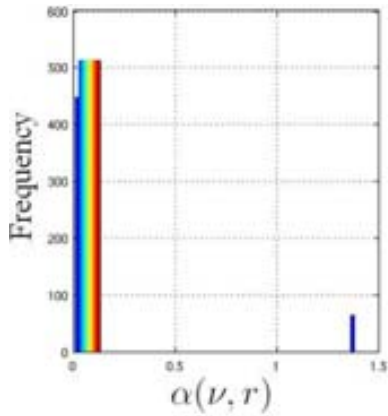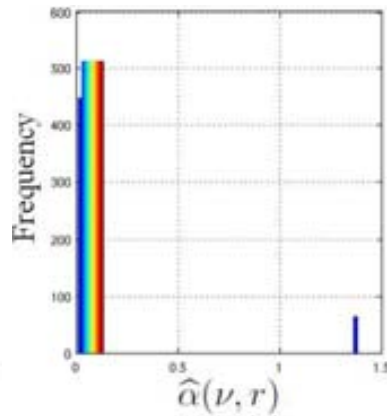
(b)   (c)

Figure 4.7: (a) Graphical representation of a whole process of compression and decompression. Histograms of (b) $\alpha(\nu, r)$ and (c) $\widehat{\alpha}(\nu, r)$ visual frequency weights for the 512×512 image *Lenna*, channel Y at 10 meters.

Experiment 2: Correlation analysis between $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$. We employ the process shown in Fig. 4.7(a) for all the images of the CMU (Figs. A.5 and A.6), CSIQ(Fig. A.4) and IVC(Fig. A.1) Image Databases. In order to obtain $\widehat{\alpha}(\nu, r)$, we measure the lineal correlation between the original $\alpha(\nu, r)$ applied during the F-ρSQ process and the recovered $\widehat{\alpha}(\nu, r)$. Table 4.2 shows that there is a high similarity between the applied VFW and the recovered one, since their correlation is 0.9849, for gray-scale images, and 0.9840, for color images.

Table 4.2: Correlation between $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$ across CMU (Figs. A.5 and A.6), CSIQ(Fig. A.4) and IVC(Fig. A.1) Image Databases.

Image Database	8 bpp gray-scale	24 bpp color
CMU	0.9840	0.9857
CSIQ	0.9857	0.9851
IVC	0.9840	0.9840
Overall	**0.9849**	**0.9844**

In this section, we only expose the results for the CMU image database. In Sections C.1.1 and C.1.2, we display the results for CSIQ and IVC image databases, respectively.

Fig. 4.8 depicts the PSNR difference (dB) of each color image of the CMU database, that is, the gain in dB of image quality after applying $\widehat{\alpha}(\nu, r)$ at $d = 2000$ centimeters to the $\widehat{Q}$ images. On average, this gain is about 15 dB. Visual examples of these results are shown by Fig. 4.9, where the right images are the original images, central images are perceptual quantized images after applying $\alpha(\nu, r)$ and left images are recovered images after applying $\widehat{\alpha}(\nu, r)$.
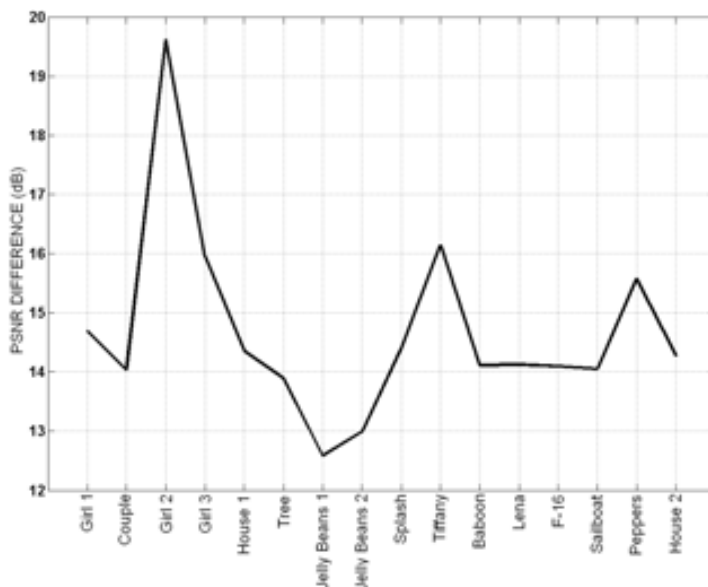
Figure 4.8: PSNR difference between $\widehat{Q}$ image after applying $\alpha(\nu, r)$ and recovered $\widehat{\mathcal{I}}$ after applying $\widehat{\alpha}(\nu, r)$ for every color image of the CMU database.
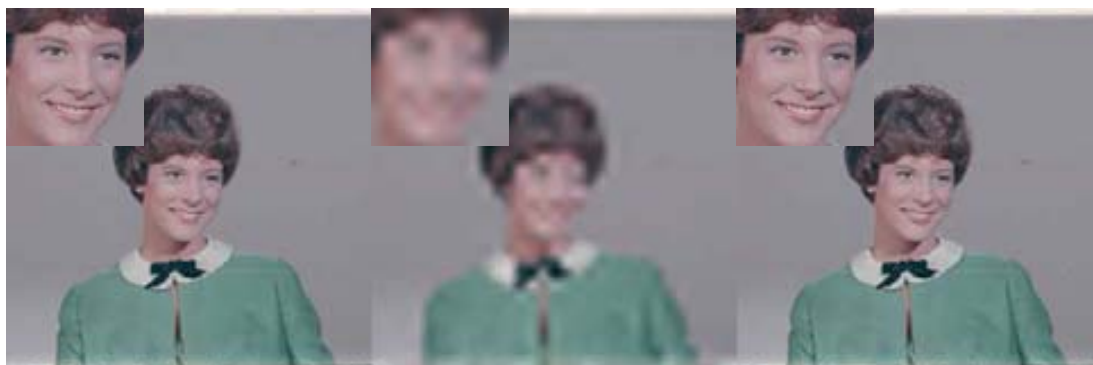
(a) Girl 2
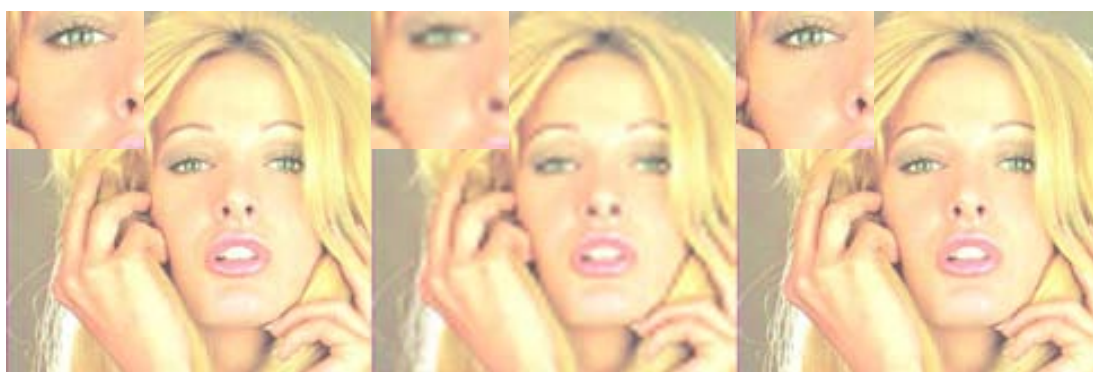
(b) Tiffany

(c) Peppers

Figure 4.9: Visual examples of Perceptual Quantization. Left images are the original images, central images are forward perceptual quantized images (F-ρSQ) after applying $\alpha(\nu, r)$ at $d = 2000$ centimeters and right images are recovered I-ρSQ images after applying $\widehat{\alpha}(\nu, r)$.

After applying $\widehat{\alpha}(\nu, r)$, a visual inspection of these sixteen recovered images show a perceptually lossless quality. We perform the same experiment experiment for gray-scale and color images with $d = 20, 40, 60, 80, 100, 200, 400, 800, 1000$ and 2000 centimeters, in addition to test their objective and subjective image quality by means of the PSNR and MSSIM metrics, respectively.

In Figs. 4.10 and 4.11, green functions denoted as F-ρSQ are the quality metrics of perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as I-ρSQ are the quality metrics of recovered images after applying $\widehat{\alpha}(\nu, r)$. Thus, either for gray-scale or color images, both PSNR and MSSIM estimations of the quantized image $\mathcal{Q}$ decrease regarding d, the longer d the greater the image quality decline. When the image decoder recovers $\widehat{\mathcal{Q}}$ and it is perceptually inverse quantized, the quality barely varies and is close to perceptually lossless, no matter the distance.

(a) PSNR (b) MSSIM
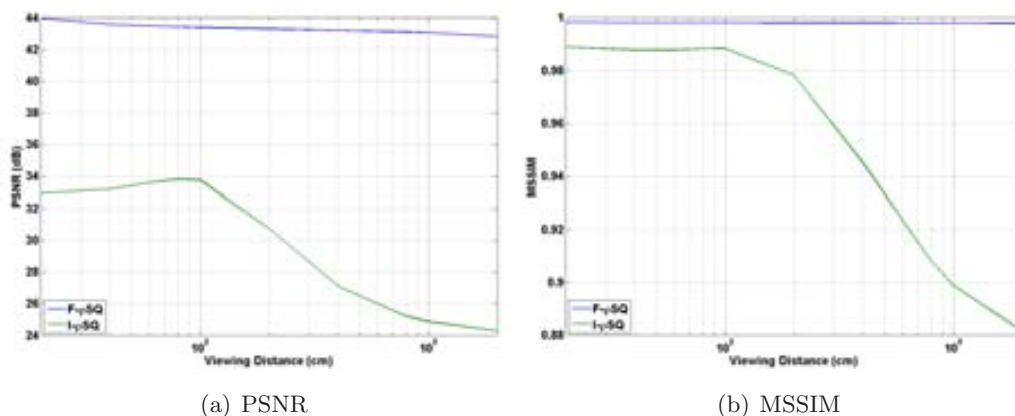
Figure 4.10: PSNR and MSSIM assessments of compression of Gray-scale Images (Y Channel) of the CMU image database. Green functions denoted as F-ρSQ are the quality metrics of forward perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as I-ρSQ are the quality metrics of recovered images after applying $\widehat{\alpha}(\nu, r)$.

4.5 Φ_{SET} Codestream Syntax

Φ_{SET} Codestream Syntax is similar to the Hi-SET one (Section 3.6), only two Markers are added inside Complemental Header (Fig. 3.10(b)), Perceptual Quantization Marker (PQ) and Observation Distance Marker (d).

(a) PSNR   (b) MSSIM
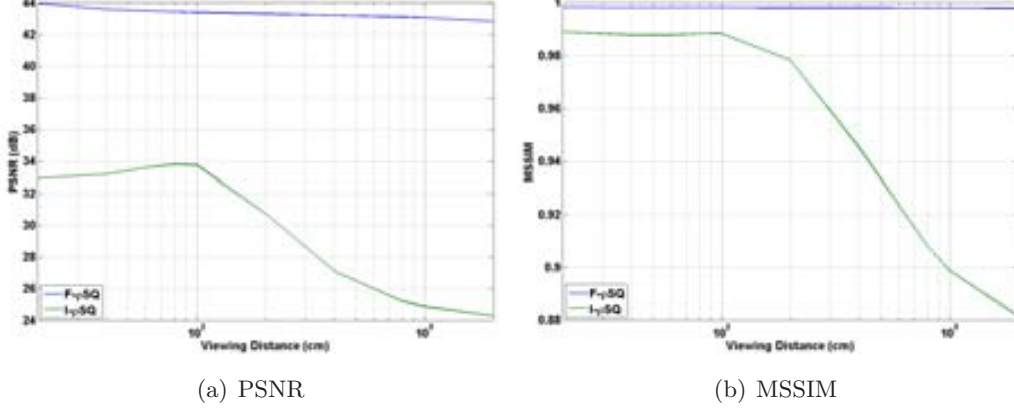
Figure 4.11: PSNR and MSSIM assessments of compression of Color Images of the CMU image database. Green functions denoted as F-ρSQ are the quality metrics of forward perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as I-ρSQ are the quality metrics of recovered images after applying $\widehat{\alpha}(\nu, r)$.

PQ (1 bit). If $Q_{step} = 1$, PQ would specify if the wavelet coefficients were perceptually quantized or not. Fig. 4.12(a) shows this marker.

d (16 bits). This marker stores the observation distance d. d is represented by a two-byte long sub-marker, which is divided in two parts: Exponent ε_d and Mantissa μ_d (Fig. 4.12(b)).

The eleven least significant bits are employed for the allocation of μ_d, which is defined as:

$$\mu_d = \left\lfloor 2^{11} \left(\frac{d}{2^{R_{d_{\max}} - \varepsilon_d}} - 1 \right) + \frac{1}{2} \right\rfloor \tag{4.3}$$

Equation (4.4) expresses how ε_d is obtained, which is stored at the 5 remaining bits of the d marker

$$\varepsilon_d = R_{d_{\max}} - \lceil \log_2 (d) \rceil \tag{4.4}$$

where $R_{d_{\max}}$ is the number of bits used to represent the peak permitted observation distance $d < 2048H$, being H the height of a 512×512 pixel image presented in an M_{size} LCD monitor with horizontal resolution of h_{res} pixels and v_{res} pixels of vertical resolution. Therefore, $R_{d_{\max}} = 11$.

1 bit

PQ

(a) *PQ* Marker

5 bits	11 bits
$\mathcal{E}_d$	μ_d
MSB	LSB
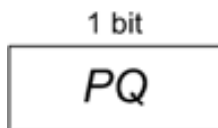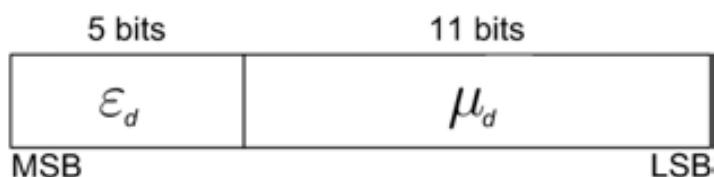
(b) *d* Marker

Figure 4.12: Markers added to Complemental Header (Fig. 3.10(b)). (a) Perceptual Quantization Marker and (b) Structure of Observation Distance Marker

4.6 Experiments and Results

4.6.1 Comparing Φ_{SET} and H*i*-SET coders
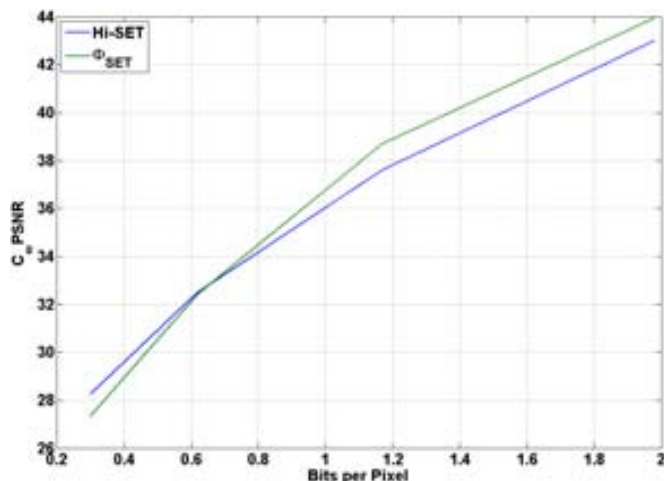
Figure 4.13: Comparison between Φ_{SET} and H*i*-SET image coders. Compression rate vs $\mathcal{C}_w$PSNR perceptual image quality of Image *Lenna* (128×128, Channel *Y*).

In this Section, we compare Φ_{SET} and H*i*-SET coders with the Image *Lenna* (Fig. 2.3(a), 128×128, Channel *Y*), in order to know if there is an improvement when ρSQ is applied to the H*i*-SET coder. Thus, for this particular case, Fig. 4.13 shows that there is a slight improvement in Φ_{SET} (Green function) in the perceptual quality of the

image of about $\mathcal{C}_w$PSNR=0.26 dB, on the average, regarding the Hi-SETcoder (Blue function).

4.6.2 Comparing Φ_{SET} and JPEG2000 coders

For the sake of comparing the performance between the JPEG2000(51) and Φ_{SET} coders, both algorithms are tested according to the process depicted in Fig. 4.14. First a Φ_{SET} compression with certain viewing conditions is performed, which gives a compressed image with a particular bit-rate (bpp). Then, a JPEG2000 compression is performed with the same bit-rate. Once both algorithms recover their distorted images, they are compared with some numerical image quality estimators such as: MSE(18), PSNR(18), SSIM(45), MSSIM(54), VSNR(12), VIF(58), VIFP(45), UQI(55), IFC(47), NQM(14), WSNR(25), SNR and $\mathcal{C}_w$PSNR(Section 2.3.1).
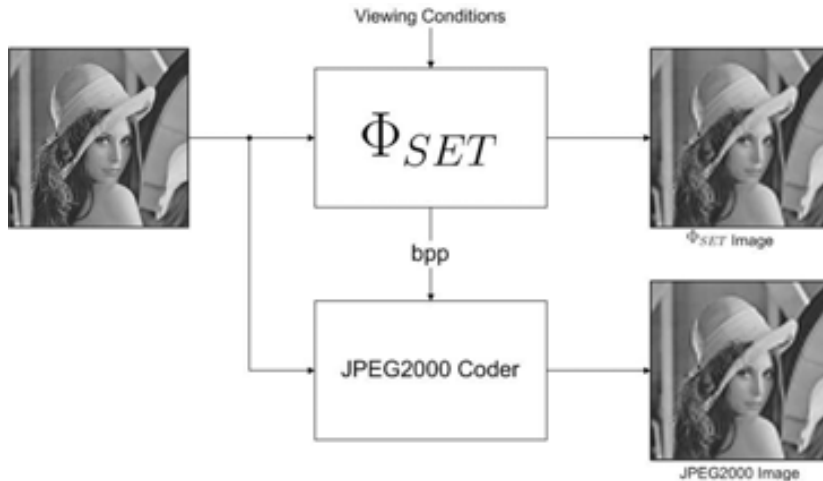
Figure 4.14: Process for comparing JPEG2000 and Φ_{SET}. Given some viewing conditions a Φ_{SET} compression is performed obtaining a particular bit-rate. Thus, a JPEG2000 compression is performed with such a bit-rate.

This experiment is performed across the CMU (Section A.5) and IVC (Section A.1) Image Databases. Image quality estimations are assessed by the thirteen metrics before mentioned, but in this section only $\mathcal{C}_w$PSNR results are exposed, the remaining metrics are shown in Sections C.2.1 and C.2.2 for the CMU and IVC Image Databases, respectively.

The parameters for estimating the $\mathcal{C}_w$PSNR assessment are: $d = 8H$, $M_{size} = 19''$, $h_{res} = 1280$ and $v_{res} = 1024$.

(a) CMU Image Database   (b) IVC Image Database
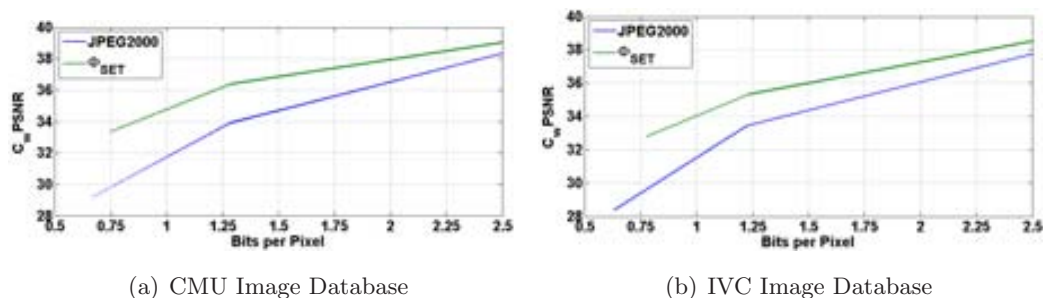
Figure 4.15: Comparison between Φ_{SET} and JPEG2000 image coders. Compression rate vs $\mathcal{C}_w$PSNR perceptual image quality, of (a) the CMU and (b) IVC image databases.

Fig. 4.15(a) shows the perceptual quality, estimated by $\mathcal{C}_w$PSNR, of the recovered color images both for JPEG2000 and Φ_{SET} as a function of their compression rate. For this experiment, we employ the CMU Image Database (Section A.5) and the *Kakadu* implementation for JPEG2000 compression(50). On the average, a color image with $\mathcal{C}_w$PSNR=36 dB is compressed by JPEG2000 coder (blue function) at 2.00 bpp (1:12 compression ratio) in 64 KBytes and by Φ_{SET} (green function) at 1.50 bpp (1:16 ratio) in 48 KBytes. In Figure 4.16, we can see these differences when images *Lenna*, *Girl2* and *Tiffany* are compressed at 0.92 bpp, 0.54 bpp and 0.93 bpp, respectively, by JPEG2000 and Φ_{SET}. Thus, on the average for this image database, Φ_{SET} is 2.38 dB better than JPEG2000.

Fig. 4.15(b) shows the perceptual quality, estimated by $\mathcal{C}_w$PSNR, of the recovered color images both for JPEG2000 and Φ_{SET} as a function of their compression rate. For this experiment, we employ the IVC Image Database (Section A.1) and the *JJ2000* implementation for JPEG2000 compression(40). On the average, a color image compressed at 1.5 bpp (1:16 ratio, stored in 48 KBytes) by JPEG2000 coder (blue function) has $\mathcal{C}_w$PSNR=34.70 dB of perceptual image quality and by Φ_{SET} (green function) has $\mathcal{C}_w$PSNR=36.86 dB. In Figure 4.17, we can see these differences when images *Barbara*, *Mandrill* and *Clown* are compressed at 0.76 bpp, 1.15 bpp and 0.96 bpp, respectively, by JPEG2000 and Φ_{SET}. Thus, on the average for this image database, Φ_{SET} is 2.33 dB better than JPEG2000.

4.7   Conclusions

We defined both forward (F-ρSQ) and inverse (I-ρSQ) perceptual quantizer using
CIWaM. We incorporated it to Hi-SET, proposing the perceptual image compres-
sion system Φ_{SET}. In order to measure the effectiveness of the perceptual quantization,
a performance analysis is done using thirteen assessments such as PSNR, MSSIM, VIF,
WSNR or $\mathcal{C}_w$PSNR, for instance, which measured the image quality between recon-
structed and original images. The experimental results show that the solely usage of
the Forward Perceptual Quantization improves the JPEG2000 compression and im-
age perceptual quality. In addition, when both Forward and Inverse Quantization are
applied into Hi-SET, it significatively improves the results regarding the JPEG2000
compression.

(a) JPEG2000, $\mathcal{C}_w$PSNR=34.40 dB

(b) Φ_{SET}, $\mathcal{C}_w$PSNR=37.81 dB

(c) JPEG2000, $\mathcal{C}_w$PSNR=33.42 dB

(d) Φ_{SET}, $\mathcal{C}_w$PSNR=38.22 dB

(e) JPEG2000, $\mathcal{C}_w$PSNR=32.88 dB

(f) Φ_{SET}, $\mathcal{C}_w$PSNR=38.10 dB

Figure 4.16: Example of reconstructed color images *Lenna*, *Girl2* and *Tiffany* of the CMU image database compressed at (a-b) 0.92 bpp, (c-d) 0.54 bpp and (e-f) 0.93 bpp, respectively.

(a) JPEG2000, $\mathcal{C}_w$PSNR=30.87 dB

(b) Φ_{SET}, $\mathcal{C}_w$PSNR=31.69 dB
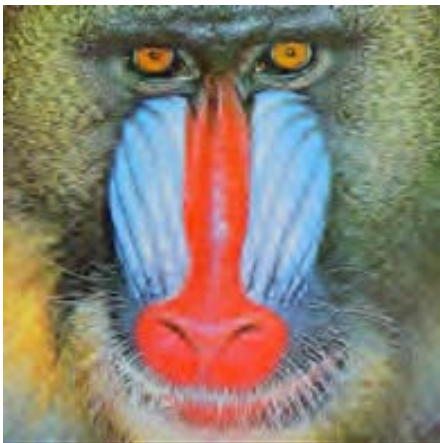
(c) JPEG2000, $\mathcal{C}_w$PSNR=27.71 dB
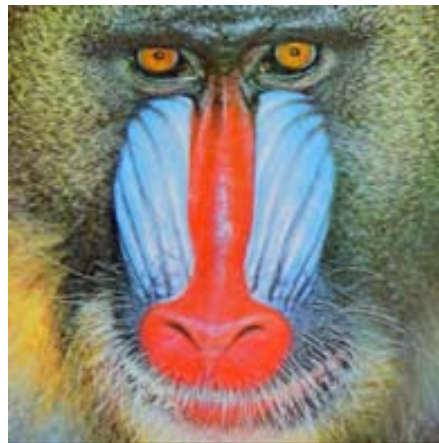
(d) Φ_{SET}, $\mathcal{C}_w$PSNR=28.86 dB

(e) JPEG2000, $\mathcal{C}_w$PSNR=31.74 dB

(f) Φ_{SET}, $\mathcal{C}_w$PSNR=33.19 dB

Figure 4.17: Example of reconstructed color images *Barbara*, *Mandrill* and *Clown* of the IVC image database compressed at (a-b) 0.76 bpp, (c-d) 1.15 bpp and (e-f) 0.96 bpp, respectively.

Chapter 5

Perceptual Generalized Bitplane-by-Bitplane Shift

5.1 Introduction

Region of interest (ROI) image coding is a feature that modern image coder have, which allows to encode an specific region with better quality than the rest of the image or background (BG). ROI coding is one of the requirements in the JPEG2000 image coding standard (10, 11, 48, 51), which defines two ROI methods(4, 13, 30, 31, 51):

1. Based on general scaling

2. Maximum shift (MaxShift)

The general ROI scaling-based method scales coefficients in such a way that the bits associated with the ROI are shifted to higher bitplanes than the bitplanes associated with the background, as shown in Figure 5.1(b). It implies that during a embedded coding process, any background bitplane of the image is located after the most significant ROI bitplanes into the bit-stream. But, in some cases, depending on the scaling value, φ, some bits of ROI are simultaneously encoded with BG. Therefore, this method allows to decode and refine the ROI before the rest of the image. No matter φ, it is posible to reconstruct with the entire bitstream a highest fidelity version of the whole image. Nevertheless, If the bitstream is terminated abruptly, the ROI will have a higher fidelity than BG.

The scaling-based method is implemented in five steps:

5. PERCEPTUAL GENERALIZED BITPLANE-BY-BITPLANE SHIFT

(a) No ROI coding   (b) Scaling Based Method with $\varphi = 3$
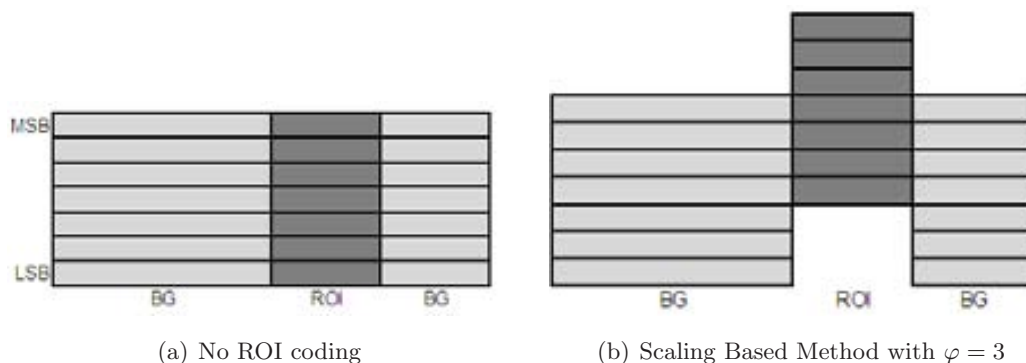
Figure 5.1: Scaling based ROI coding method. Background is denoted as BG and Region of Interest as ROI. MSB is the most significant bitplane and LSB is the least significant bitplane.

1. A wavelet transform of the original images is performed.

2. A ROI mask is defined, indicating the set of coefficients that are necessary for reaching a lossless ROI reconstruction, Figure 5.2.

3. Wavelet coefficients are quantized and stored in a sign magnitude representation, using the most significant part of the precision. It will allow to downscale BG coefficients.

4. A specified scaling value, $\tilde{\varphi}$, downscales the coefficients inside the BG.

5. The most significant bitplanes are progressively entropy encoded.
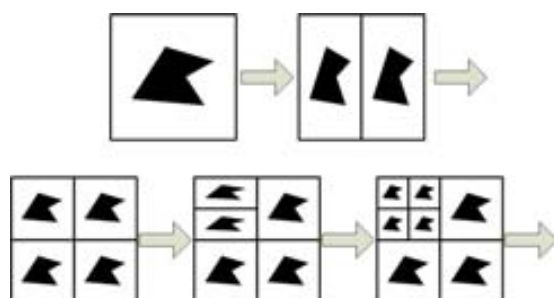
Figure 5.2: ROI mask generation, wavelet domain.

The input of ROI scaling-based method is the scaling value φ, while MaxShift method calculates it. Hence, the encoder defines from quantized coefficients this scaling

value such that:

$$\varphi = \lceil \log_2 \left(\max \left\{ \mathcal{M}_{\mathcal{BG}} \right\} + 1 \right) \rceil \tag{5.1}$$

where $\max \left\{ \mathcal{M}_{\mathcal{BG}} \right\}$ is the maximum coefficient in the BG. Thus, when ROI is scaled up φ bitplanes, the minimum coefficient belonging to ROI will be place one bitplane up of BG (Fig. 5.3). Namely, 2^φ is the smallest integer that is greater than any coefficient in the BG. MaxShift method is shown in Figure 5.3. Bitplane mask (BP_{mask}) will be explained in section 5.2.2.
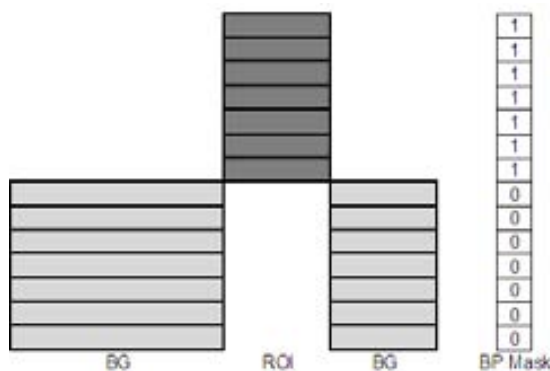
Figure 5.3: MaxShift method, $\varphi = 7$. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP_{mask}.

At the decoder side, the ROI and BG coefficients are simply identified by checking the coefficient magnitudes. All coefficients that are higher or equal than the φth bitplane belong to the ROI otherwise they are a part of BG. Hence, it is not important to transmit the shape information of the ROI or ROIs to the decoder. The ROI coefficients are scaled down φ bitplanes before inverse wavelet transformation is applied.

5.2 Related Work

5.2.1 BbBShift

Wang and Bovik proposed the bitplane-by-bitplane shift (BbBShift) method in (60). BbBShift shifts bitplanes on a bitplane-by-bitplane strategy. Figure 5.4 shows an illustration of the BbBShift method. BbBShift uses two parameters, φ_1 and φ_2, whose sum is equal to the number of bitplanes for representing any coefficient inside the image, indexing the top bitplane as bitplane 1.
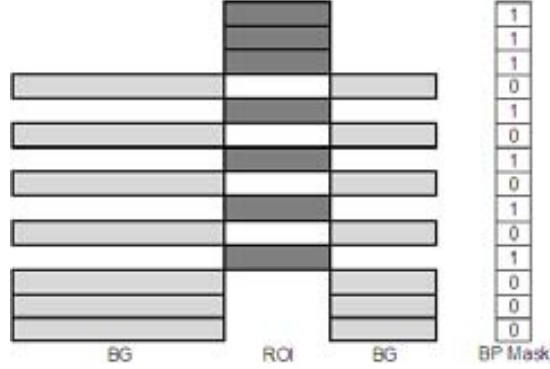
Figure 5.4: BbBShift ROI coding method, $\varphi_1 = 3$ and $\varphi_2 = 4$. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP_{mask}.

The encoding process of the BbBShift method is defined as:

1. For a given bitplane bpl with at least one ROI coefficient:

   - If $bpl \leq \varphi_1$, bpl is not shifted.

   - If $\varphi_1 < bpl \leq \varphi_1 + \varphi_2$, bpl is shifted down to $\varphi_1 + 2\,(bpl - \varphi_1)$

2. For a given bitplane bpl with at least one BG coefficient:

   - If $bpl \leq \varphi_2$, bpl is shifted down to $\varphi_1 + 2bpl - 1$

   - If $bpl > \varphi_2$, bpl is shifted down to $\varphi_1 + \varphi_2 + bpl$

Summarizing, the BbBShift method encodes the first φ_1 bitplanes with ROI coefficients, then, BG and ROI bitplanes are alternately shifted, refining gradually both ROI and BG of the image (Fig. 5.4).

5.2.2 GBbBShift

In practice, the quality refinement pattern of the ROI and BG used by BbBShift method is similar to the general scaling based method. Thus, when the image is encoded and this process is truncated in a specific point the quality of the ROI is high while there is no information of BG.

Hence, Wang and Bovik (56) modified BbBShift method and proposed the generalized bitplane-by-bitplane shift (GBbBShift) method, which introduces the option to improve visual quality either of ROI or BG or both. Figure 5.5 shows that with

GBbBShift method it is possible to decode some bitplanes of BG after the decoding of same ROI bitplanes. It allows to improve the overall quality of the recovered image. This is posible gathering BG bitplanes. Thus, when the encoding process achieves the lowest bitplanes of ROI, the quality of BG could be good enough in order to portray an approximation of BG.
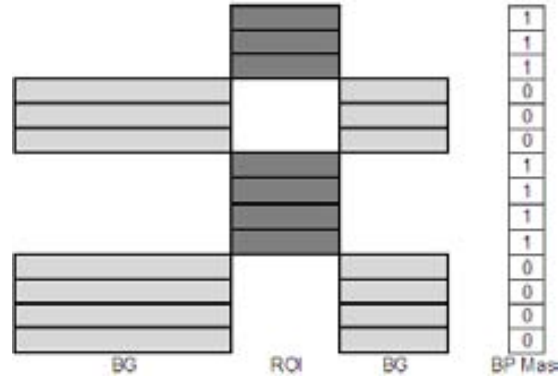
Figure 5.5: GBbBShift ROI coding method. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP_{mask}.

Therefore, the main feature of GBbBShift is to give the opportunity to arbitrary chose the order of bitplane decoding, grouping them in ROI bitplanes and BG bitplanes. This is posible using a binary bitplane mask or BP_{mask}, which contains one bit per each bitplane, that is, twice the amount of bitplanes of the original image. A ROI bitplane is represented by 1, while a BG bitplane by 0. For example, the BP_{mask} for MaxShift method in Figure 5.3 is 11111110000000, while for BbBShift in Figure 5.4 and GBbBShift in Figure 5.5 are 11101010101000 and 11100011110000, respectively.

At the encoder side, the BP_{mask} has the order of shifting both the ROI and BG bitplanes. Furthermore, BP_{mask} is encoded in the bitstream, while the scaling values φ or φ_1 and φ_2 from the MaxShift and BbBShift methods, respectively, have to be transmitted.

5.3 ρGBbBShift Method

In order to have several kinds of options for bitplane scaling techniques, a perceptual generalized bitplane-by-bitplane shift(ρGBbBShift) method is proposed. The ρGBbBShift method introduces to the GBbBShift method perceptual criteria when

bitplanes of ROI and BG areas are shifted. This additional feature is intended for balancing perceptual importance of some coefficients regardless their numerical importance and for not observing visual difference at ROI regarding MaxShift method, improving perceptual quality of the entire image.
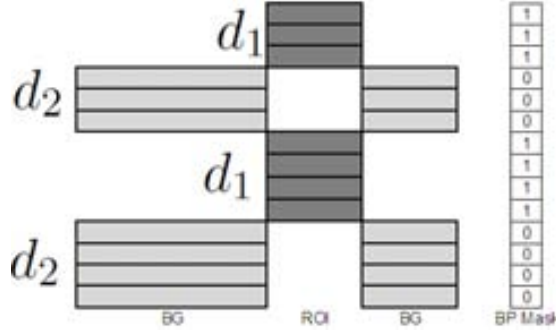
Figure 5.6: ρGBbBShift ROI coding method. Background is denoted as BG (perceptually quantized by ρSQ at d_2), Region of Interest as ROI (perceptually quantized at d_1 by ρSQ)and Bitplane mask as BP_{mask}.

Thus, ρGBbBShift uses a binary bitplane mask or BP_{mask} in the same way that GBbBShift (Figure 5.6). At the encoder, shifting scheme is as follows:

1. Calculate φ using Equation 5.1.

2. Verify that the length of BP_{mask} is equal to 2φ.

3.
   - For all ROI Coefficients, forward perceptual quantize them using Equation 4.1 (F-ρSQ) with viewing distance d_1.
   - For all BG Coefficients, forward perceptual quantize them using Equation 4.1(F-ρSQ) with viewing distance d_2, being $d_2 \gg d_1$.

4. Let τ and η be equal to 0.

5. For every element i of BP_{mask}, starting with the least significant bit:
   - If $BP_{mask}(i) = 1$, Shift up all ROI perceptual quantized coefficients of the $(\varphi - \eta)$-th bitplane by τ bitplanes and increment η.
   - Else: Shift up all BG perceptual quantized coefficients of the $(\varphi - \tau)$-th bitplane by η bitplanes and increment τ.

At the decoder, shifting scheme is as follows:

1. Let $\varphi = \frac{length\ of\ BP_{mask}}{2}$ be calculated.

2. Let τ and η be equal to 0.

3. For every element i of BP_{mask}, starting with the least significant bit:

   - If $BP_{mask}(i) = 1$, Shift down all perceptual quantized coefficients by τ bit-planes, which pertain to the $(2\varphi - (\tau + \eta))$-th bitplane of the recovered image and increment η.

   - Else: Shift down all perceptual quantized coefficients by η bitplanes, which pertain to the $(2\varphi - (\tau + \eta))$-th bitplane of the recovered image and increment τ.

4. Let us denote as $c_{i,j}$ a given non-zero wavelet coefficient of the recovered image with 2φ bitplanes and $\bar{c}_{i,j}$ as a shifted down c obtained in the previous step, with φ bitplanes.

   - If $(c_{i,j}\ \&\ BP_{mask}) > 0$, inverse perceptual quantize $\bar{c}_{i,j}$ using Equation 4.2(I-ρSQ) with d_1 as viewing distance.

   - If $(c_{i,j}\ \&\ BP_{mask}) = 0$, inverse perceptual quantize $\bar{c}_{i,j}$ using Equation 4.2(I-ρSQ) with d_2 as viewing distance.

5.4 Experimental Results

The ρGBbBShift method, as the other methods presented here, can be applied to many image compression algorithms such as JPEG2000 or Hi-SET. We test our method applying it to Hi-SET and the results are contrasted with MaxShift method in JPEG2000 and Hi-SET. The setup parameters are $\varphi = 8$ for MaxShift and $BP_{mask} = 1111000110110000$, $d_1 = 5H$ and $d_2 = 50H$, where H is picture height (512 pixels) in a 19-inch LCD monitor, for ρGBbBShift. Also, we use the *JJ2000* implementation when an image is compressed by JPEG2000 standard(40).

5.4.1 Experiments

Figure 5.7 shows a comparison among methods MaxShift and GBbBShift applied to JPEG2000, in addition to, ρGBbBShift applied to Hi-SET. The 24-bpp image *Barbara* is compressed at 0.5 bpp.

(a) MaxShift in JPEG2000 coder, 0.5 bpp

(b) GBbBShift in JPEG2000 coder, 0.5 bpp   (c) ρGBbBShift in Hi-SET coder, 0.5 bpp

Figure 5.7: 512×640 pixel Image *Barbara* with 24 bits per pixel. ROI is a patch of the image located at [341 280 442 442], whose size is 1/16 of the image. Decoded images at 0.5 bpp using MaxShift method in JPEG2000 coder((a) $\varphi = 8$), GBbBShift method in JPEG2000 coder ((b)$BP_{mask} = 1111000110110000$) and ρGBbBShift method in Hi-SET coder ((c)$BP_{mask} = 1111000110110000$).

It can be observed that without visual difference at ROI, the ρGBbBShift method provide better image quality at the BG than the general based methods defined in

JPEG2000 Part II(11).

In order to better qualify the performance of MaxShift, GBbBShift and ρGBbBShift methods, first, we compared these methods applied to the Hi-SET coder and then, we compare MaxShift and ρGBbBShift methods applied to the JPEG2000 standard and Hi-SET, respectively .We compress two different gray-scale and color images of *1600*, from CSIQ image database (Fig A.4), and *Lenna* at different bit-rates. ROI area is a patch at the center of these images, whose size is 1/16 of the image.
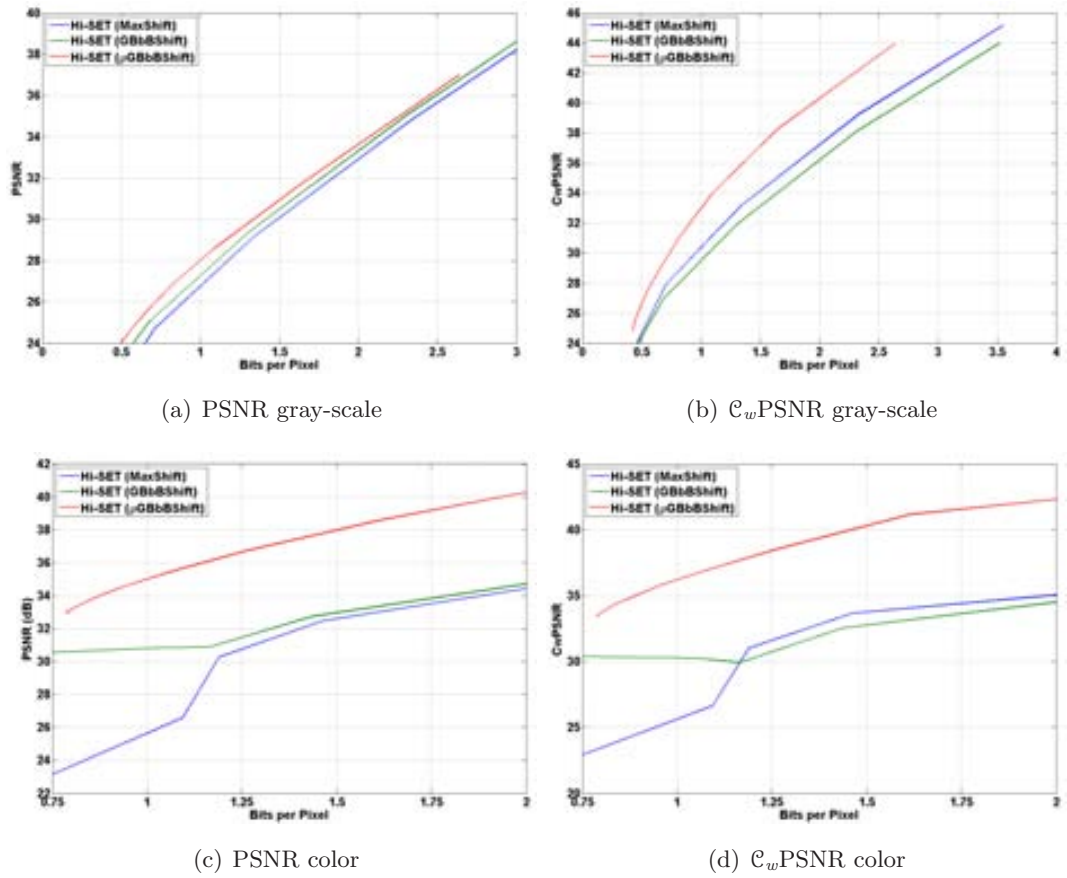

(a) PSNR gray-scale

(b) $\mathcal{C}_w$PSNR gray-scale

(c) PSNR color

(d) $\mathcal{C}_w$PSNR color

Figure 5.8: Comparison among MaxShift(Blue Function), GBbBShift(Green Function) and ρGBbBShift(Red Function) methods applied to Hi-SET coder. 512×512 pixel Image *1600* with (a-b) 8 and (c-d) 24 bits per pixel are employed for this experiment. ROI is a patch at the center of the image, whose size is 1/16 of the image. The overall image quality of decoded images at different bits per pixel are contrasted both (a and c) objectively and (b and d) subjectively.

Figure 5.8 shows the comparison among MaxShift(Blue Function), GBbBShift(Green

Function) and ρGBbBShift(Red Function) methods applied to Hi-SET coder. 512×512 pixel Image *1600* both for gray-scale and color are employ for this experiment. These Figures also show that the ρGBbBShift method gets the better results both in PSNR(objective image quality) and $\mathcal{C}_w$PSNR(subjective image quality) in contrast to MaxShift and GBbBShift methods.
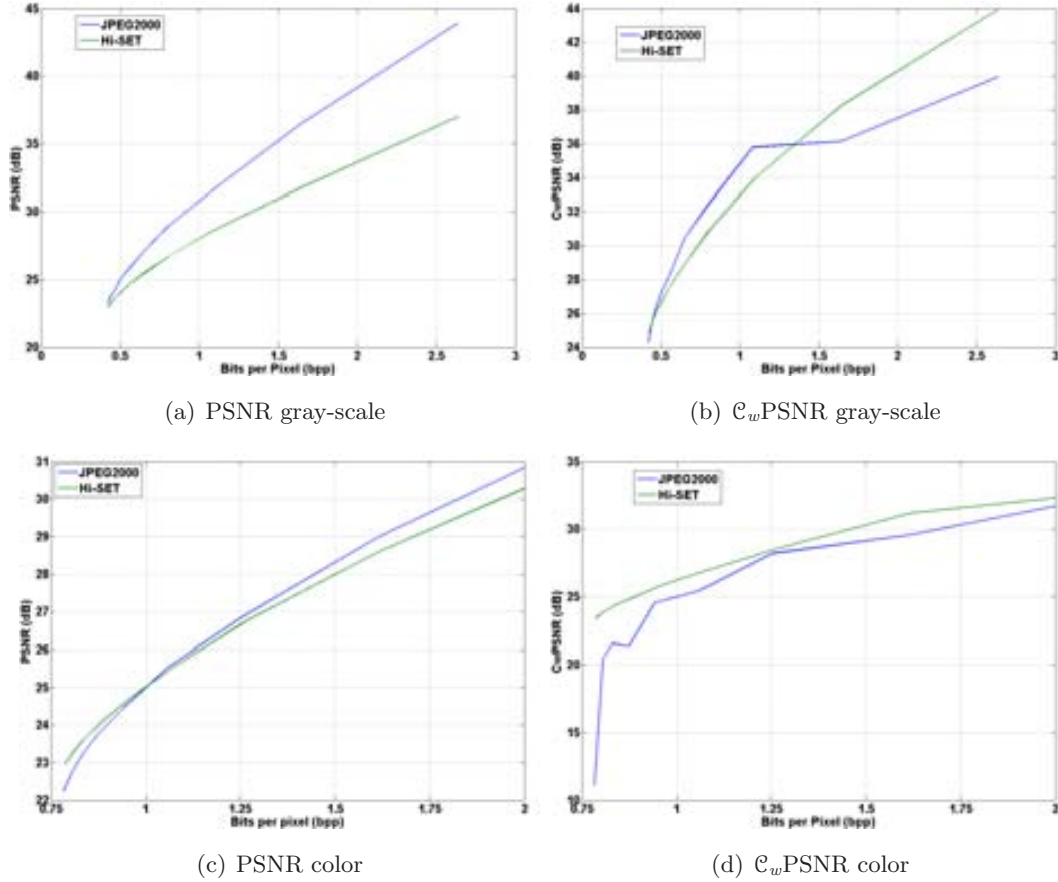

<table>
<tr><td>(a) PSNR gray-scale</td><td>(b) $\mathcal{C}_w$PSNR gray-scale</td></tr>
<tr><td>(c) PSNR color</td><td>(d) $\mathcal{C}_w$PSNR color</td></tr>
</table>

Figure 5.9: Comparison between MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to Hi-SET coder. 512×512 pixel Image *1600* with (a-b) 8 and (c-d) 24 bits per pixel are employed for this experiment. ROI is a patch at the center of the image, whose size is 1/16 of the image. The overall image quality of decoded images at different bits per pixel are contrasted both (a and c) objectively and (b and d) subjectively .

When MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to Hi-SET coder are compared, in the whole image quality assessment of image *1600*, JPEG2000 obtains better objective quality both for gray-scale and color images (Fig-

ures 5.9(a) and 5.9(c), respectively). But when the subjective quality is estimated ρGBbBShift coded images are perceptually better.

A visual example is depicted by Figure 5.10, where it can be shown that there is no perceptual difference between ROI areas besides the perceptual image quality at BG is better when ρGBbBShift is applied to the Hi-SET coder (Fig. 5.10(d)). Furthermore, Figs. 5.10(b) and 5.10(c) show the examples when MaxShift and GBbBShift methods, respectively, are applied to the Hi-SET coder.

Similarly, when a ROI area is defined in Image *Lenna*, Fig. 5.11 shows the comparison among MaxShift(Blue Function), GBbBShift(Green Function) and ρGBbBShift(Red Function) methods applied to Hi-SET coder. 512×512 pixel Image *Lenna* both for gray-scale and color are employ for this experiment. These Figures also show that the ρGBbBShift method gets the better results both in PSNR(objective image quality) and $\mathcal{C}_w$PSNR(subjective image quality) in contrast to MaxShift and GBbBShift methods. In addition, When MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to Hi-SET coder are compared, ρGBbBShift obtains less objective quality (Figures 5.12(a) and 5.12(c)), but better subjective quality both for gray-scale and color images (Figures 5.12(b) and 5.12(d), respectively).

Figure 5.13 shows a visual example, when image *Lenna* is compressed at 0.34 bpp by JPEG2000 and Hi-SET. Thus, it can be observed that ρGBbBShift provides an important perceptual difference regarding the MaxShift method(Fig. 5.13(d)). Furthermore, Figs. 5.13(b) and 5.13(c) show the examples when MaxShift and GBbBShift methods, respectively, are applied to the Hi-SET coder.

5.4.2 Application in other image compression fields

The usage of ROI coded images depends on an specific application, but in some fields such as manipulation and transmission of images is important to enhance the image quality of some areas and to reduce it in others(7, 15). In Telemedicine or in Remote Sensing (RS) it is desirable to maintain the best quality of the ROI area, preserving relevant information of BG, namely the most perceptual frequencies.

Thus, in medical applications an image is by itself a ROI_ϕ area of the human body, a mammography is an area of chest, for instance. That is why, it is important to know where is this ROI_ϕ located, in order to ease the interpretation of a given ROI

(a) MaxShift method in JPEG2000 coder, 0.42 bpp

(b) MaxShift method in H*i*-SET coder, 0.42 bpp.

(c) GBbBShift method in H*i*-SET coder, 0.42 bpp.

(d) ρGBbBShift method in H*i*-SET coder, 0.42 bpp

Figure 5.10: 512×512 pixel Image *1600* from CSIQ image database with 8 bits per pixel. ROI is a patch at the center of the image, whose size is 1/16 of the image. Decoded images at 0.42 bpp using $\varphi = 8$ for MaxShift method (a) in JPEG2000 coder and (b) in H*i*-SET coder, and $BP_{mask} = 1111000110110000$ for (c) GBbBShift and (d) ρGBbBShift methods in H*i*-SET coder.

(a) PSNR gray-scale

(b) $\mathcal{C}_w$PSNR gray-scale
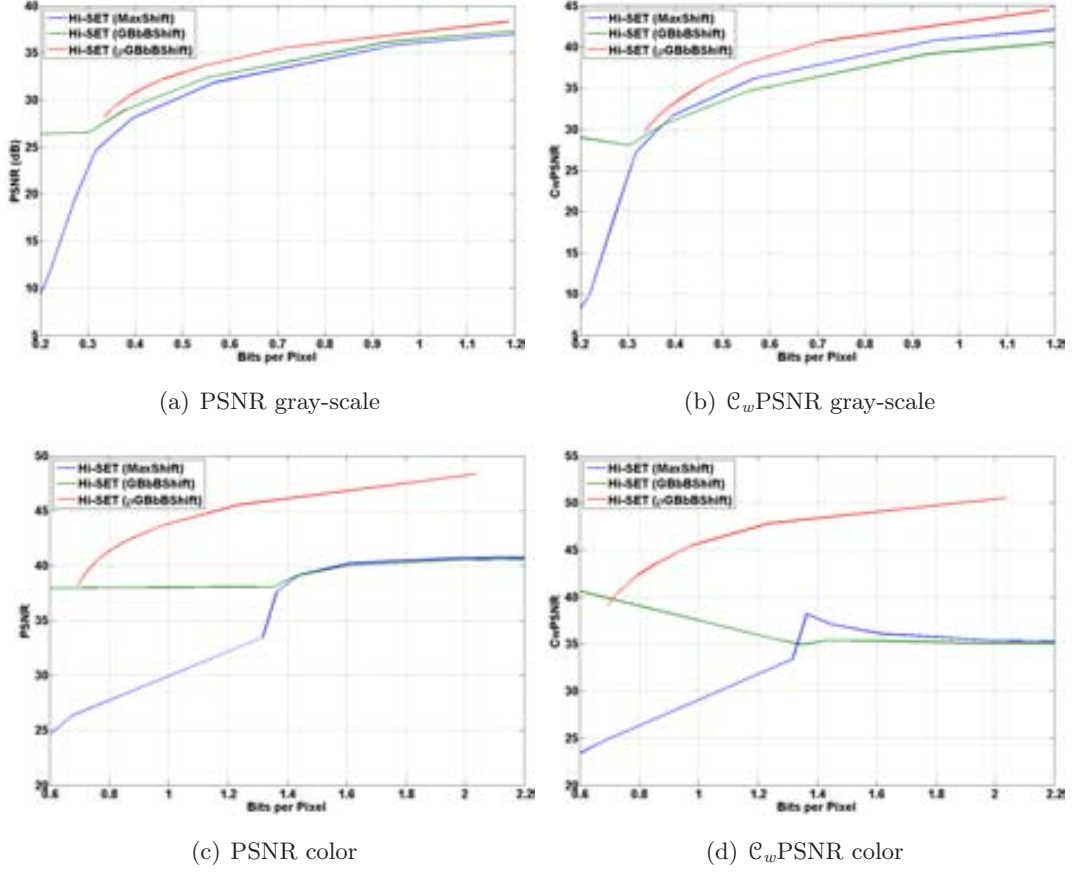
(c) PSNR color

(d) $\mathcal{C}_w$PSNR color

Figure 5.11: Comparison among MaxShift(Blue Function), GBbBShift(Green Function) and ρGBbBShift(Red Function) methods applied to Hi-SET coder. 512×512 pixel Image *Lenna* with (a-b) 8 and (c-d) 24 bits per pixel are employed for this experiment. ROI is a patch at the center of the image, whose size is 1/16 of the image. The overall image quality of decoded images at different bits per pixel are contrasted both (a and c) objectively and (b and d) subjectively.

coded image. In addition, according Federal laws in some countries, ROI areas must be lossless areas(62). ρGBbBShiftis able to accomplish these two features.

Figure 5.14 shows an example of medical application. A rectangular ROI of the Image *mdb202* from PEIPA image database(37) , coordinates [120 440 376 696], is coded at 0.12 bpp by JPEG2000 and Hi-SET, employing MaxShift and ρGBbBShift methods, respectively. The overall image quality measured by PSNR in Figure 5.14(a) (MaxShift method applied to JPEG2000) is 37.21 dB, while in Figure 5.14(c) (ρGBbBShift method applied to Hi-SET) is 36.76 dB. Again, PSNR does not reflect perceptual differences

(a) PSNR gray-scale

(b) $\mathcal{C}_w$PSNR gray-scale

(c) PSNR color

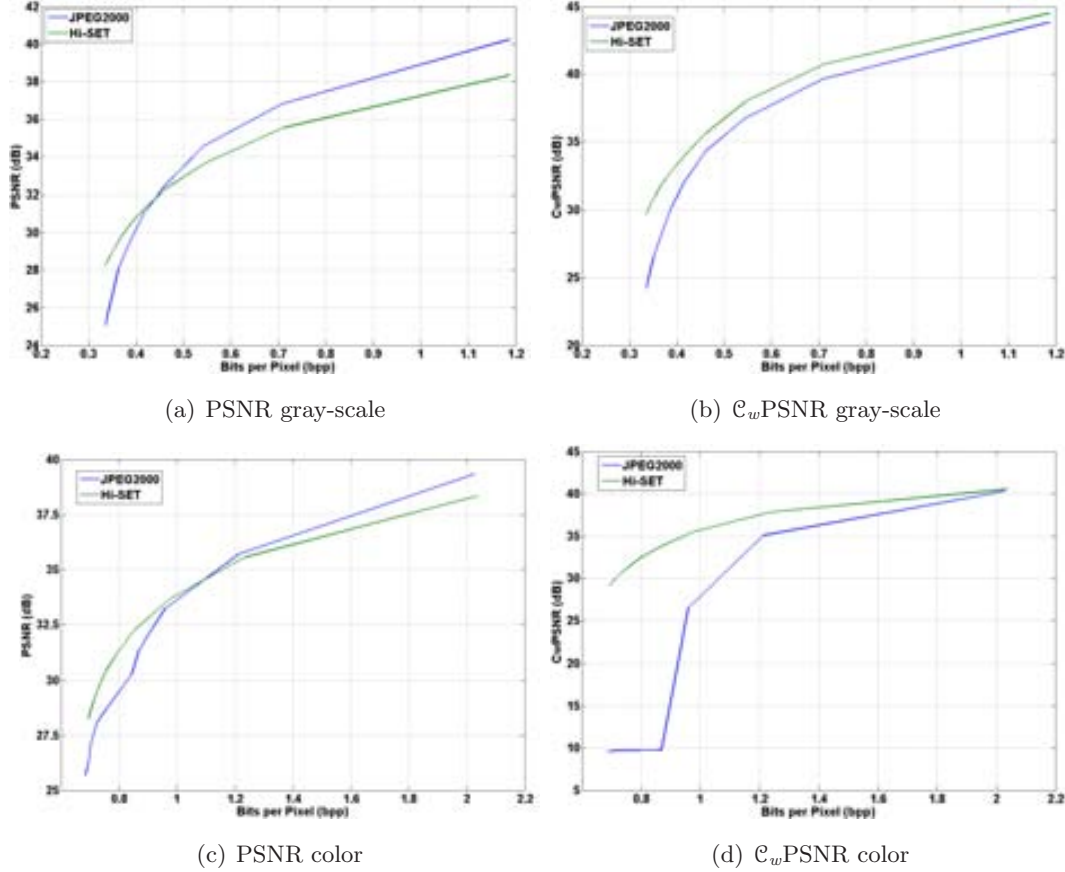(d) $\mathcal{C}_w$PSNR color

Figure 5.12: Comparison between MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to Hi-SET coder. 512×512 pixel Image *Lenna* with (a-b) 8 and (c-d) 24 bits per pixel are employed for this experiment. ROI is a patch at the center of the image, whose size is 1/16 of the image. The overall image quality of decoded images at different bits per pixel are contrasted both (a and c) objectively and (b and d)subjectively.

between images (Figures 5.14(b) and 5.14(d)). When perceptual metrics assess the image quality of the ρGBbBShift coded image, for example, VIFP=0.6359, WSNR=34.24 and $\mathcal{C}_w$PSNR=40.88, while for MaxShift coded image VIFP=0.3561, WSNR=31.34 and $\mathcal{C}_w$PSNR=37.18. Thus, these metrics predicts that there is an important perceptual difference between ROI methods, being ρGBbBShift method better than MaxShift method.

Remote Sensing Images (RSI) are widely used in agriculture, mapping, water conservancy, etc. An RSI database is usually very huge in size, since the saved images have abundant details. Thus, an important goal for compressing RSI is to code the images
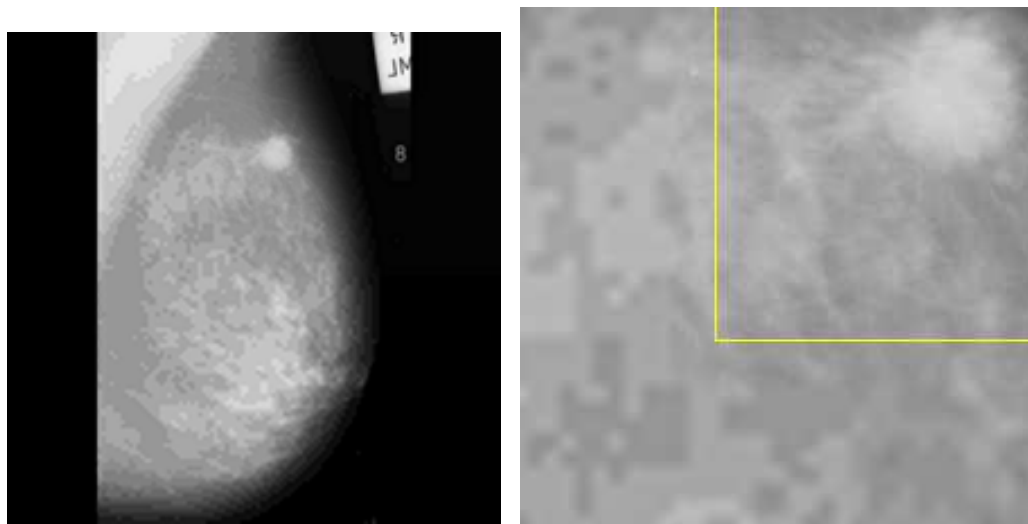
(a) MaxShift method in JPEG2000 coder, 0.34 bpp.

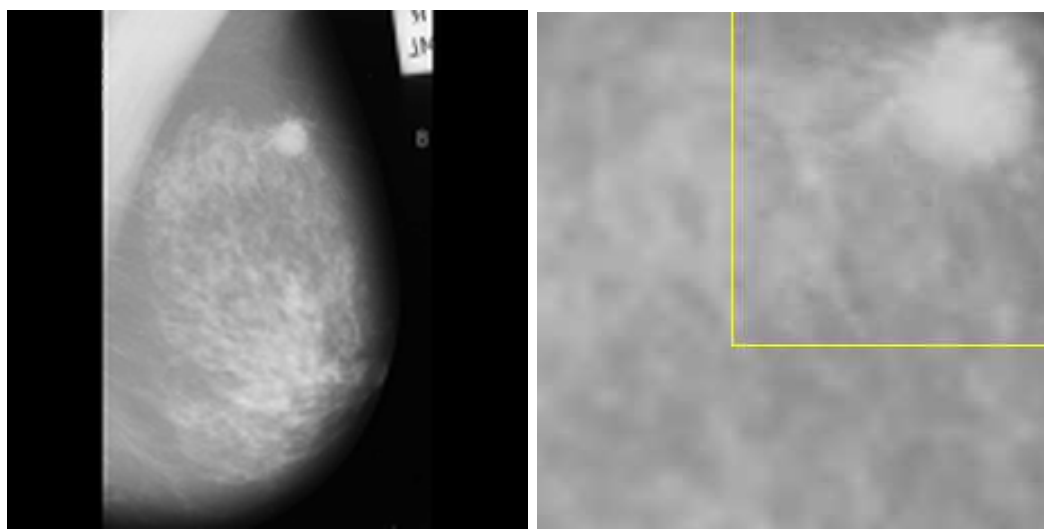(b) MaxShift method in Hi-SET coder, 0.34 bpp.

(c) GBbBShift method in Hi-SET coder, 0.34 bpp.

(d) ρGBbBShift method in Hi-SET coder, 0.34 bpp.

Figure 5.13: 512×512 pixel Image *Lenna* from CMU image database with 8 bits per pixel. ROI is a patch at the center of the image, whose size is 1/16 of the image. Decoded images at 0.34 bpp using $\varphi = 8$ for MaxShift method (a) in JPEG2000 coder and (b) in Hi-SET coder, and $BP_{mask} = 1111000110110000$ for (c) GBbBShift and (d) ρGBbBShift methods in Hi-SET coder.

(a) MaxShift method in JPEG2000 coder, 0.12 bpp

(b) Patch of (a) portrayed both ROI and BG areas.

(c) ρGBbBShift method in Hi-SET coder, 0.12 bpp

(d) Patch of (c) portrayed both ROI and BG areas.

Figure 5.14: Example of a medical application. 1024×1024 pixel Image $mdb202$ from PEIPA image database. ROI is a patch with coordinates [120 440 376 696], whose size is 1/16 of the image. Decoded images at 0.12 bpp using MaxShift method ((a-b) $\varphi = 8$) in JPEG2000 coder and ρGBbBShift method ((c-d)$BP_{mask} = 1111000110110000$) in H$i$-SET coder.

in advance, in order to transfer them and store them. However, only a small part of the image is useful and therefore some regions are only sketched(63).

Figure 5.15 shows an example of the application of ROI in Remote Sensing. Image *2.1.05*, from *Volumen 2: aerials* of USC-SIPI image database 8 bits per pixel(2), is compressed at 0.42 bpp. MaxShift method spends all the bit-ratio for coding ROI, located at [159 260 384 460], while ρGBbBShift balances a perceptually lossless ROI area with an acceptable representation of the BG. Hence, the overall image quality measured by PSNR in Figure 5.15(a) is 16.06 dB, while in Figure 5.15(b) is 24.28 dB. When perceptual metrics assess the image quality of the ρGBbBShift coded image, for example, VIFP=0.4982, WSNR=24.8469 and $\mathcal{C}_w$PSNR=27.07, while for MaxShift coded image VIFP=0.2368, WSNR=11.33 and $\mathcal{C}_w$PSNR=16.72. Thus, for this example, both PSNR and these subjective metrics reflect important perceptual differences between ROI methods, being ρGBbBShift method better than MaxShift method..

(a) MaxShift in JPEG2000 coder, 0.42 bpp

(b) ρGBbBShift method in H*i*-SET coder, 0.42 bpp

Figure 5.15: Example of a remote sensing application. 512×512 pixel Image *2.1.05* from *Volumen 2: aerials* of USC-SIPI image database at 8 bits per pixel. ROI is a patch with coordinates [159 260 384 460], whose size is 225×200 pixels. Decoded images at 0.42 bpp using MaxShift method ((a) $\varphi = 8$) in JPEG2000 coder and ρGBbBShift method ((b)$BP_{mask} = 1111000110110000$) in H*i*-SET coder.

5.5   Conclusions

A perceptual implementation of the Region of Interest, ρGBbBShift(), is proposed, which is a generalized method that can be applied to any wavelet-based compressor. We introduced ρGBbBShift method to the Hi-SET coder and it visually improves the results obtained by previous methods like MaxShift and GBbBShift. Our experiments show that ρGBbBShift into Hi-SET provides an important perceptual difference regarding the MaxShift method into JPEG2000, when it is applied not only to conventional images like *Lenna* or *Barbara*, but also to another image compression fields such as Telemedicine or Remote Sensing.

Chapter 6

Conclusions and Future work

The main goal of this thesis was to introduce perceptual criteria on two aspects of the image compression process. One the one hand, perceptual criteria was used on image quality estimation. On the other hand, these perceptual criteria were used to identify and to remove non-perceptual information of an image. These two aspects were used to propose a perceptual image compression system and an image quality assessment. Additionally, a new coder based on Hilbert Scanning (Hi-SET) is also presented.

6.1   Conclusions

In Chapter 2, we present a new metric for full-reference image quality based on perceptual weighting of PSNR by using a perceptual low-level model of the Human Visual System (CIWaM model). The proposed $\mathcal{C}_w$PSNR metrics is based on three concepts. First, the Relative Energy Ratio, measured at the point where an observer can better perceive differences among images, e.g. $\varepsilon\mathcal{R}\left(n\mathcal{P}\right)$, Sec. 2.3.2.1. This is a good enough approximation to image quality when different distorted versions of the same image are evaluated. Second, the distance D where the observer cannot perceive differences between the energies of distorted and reference images. The shorter it is, the better the quality of the distorted image. It is a good approximation to image quality when the same distortion is applied to several images. Finally, the generalization to any image and for JPEG and JPEG2000 distortions is performed by measuring the objective numerical quality (i.e. the PSNR) of the perceptual images predicted by CIWaM at D cm.

The $\mathcal{C}_w$PSNR assessment was tested in four well-known image databases such as TID2008, LIVE, CSIQ and IVC. It is the best-ranked image quality method in these databases for JPEG and JPEG2000 distortions when compared to several state-of-the-art metrics. Concretely, it is 2.5% and 1.5% better that MSSIM (the second best performing method) for JPEG and JPEG2000 distortions, respectively. $\mathcal{C}_w$PSNR significantly improves the correlation of PSNR with perceived image quality. On average, when $\mathcal{C}_w$PSNR is applied on the same distortion, it improves the results obtained by PSNR and MSE by 14% and 11.5%,respectively.

The Hi-SET coder, presented in Chapter 3, is based on Hilbert scanning of embedded quadTrees. It has low computational complexity and some important properties of modern image coders such as embedding and progressive transmission. This is achieved by using the principle of partial sorting by magnitude when a sequence of thresholds decreases. The desired compression rate can be controlled just by chunking the stream at the desired file length. When compared to other algorithms that use Hilbert scanning for pixel ordering, Hi-SET improves image quality by around 6.20 dB. Hi-SET achieves higher compression rates than JPEG2000 coder not only for high and medium resolution images but also for low resolution ones where it is difficult to find redundancies among spatial frequencies. Table 6.1 summarize the average improvements when compressing the TID2008 Image Database.

Table 6.1: Average PSNR(dB) improvement of Hi-SET in front of JPEG2000 for TID2008 image database.

Components	Y		YC_bC_r	
Resolution	*Low*	*Medium*	*Low*	*Medium*
Compression Ratio (bpp)	0.55	0.17	0.93	0.33
Image Quality (dB)	1.84	0.43	1.79	1.06

The Hi-SET coder improves the image quality of the JPEG2000 coder around PSNR=1.16 dB for gray-scale images and 1.43 dB for color ones. Furthermore, it saves around 0.245 bpp for high resolution gray-scale *Bicycle* images. We extended our experiments to another four image database such as CMU, CSIQ, IVC and LIVE. Thus,the results across these databases resulted Hi-SET improved the results of JPEG2000 not

only objectively but also metrics like MSSIM, UQI or VIF, which are perceptual indicators.

In Chapter 4 we defined both forward (F-ρSQ) and inverse (I-ρSQ) perceptual quantizer using CIWaM. We incorporated it to Hi-SET, proposing the perceptual image compression system Φ_{SET}. In order to measure the effectiveness of the perceptual quantization, a performance analysis is done using thirteen assessments such as PSNR, MSSIM, VIF, WSNR or $\mathcal{C}_w$PSNR, for instance, which measured the image quality between reconstructed and original images. The experimental results show that the solely usage of the Forward Perceptual Quantization improves the JPEG2000 compression and image perceptual quality. In addition, when both Forward and Inverse Quantization are applied into Hi-SET, it significatively improves the results regarding the JPEG2000 compression.

In Chapter 5, we propose a perceptual implementation of the Region of Interest, ρGBbBShift(), which is a generalized method that can be applied to any wavelet-based compressor. We introduced ρGBbBShift method to the Hi-SET coder and it visually improves the results obtained by previous methods like MaxShift and GBbBShift. Our experiments show that ρGBbBShift into Hi-SET provides an important perceptual difference regarding the MaxShift method into JPEG2000, when it is applied not only to conventional images like *Lenna* or *Barbara*, but also to another image compression fields such as Telemedicine or Remote Sensing.

6.2 Contributions

The main contribution od this Ph.D thesis are:

- Definition of a metrics that uses the loss of perceptual energy as a tool of assessing image quality. This indicator can be considered as a set of three gauges, which can be used for different purposes.

- Development of a image coder, which is a serious alternative of JPEG2000 exploiting the recursion of fractal, avoiding the massive storage of pixel coordinates.

- Development of a perceptual quantizer algorithm , unlike the JPEG2000 global Frequency weighting, our method quantizes locally, that is pixel-by-pixel. Similarly JPEG2000, it is not necessary to store the applied weighting for inverse quan-

tizing, this is because CIWaM properties permits to predict perceptual weighting *a posteriori.*

- Proposal of a new method for coding of Region of Interest areas, which can be applied to any wavelet based compression scheme.

These contributions show that CIWaM is a perceptual low-level model of HVS that helps in some areas of image compression field.

6.3 Future Work

$\mathcal{C}_w$PSNR is mainly developed for estimation of perceptual image quality, but its usage can be extended to other applications such as image quantization in image compression algorithms, optimizing the perceptual error under the constraint of a limited bit-budget. Since the CIWaM algorithm applies a perceptual weighting to every wavelet coefficient, it can quantize a particular coefficient during the bit allocation procedure, allowing to define a perceptual bit allocation algorithm. Hence, $\mathcal{C}_w$PSNR can be incorporated into embedded compression schemes such as EZW(44), SPIHT(41), JPEG2000 (48) and Hi-SET (26).

We are currently exploring extensions of $\mathcal{C}_w$PSNR to non-referenced or blind image quality assessment and perceptual rate allocation for the Hi-SET coder.

In addition to propose a image compression algorithm that makes use of a threshold based on the e-CSF properties, namely a threshold based on the perceptual importance of a coefficient, regardless of its numerical value.

Appendix A

Image Databases

A.1 Image and Video-Communication Image Database

IVC Database includes 10 original images (Fig. A.1) with 4 different distortions (JPEG, JPEG2000, LAR coding and Blurring) and 5 distortion degrees, that is, there are 50 degraded images by distortion(23).
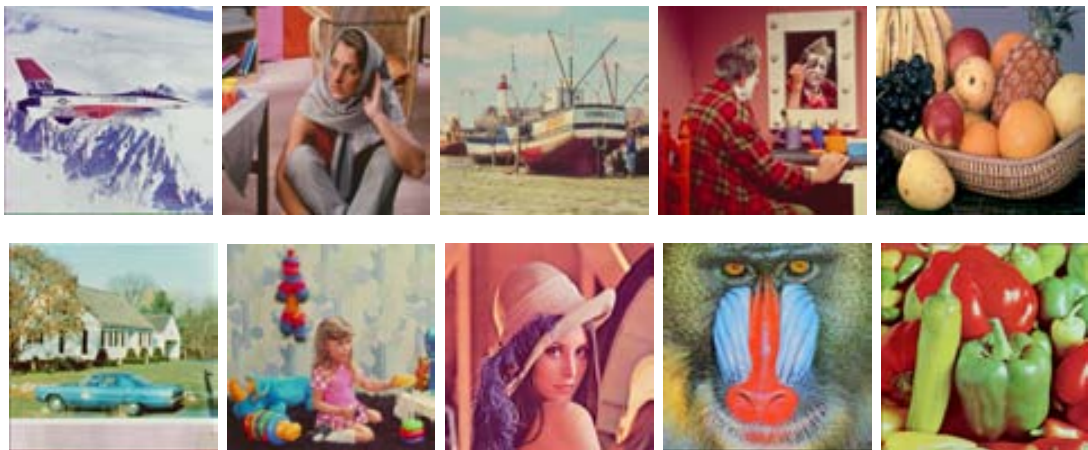
Figure A.1: Tested 512×512 pixel 24-bit color images, belonging to the IVC Image database.

A.2 Tampere Image Database

TID2008 Database contains 25 original images (Fig. A.2). They are distorted by 17 different types of distortions, and each distortion has 4 degrees of intensity, that is,

there are 68 distorted versions for every original image (38, 39).
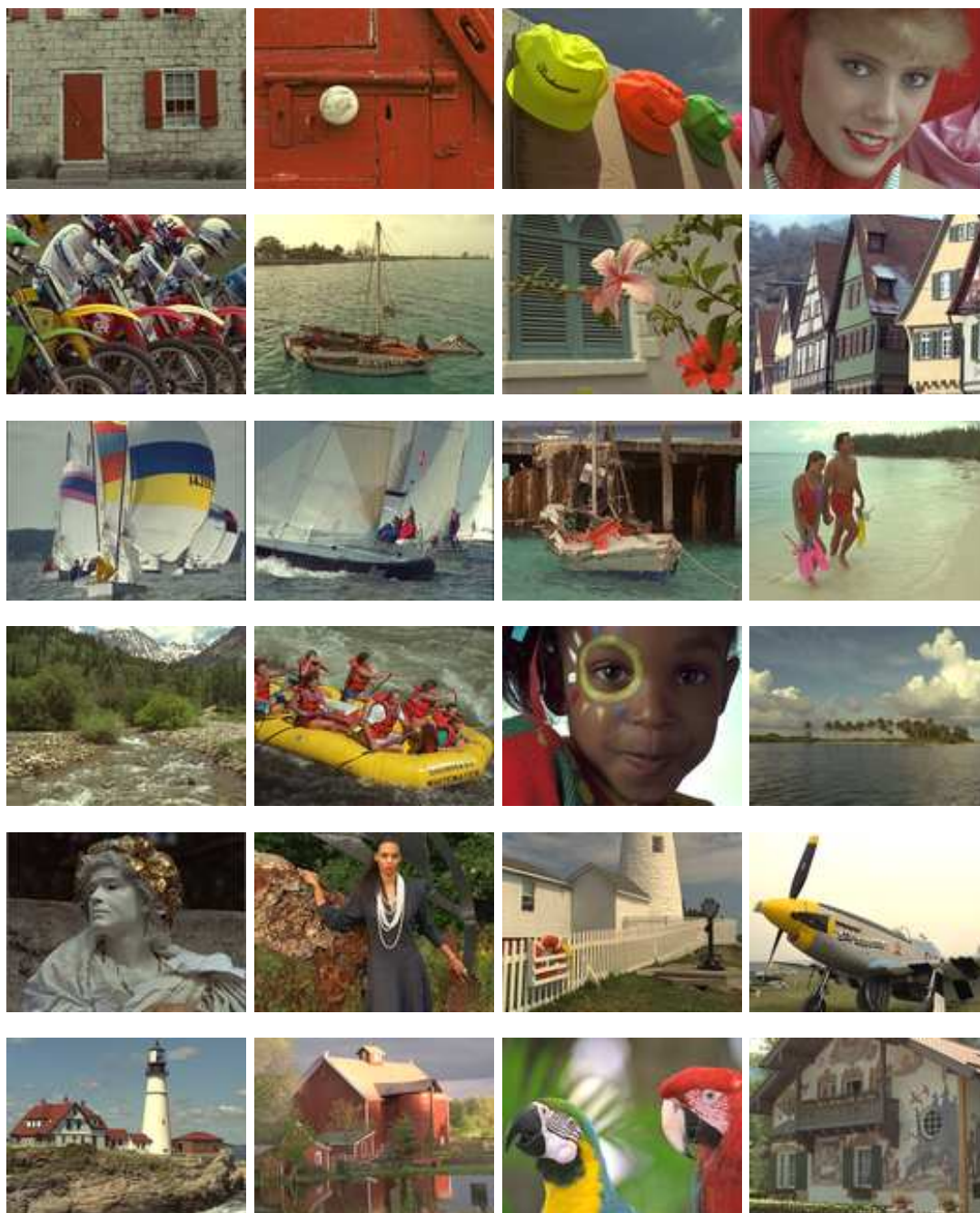
Figure A.2: Tested 512×384 pixel 24-bit color images, belonging to the Tampere test set.

A.3 Image Database of the Laboratory for Image and Video Engineering

LIVE Database contains 29 original images (Fig. A.3), with 26 to 29 altered versions for every original image. LIVE includes 234 and 228 distorted images for JPEG and JPEG2000 compression degradation, respectively(46).
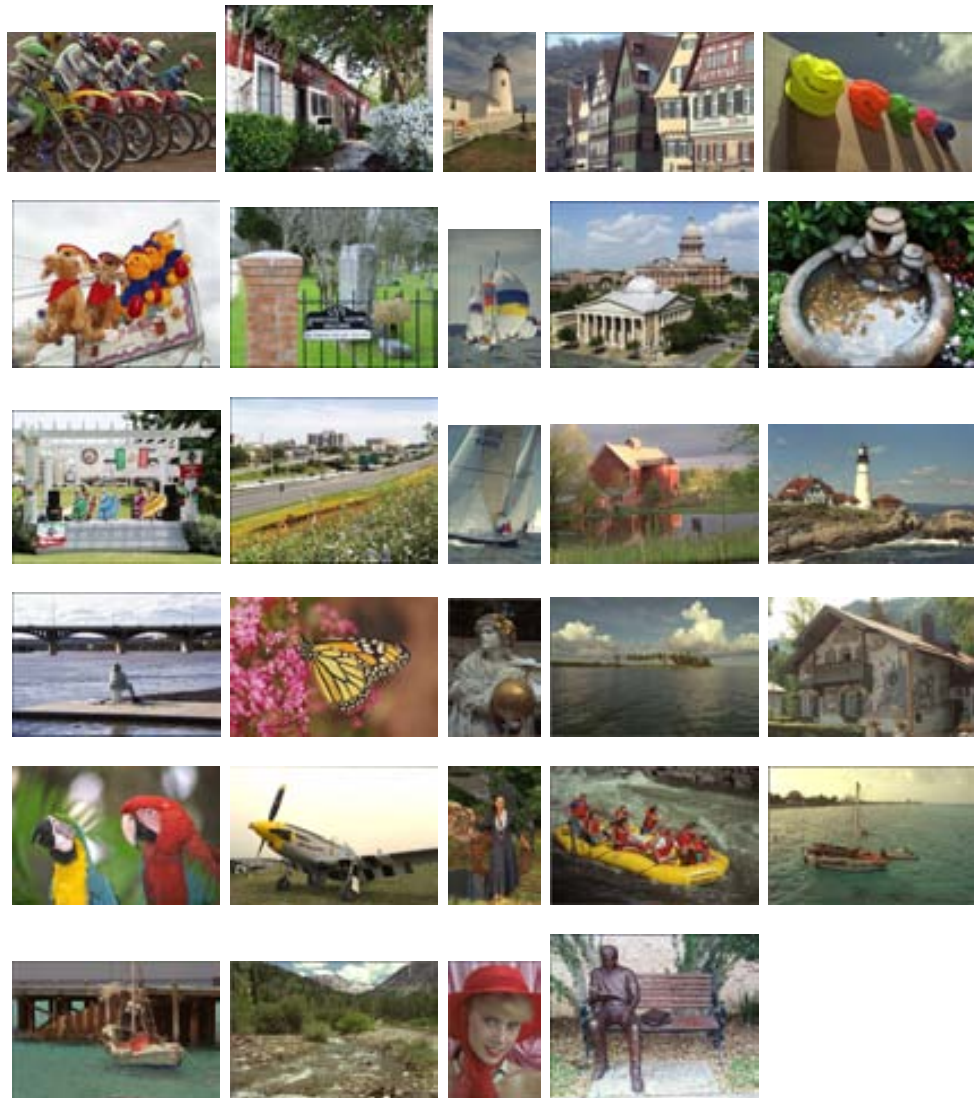
Figure A.3: Set of 29 tested images of 24-bit color, belonging to the LIVE Image database.

A.4 Categorical Subjective Image Quality Image Database

CSIQ Database includes 30 original images (Fig. A.4), which are distorted by 6 different types of distortions at 4 or 5 degrees. CSIQ Database has 5000 perceptual evaluations of 25 observers(22).
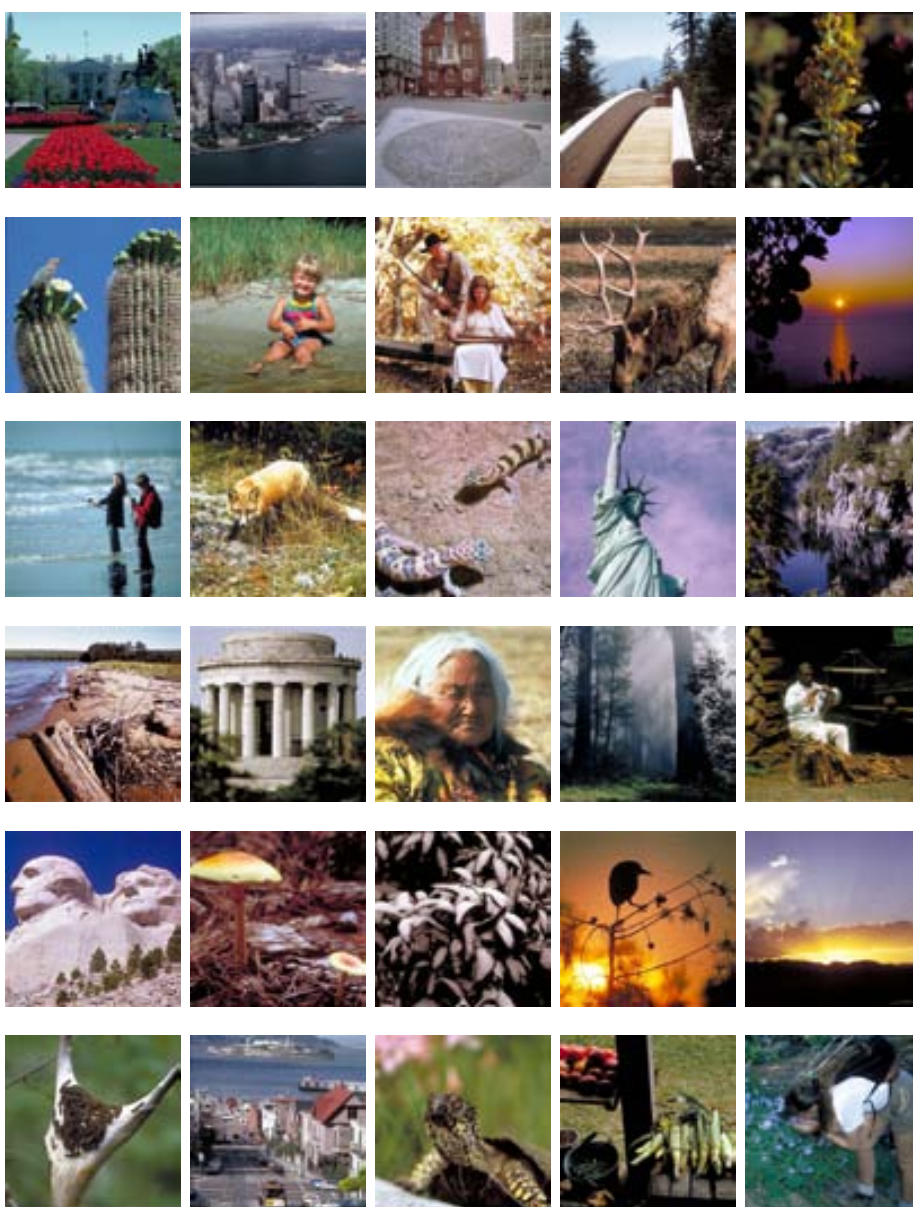
Figure A.4: Tested 512×512 pixel 24-bit color images, belonging to the CSIQ Image database.

A.5 University of Southern California Image Database

The University of Southern California Image Data Base, *Miscellaneous volume*(2). The database contains eight 256×256 pixel images (Figure A.5) and eight 512×512 pixel images (Figure A.6)(2).
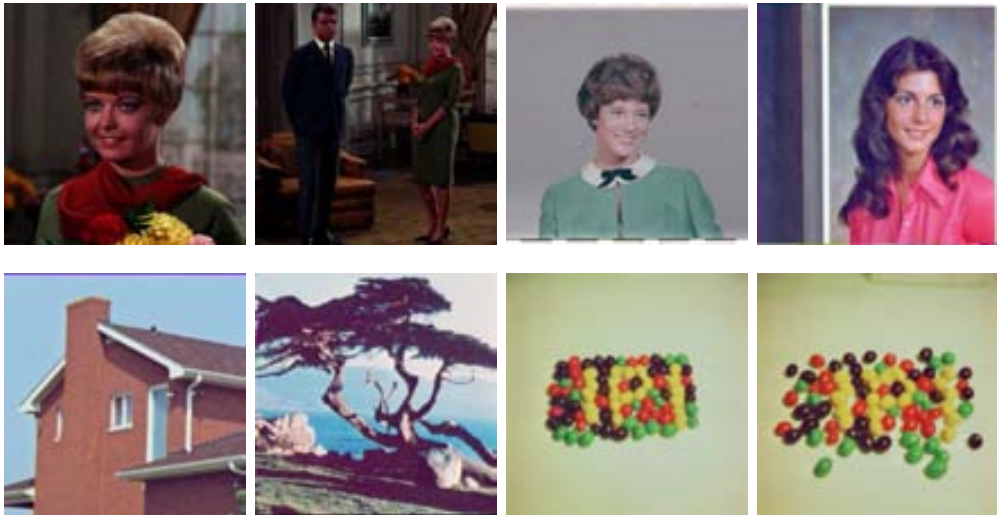
Figure A.5: Tested 256×256 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.
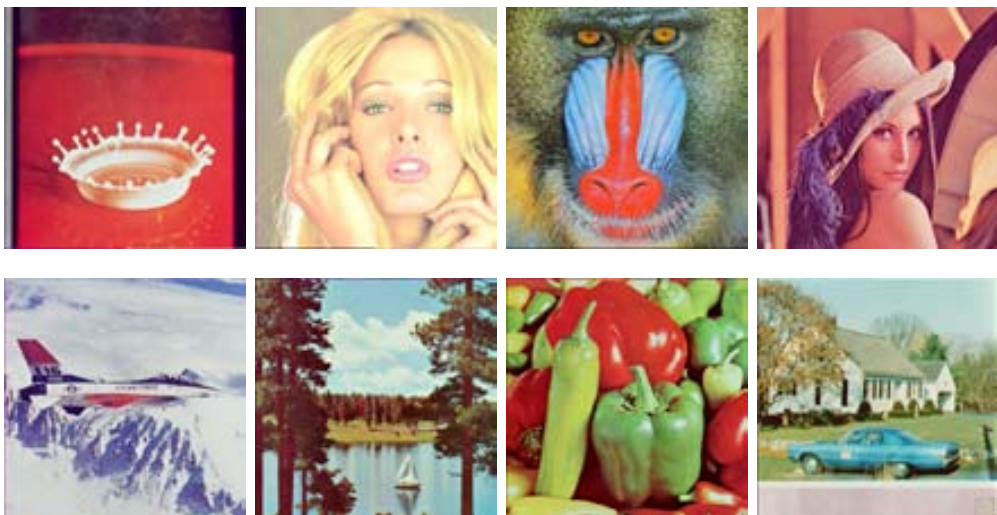
Figure A.6: Tested 512×512 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.

Appendix B

JPEG2000 vs H*i*-SET: Complementary Results of Chapter 3

B.1 University of Southern California Image Database

B.1.1 Gray-Scale (Y Channel)

Compression of Gray-Scale Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder(*Kakadu* Implementation(50)).
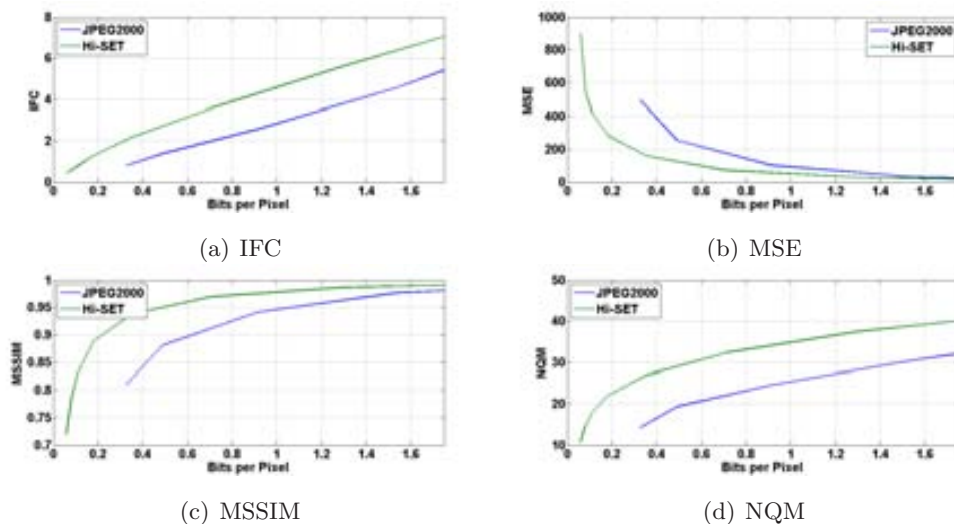

(a) IFC

(b) MSE
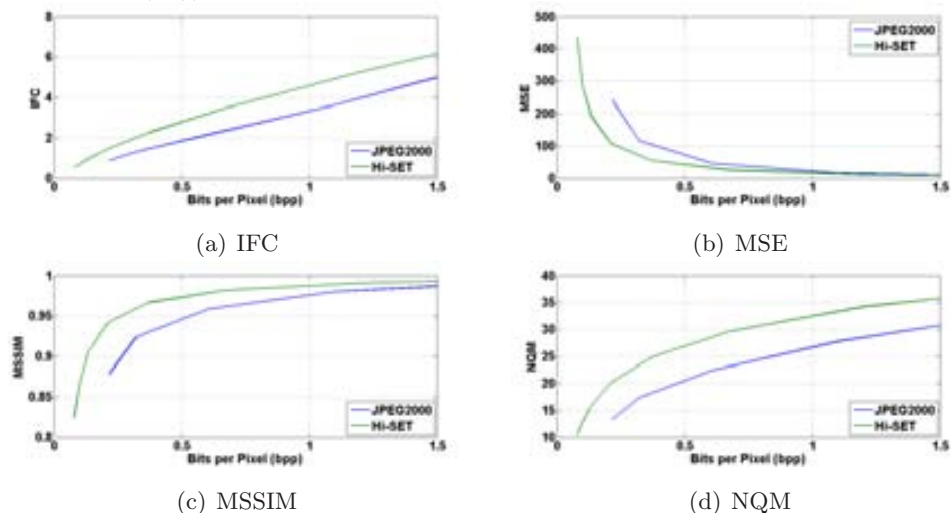
(c) MSSIM

(d) NQM

Figure B.1: Gray-Scale CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM

(d) UQI

Figure B.2: Gray-Scale CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed:
PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
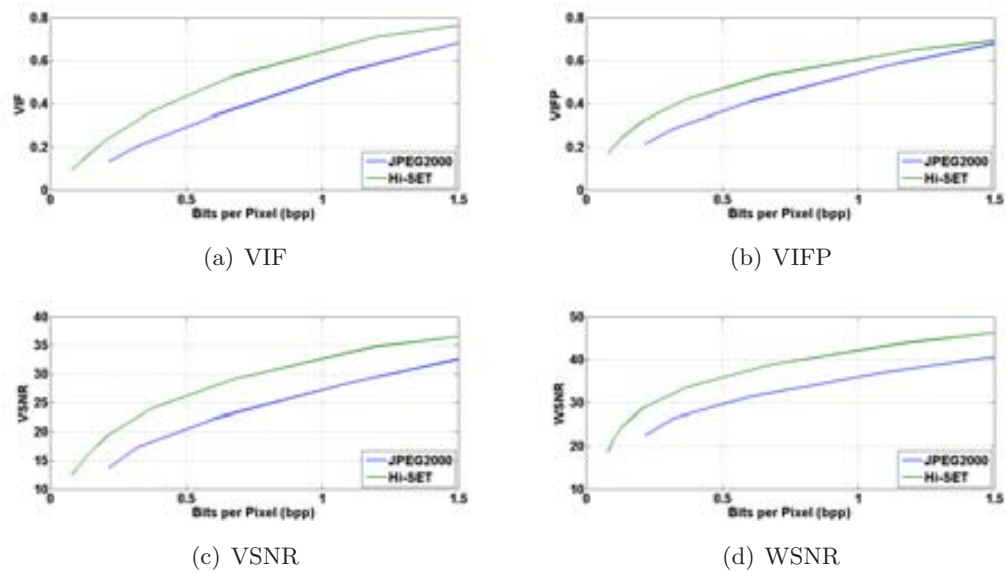
(d) WSNR

Figure B.3: Gray-Scale CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed:
VIF, VIFP, VSNR and WSNR.

B.1.2 Color Images

Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*Kakadu* Implementation(50)).

(a) IFC

(b) MSE
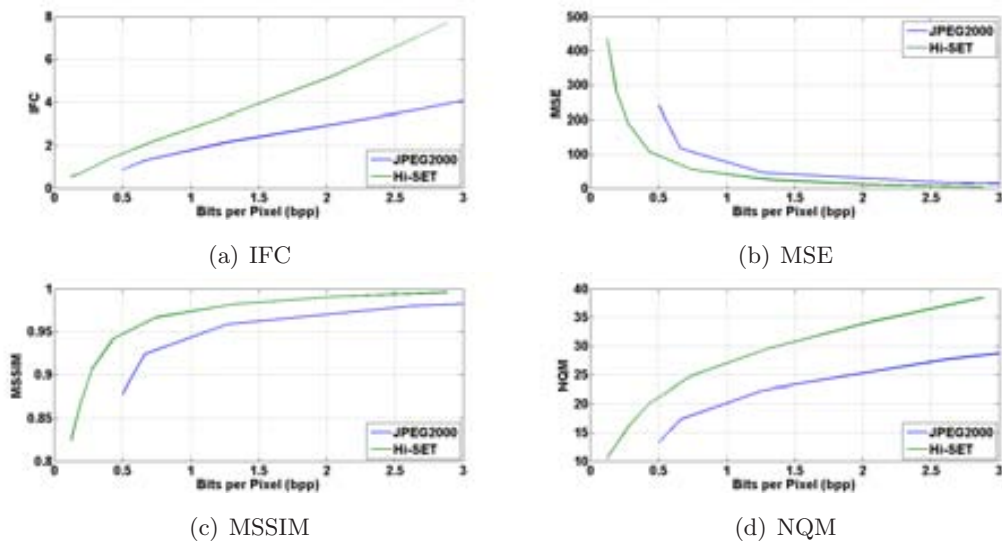
(c) MSSIM

(d) NQM

Figure B.4: Color CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
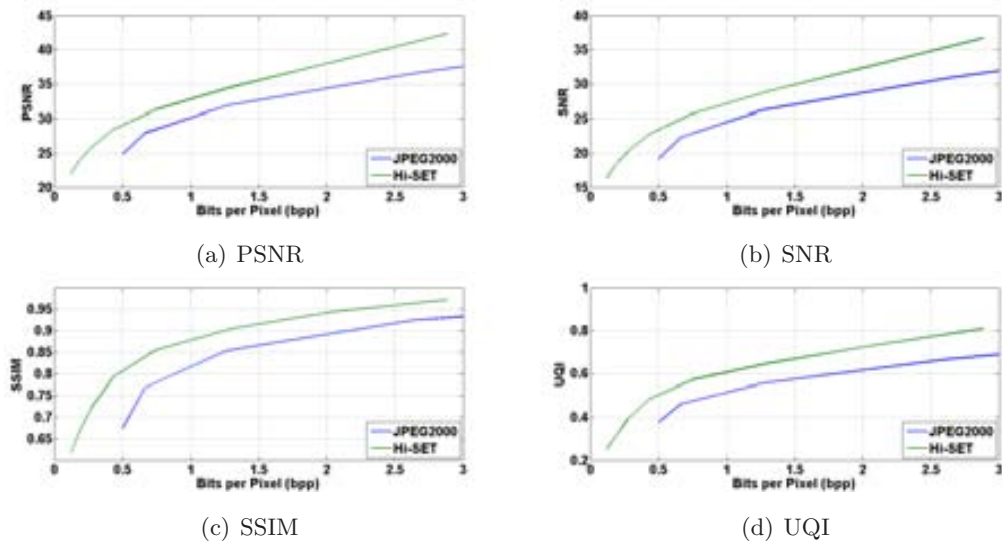
(d) UQI

Figure B.5: Color CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF	(b) VIFP
(c) VSNR	(d) WSNR

Figure B.6: Color CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.2 Categorical Subjective Image Quality Image Database

B.2.1 Gray-Scale (Y Channel)

Compression of Gray-Scale Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).
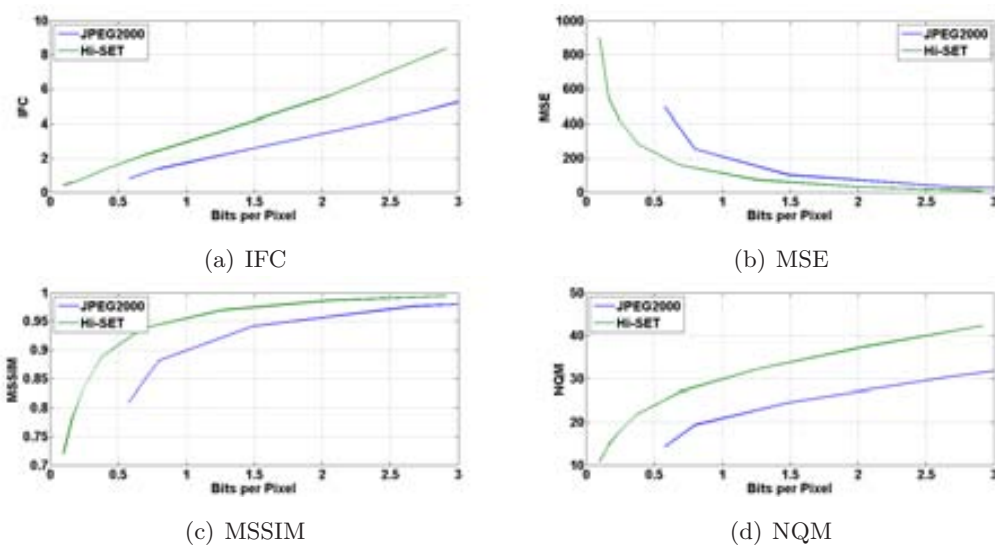

(a) IFC	(b) MSE
(c) MSSIM	(d) NQM

Figure B.7: Gray-Scale CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
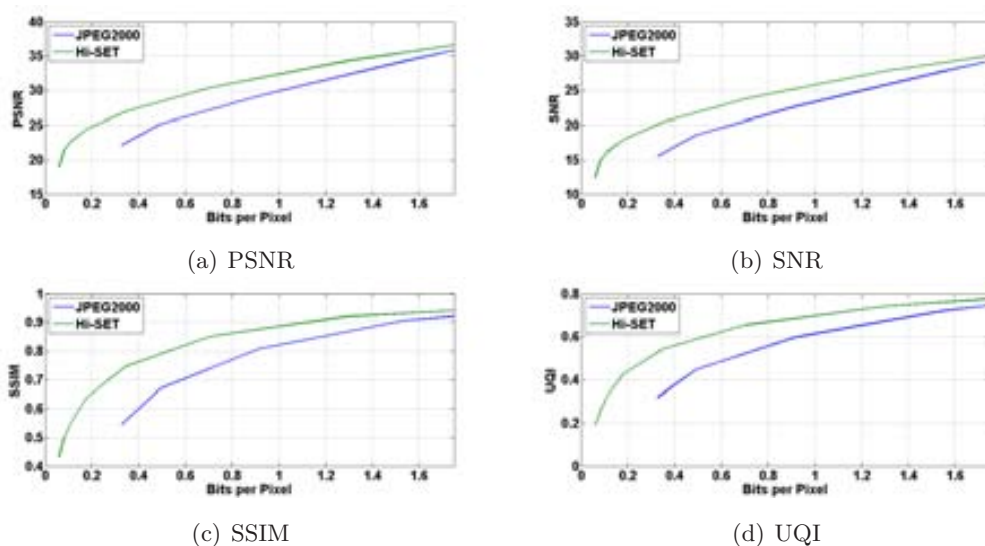
(d) UQI

Figure B.8: Gray-Scale CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
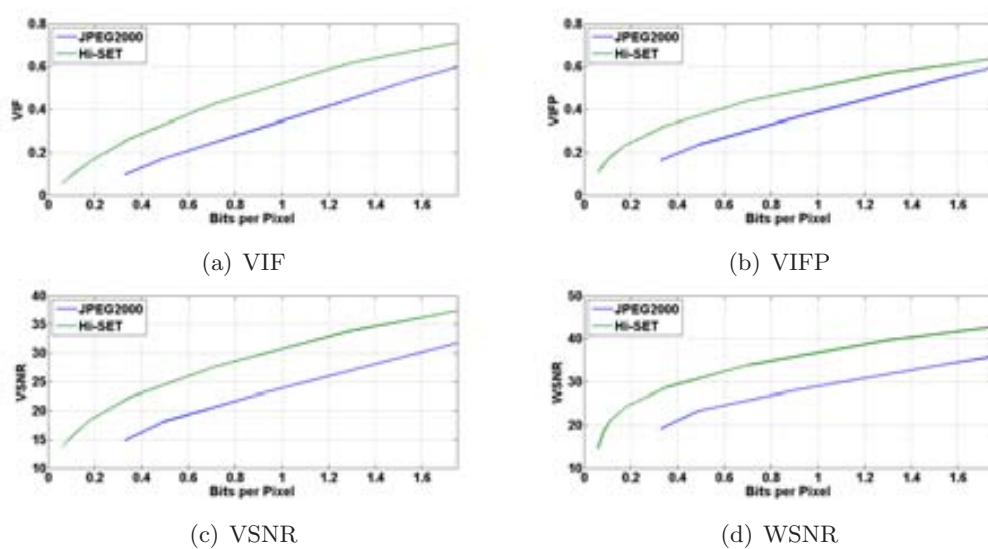
(d) WSNR

Figure B.9: Gray-Scale CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.2.2 Color Images

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).


(a) IFC

(b) MSE

(c) MSSIM

(d) NQM

Figure B.10: Color CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
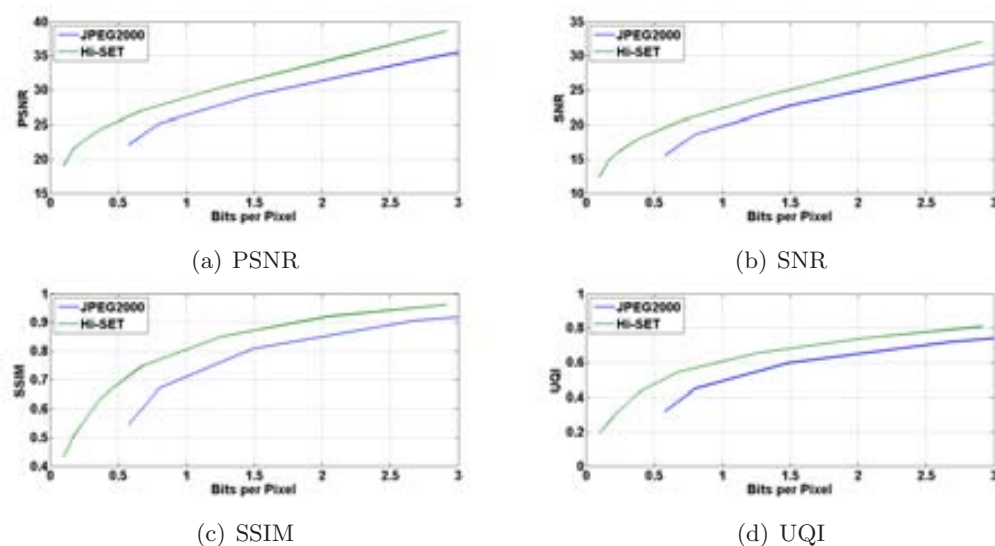
(d) UQI

Figure B.11: Color CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
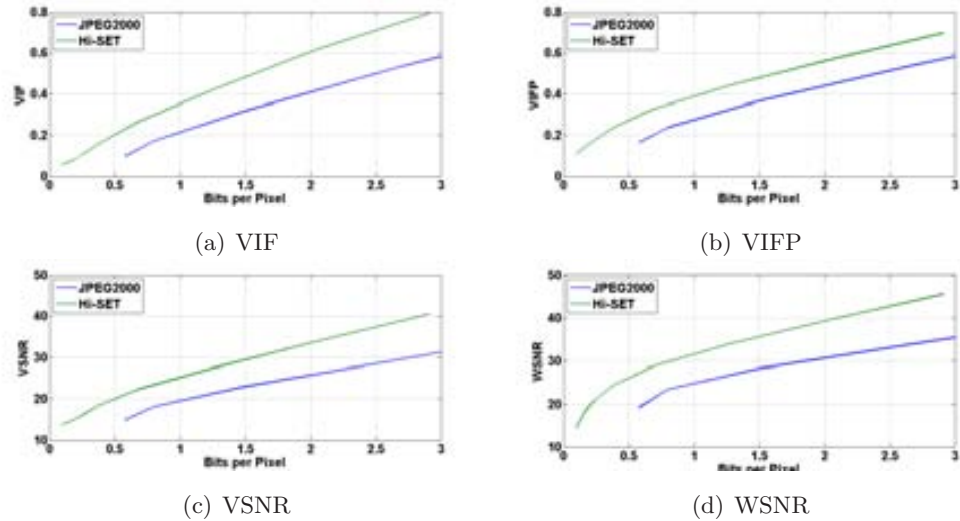
(d) WSNR

Figure B.12: Color CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.3   Image and Video-Communication Image Database
B.3.1   Gray-Scale (Y Channel)

Compression of Gray-Scale Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).

(a) IFC

(b) MSE
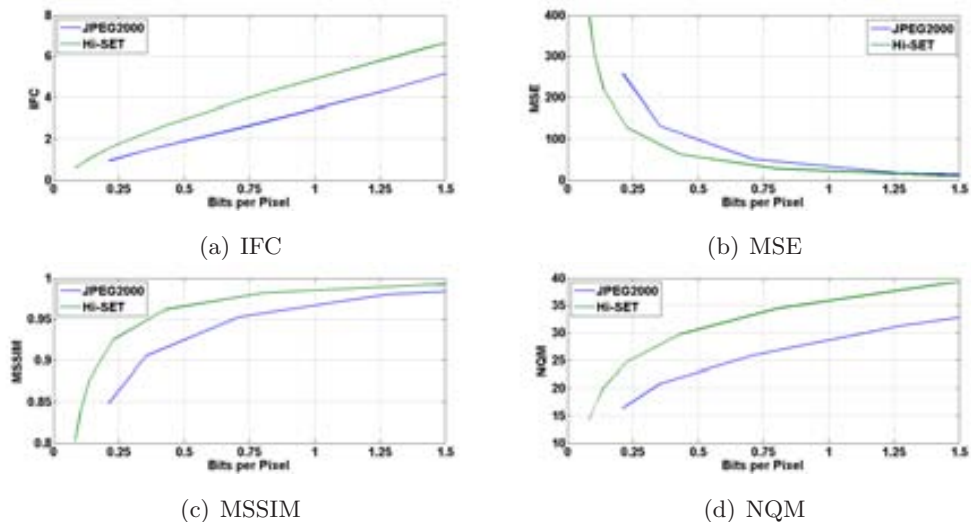
(c) MSSIM

(d) NQM

Figure B.13: Gray-Scale IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
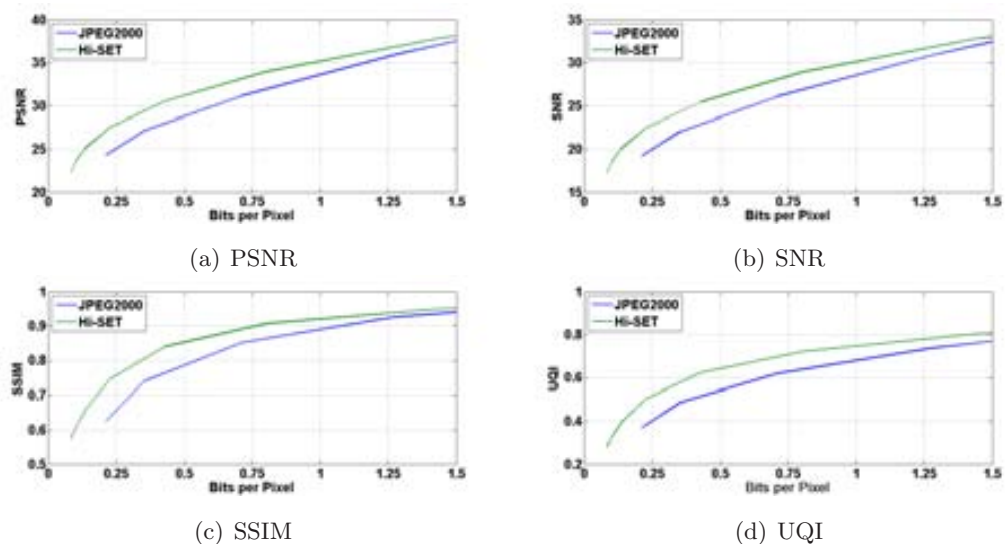
(d) UQI

Figure B.14: Gray-Scale IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed:
PSNR, SNR, SSIM and UQI.

(a) VIF

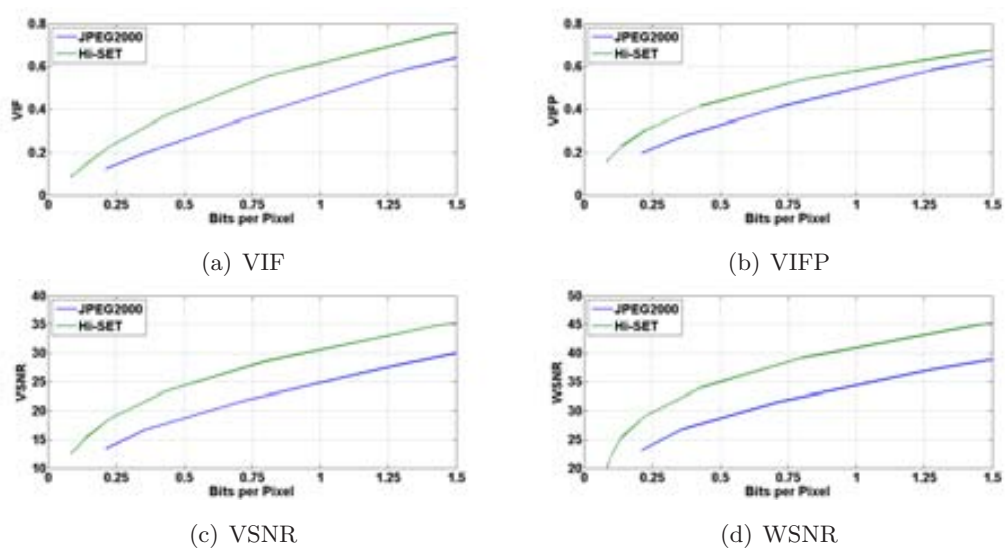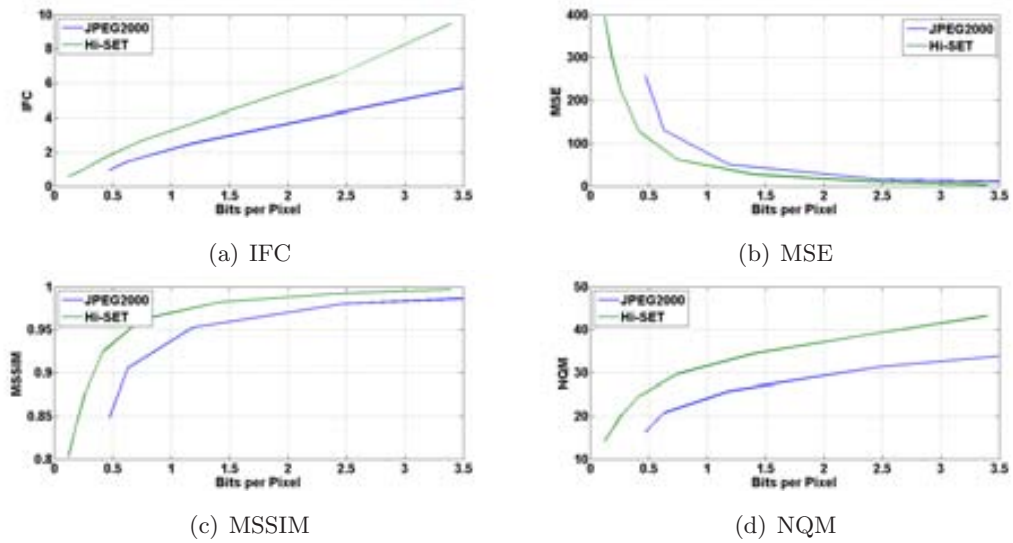(b) VIFP

(c) VSNR

(d) WSNR

Figure B.15: Gray-Scale IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed:
VIF, VIFP, VSNR and WSNR.

B.3.2 Color Images

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).

(a) IFC

(b) MSE

(c) MSSIM

(d) NQM

Figure B.16: Color IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
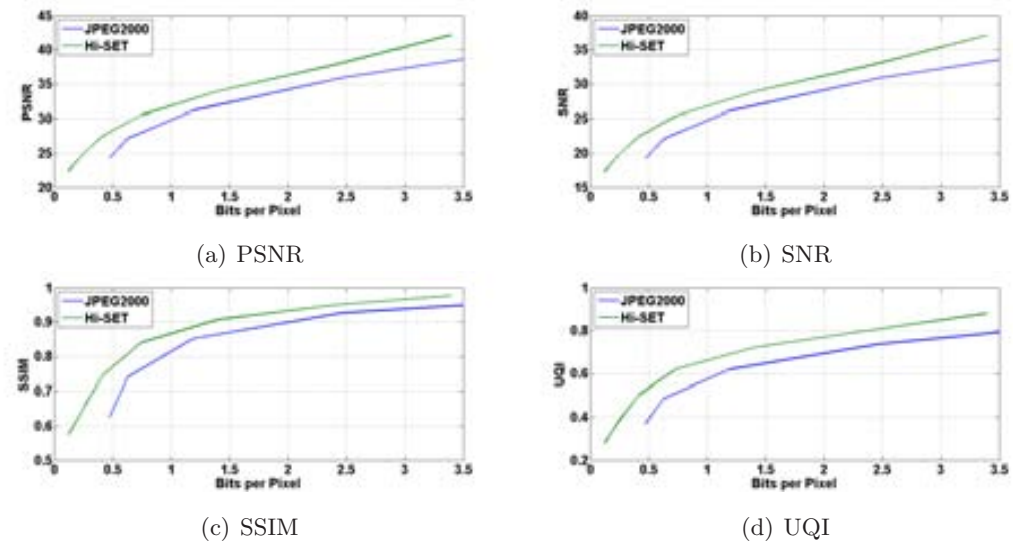
(d) UQI

Figure B.17: Color IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

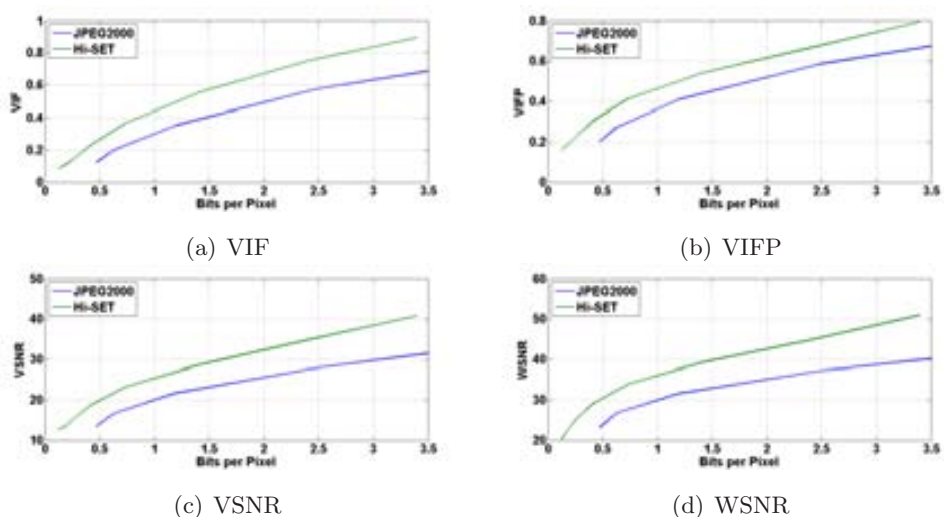(a) VIF

(b) VIFP

(c) VSNR

(d) WSNR

Figure B.18: Color IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.4 Image Database of the Laboratory for Image and Video Engineering

B.4.1 Gray-Scale (*Y* Channel)

Compression of Gray-Scale Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*Kakadu* Implementation(50)).
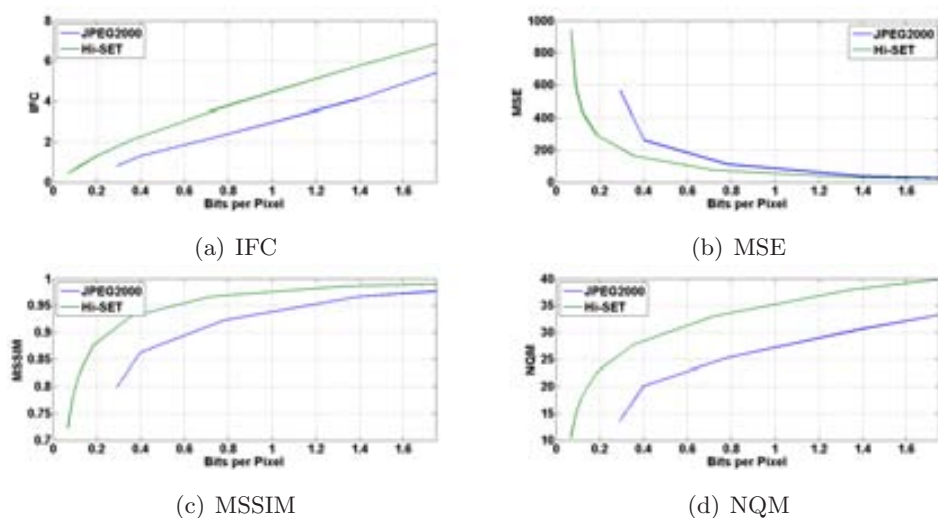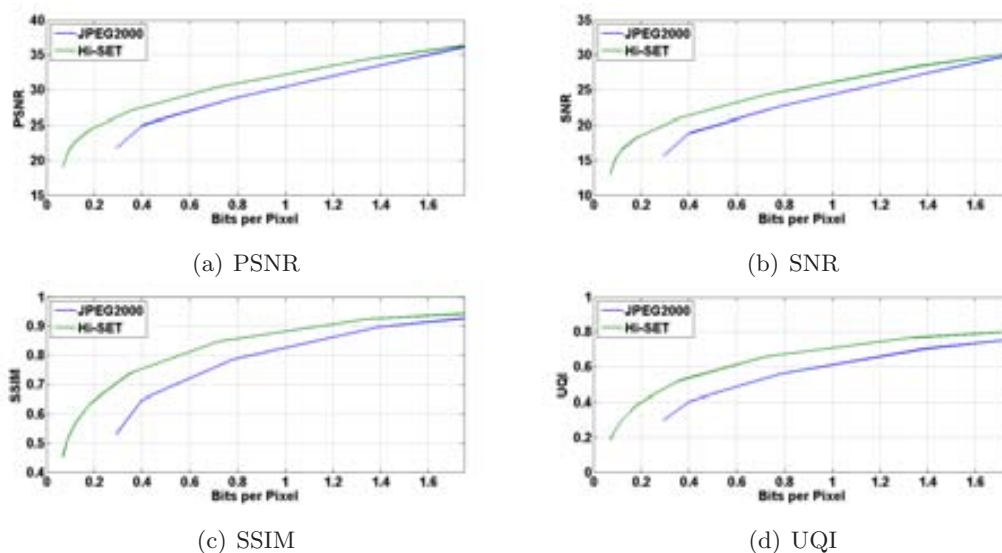

(a) IFC

(b) MSE

(c) MSSIM

(d) NQM

Figure B.19: Gray-Scale LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM

(d) UQI

Figure B.20: Gray-Scale LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
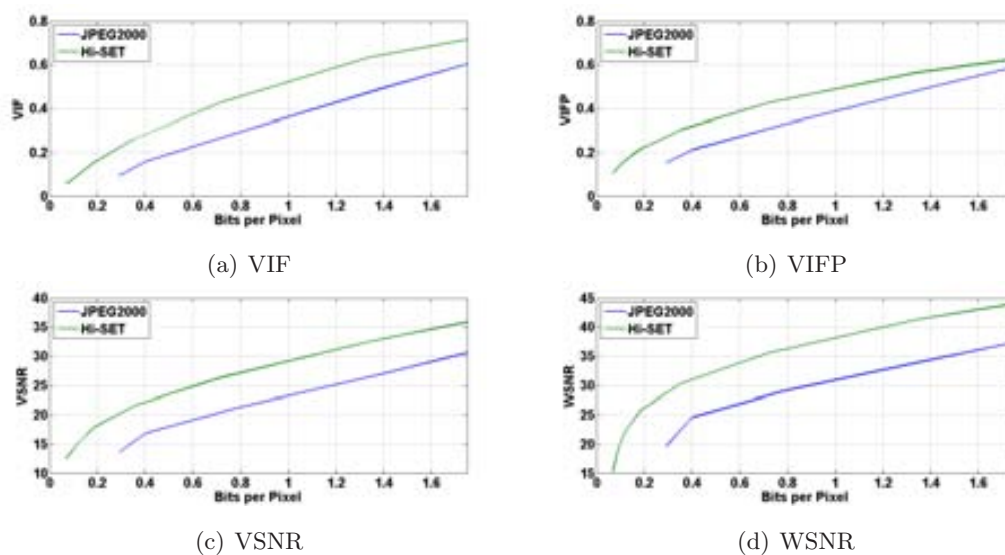
(d) WSNR

Figure B.21: Gray-Scale LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.4.2 Color Images

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*Kakadu* Implementation(50)).
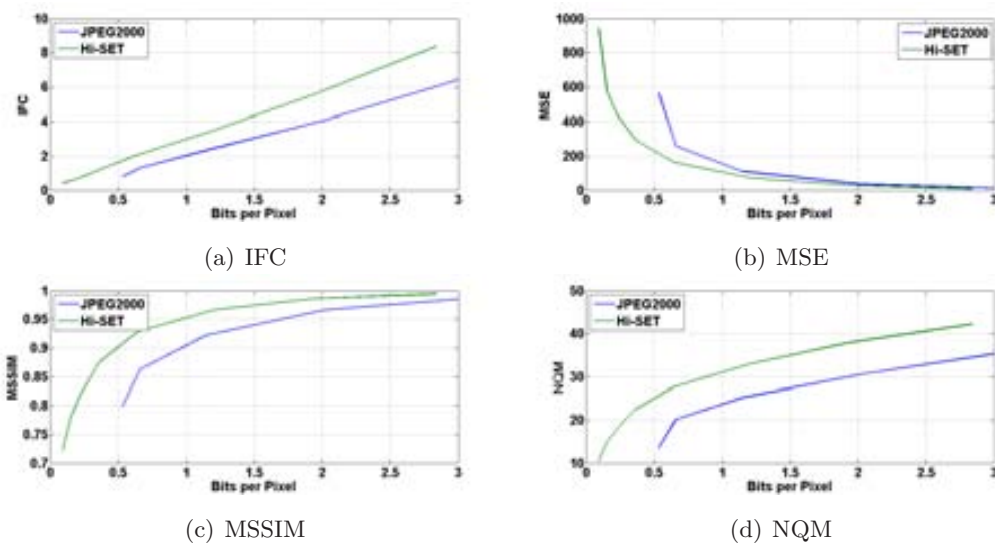

(a) IFC

(b) MSE

(c) MSSIM

(d) NQM

Figure B.22: Color LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

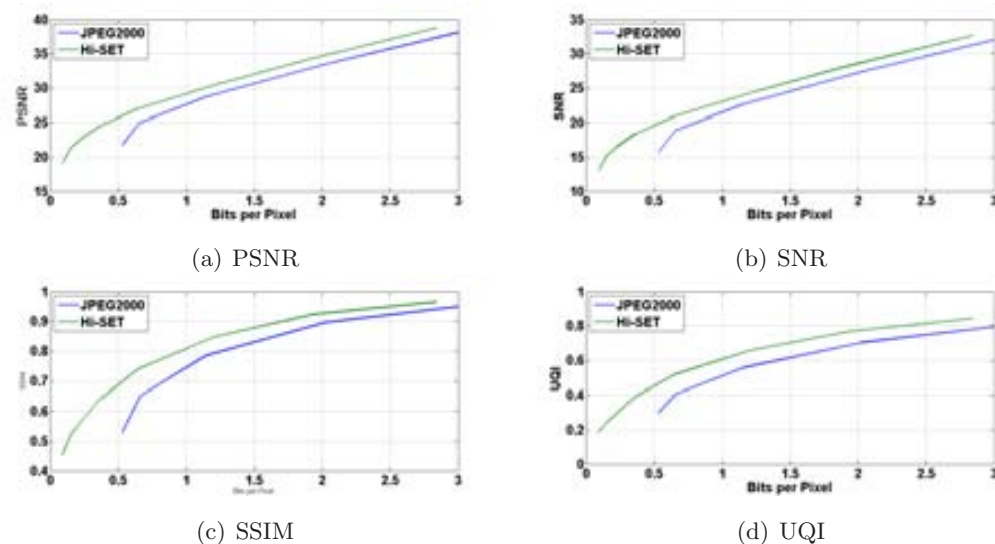
(a) PSNR

(b) SNR

(c) SSIM

(d) UQI

Figure B.23: Color LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.
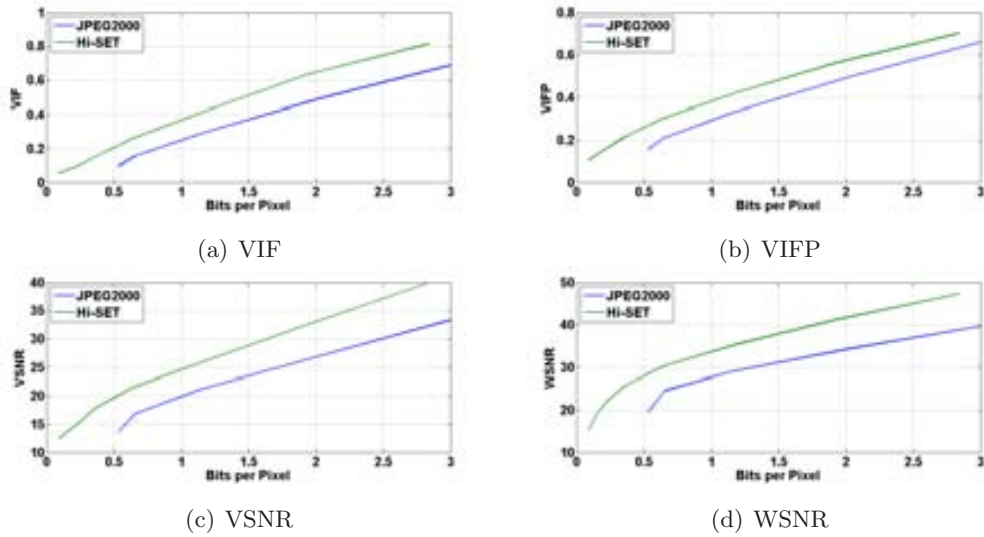
Figure B.24: Color LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.5 Tampere Image Database

B.5.1 Gray-Scale (Y Channel)

Compression of Gray-Scale Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).
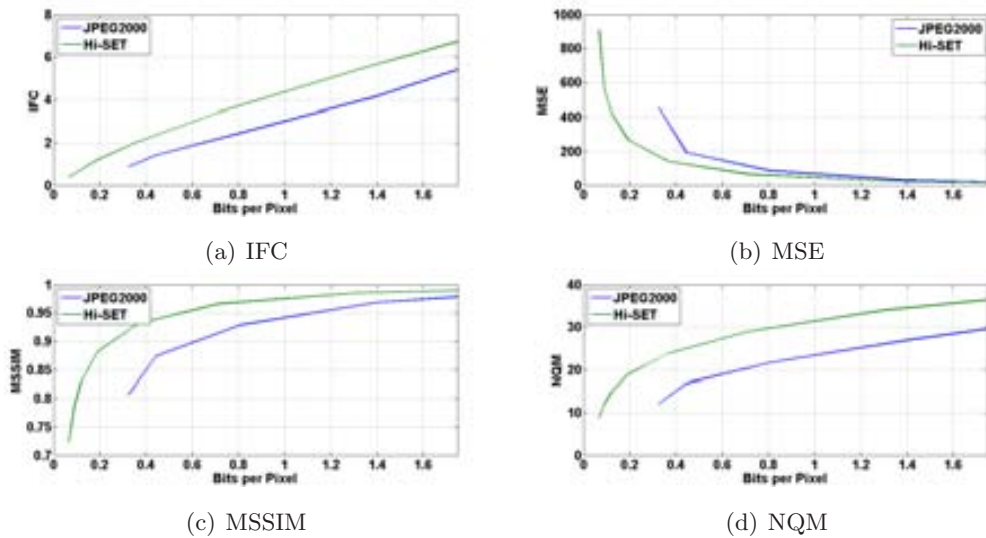

Figure B.25: Gray-Scale TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.

117

(a) PSNR

(b) SNR

(c) SSIM
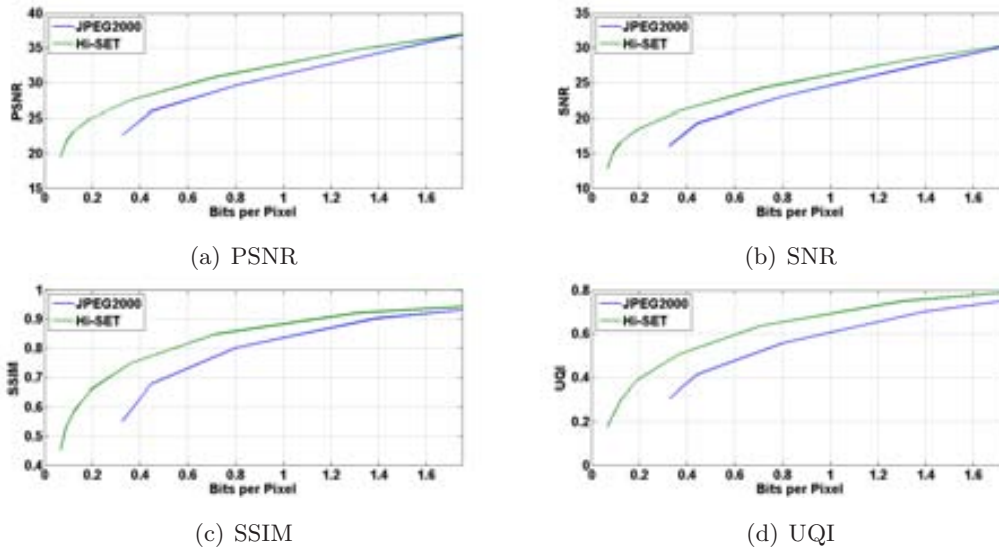
(d) UQI

Figure B.26: Gray-Scale TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.

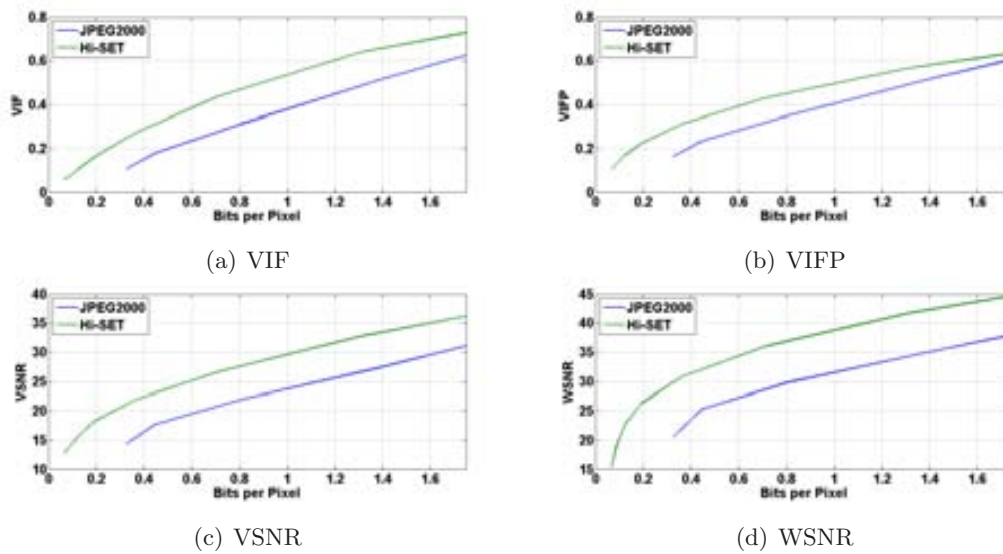
(a) VIF

(b) VIFP

(c) VSNR

(d) WSNR

Figure B.27: Gray-Scale TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.5.2 Color Images

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).
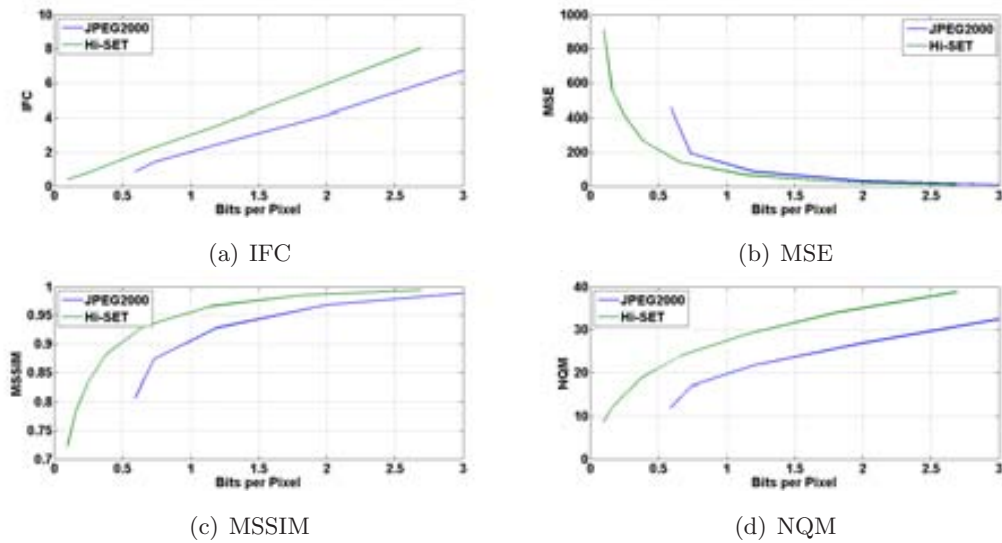

(a) IFC

(b) MSE

(c) MSSIM

(d) NQM

Figure B.28: Color TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.
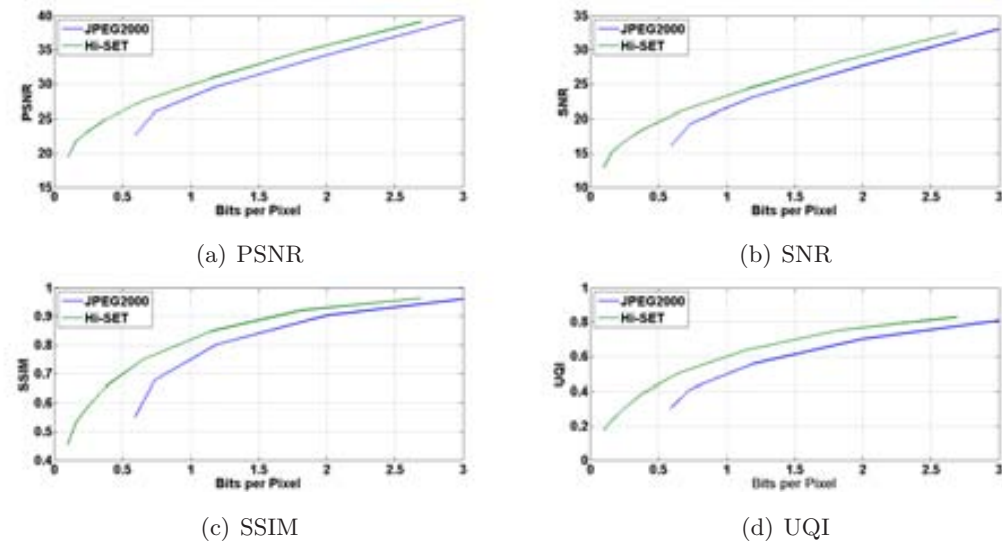

(a) PSNR

(b) SNR

(c) SSIM

(d) UQI

Figure B.29: Color TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.
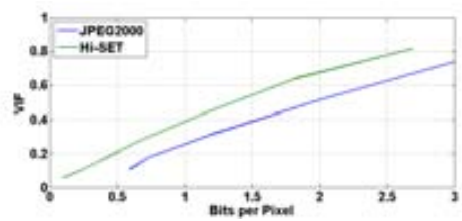
(a) VIF
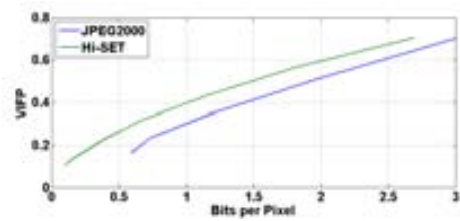
(b) VIFP
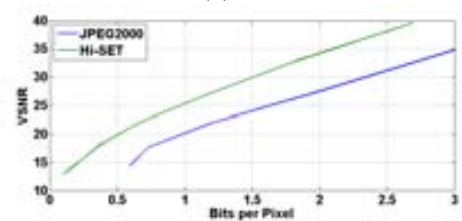
(c) VSNR
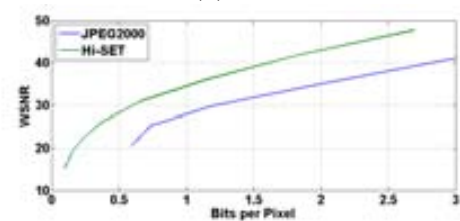
(d) WSNR

Figure B.30: Color TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed:
VIF, VIFP, VSNR and WSNR.

Appendix C

Complementary Results of Chapter 4

C.1   Correlation between $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$.

Green functions denoted as F-ρSQ are the quality metrics of forward perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as I-ρSQ are the quality metrics of recovered images after applying $\widehat{\alpha}(\nu, r)$.

C.1.1   Categorical Subjective Image Quality Image Database

Results obtained in the CSIQ (Fig. A.4) image database.

(a) PSNR   (b) MSSIM
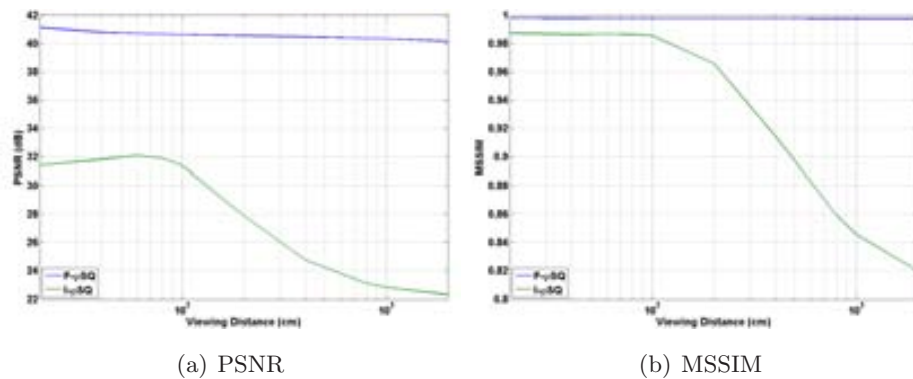
Figure C.1: Compression of Gray-scale Images (Y Channel) of the CSIQ image database.

(a) PSNR   (b) MSSIM
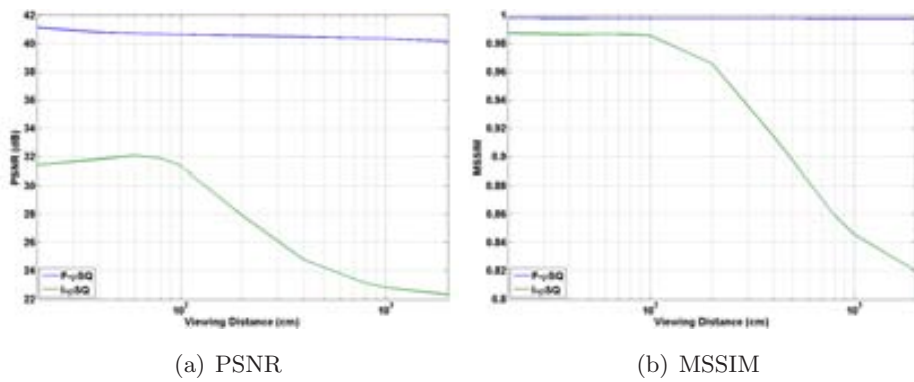
Figure C.2: Perceptual Quantization of Color Images of the CSIQ image database.

C.1.2   Image and Video-Communication Image Database

Results obtained in the IVC (Fig. A.1) image database.

(a) PSNR   (b) MSSIM
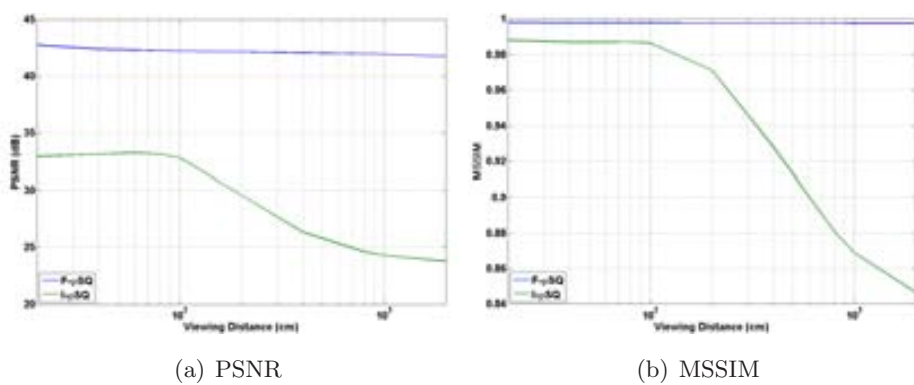
Figure C.3: Perceptual Quantization of Gray-scale Images (Y Channel) of the IVC image database.

(a) PSNR

(b) MSSIM
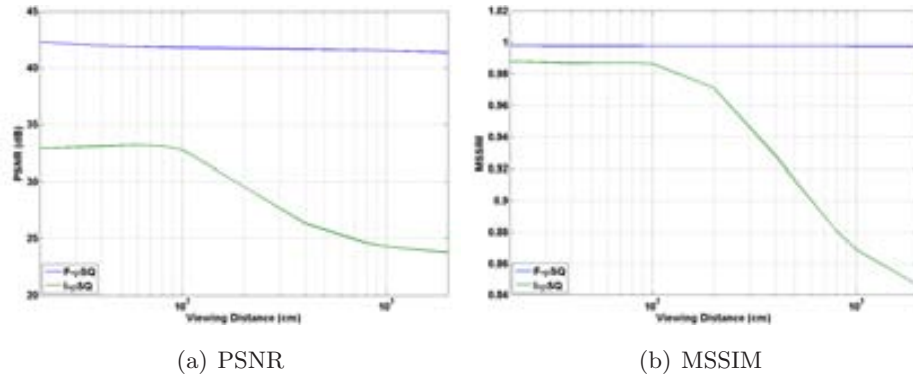
Figure C.4: Perceptual Quantization of Color Images of the IVC image database.

C.2 JPEG2000 vs Φ_{SET}

C.2.1 University of Southern California Image Database

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by Φ_{SET} coder, while blue functions by JPEG2000 coder (*Kakadu* Implementation(50)).

(a) IFC

(b) MSE
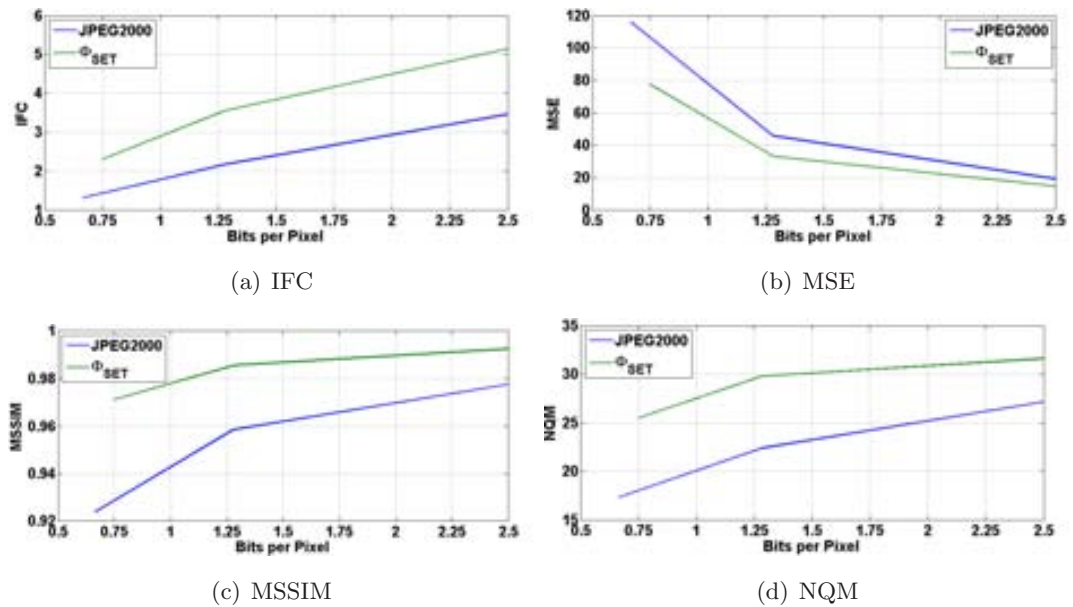
(c) MSSIM

(d) NQM

Figure C.5: Color CMU Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
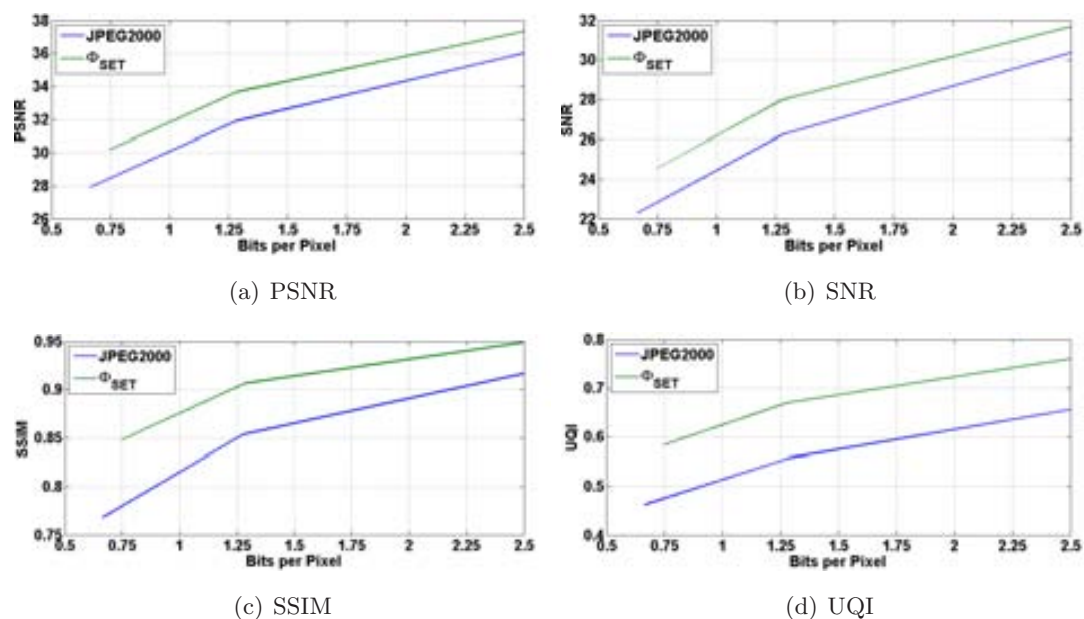
(d) UQI

Figure C.6: Color CMU Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
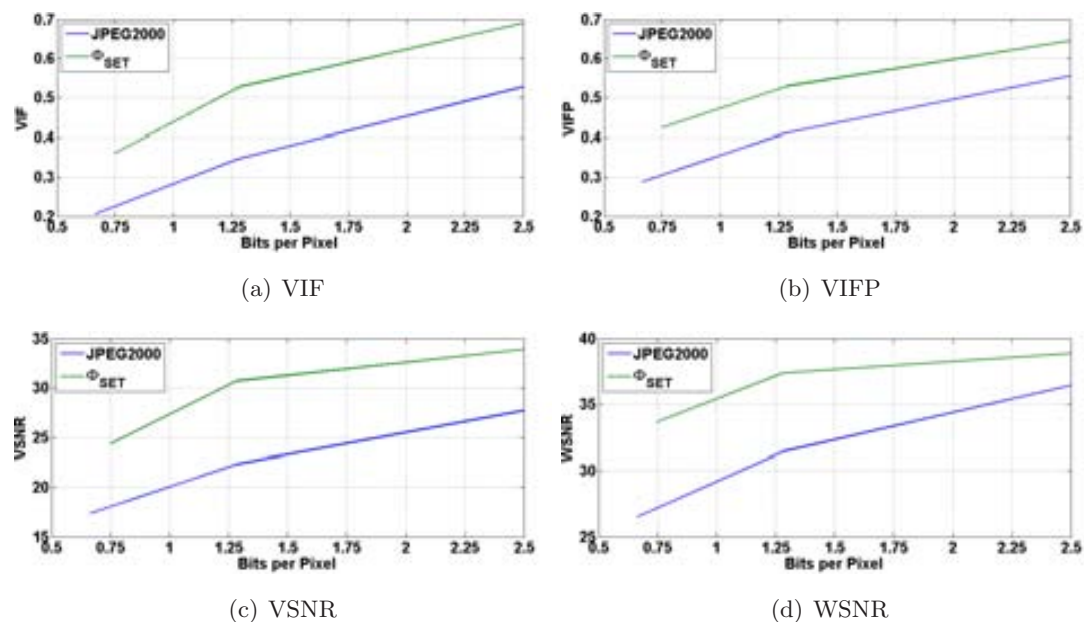
(d) WSNR

Figure C.7: Color CMU Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: VIF, VIFP, VSNR and WSNR.

C.2.2 Image and Video-Communication Image Database

Compression of Color Images (bits-per-pixel) vs Image Quality Assessment. Green functions represent results obtained by Φ_{SET} coder, while blue functions by JPEG2000 coder (*JJ2000* Implementation(40)).

(a) IFC   (b) MSE

(c) MSSIM   (d) NQM
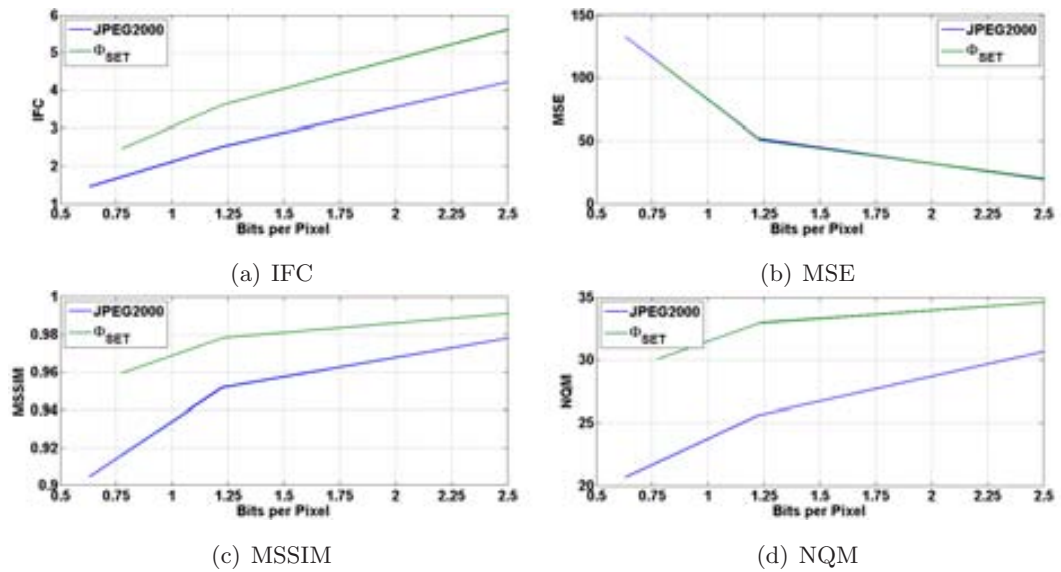
Figure C.8: Color IVC Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: IFC, MSE, MSSIM and NQM.

(a) PSNR

(b) SNR

(c) SSIM
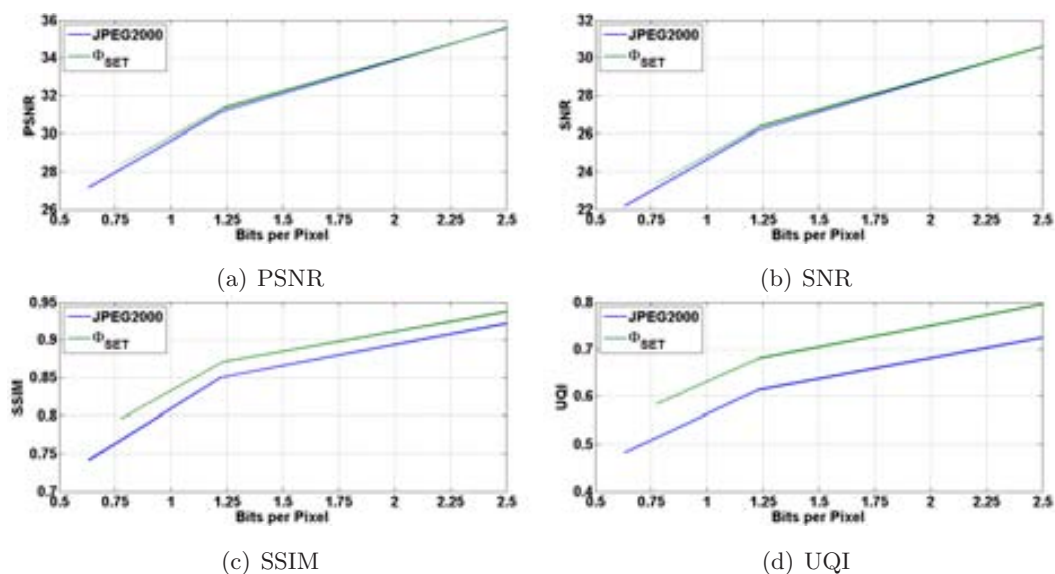
(d) UQI

Figure C.9: Color IVC Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: PSNR, SNR, SSIM and UQI.

(a) VIF

(b) VIFP

(c) VSNR
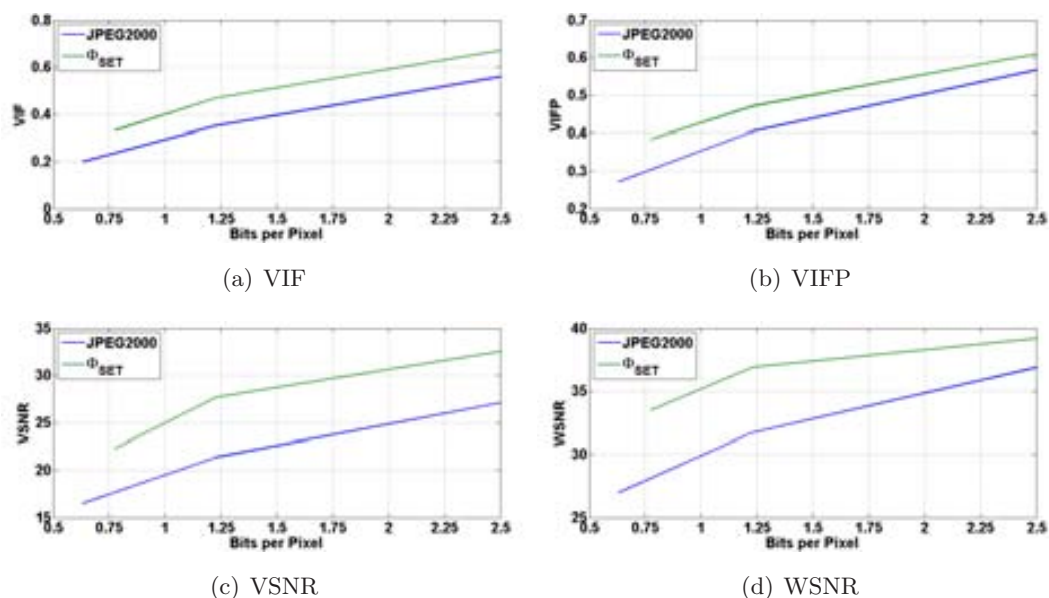
(d) WSNR

Figure C.10: Color IVC Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: VIF, VIFP, VSNR and WSNR.

References

[1] HERV ABDI. *Kendall rank correlation.* Encyclopedia of Measurement and Statistics. Thousand Oaks (CA), 2007. 25

[2] SIGNAL AND IMAGE PROCESSING INSTITUTE OF THE UNIVERSITY OF SOUTHERN CALIFORNIA. The USC-SIPI image database, available at http://sipi.usc.edu/database/, 1997. 18, 61, 93, 103

[3] M. ANTONINI, M. BARLAUD, P. MATHIEU, AND I. DAUBECHIES. Image coding using wavelet transform. *IEEE Transactions on Image Processing*, **1**(2):205 − 220, April 1992. 32

[4] E. ATSUMI AND N. FARVARDIN. Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees. In *International Conference on Image Processing*, **1**, pages 87 −91 vol.1, oct 1998. 77

[5] F. AULI-LLINAS AND J. SERRA-SAGRISTA. Low complexity JPEG2000 rate control through reverse subband scanning order and coding passes concatenation. *IEEE Signal Processing Letters*, **14**(4):251 −254, april 2007. 8

[6] J. BARTRINA-RAPESTA, F. AULI-LLINAS, J. SERRA-SAGRISTA, AND J.L. MONTEAGUDO-PEREIRA. JPEG2000 Arbitrary ROI coding through rate-distortion optimization techniques. In *Data Compression Conference*, pages 292 −301, 25-27 2008. 8

[7] J. BARTRINA-RAPESTA, F. AULI-LLINAS, J. SERRA-SAGRISTA, A. ZABALA-TORRES, X. PONS-FERNANDEZ, AND J. MASO-PAU. Region of interest coding

REFERENCES

applied to map overlapping in geographic information systems. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 5001 –5004, 23-28 2007. 87

[8] LUDWIG VON BERTALANFFY. *Teoría General de los Sistemas*. 1989. 2

[9] SAMBHUNATH BISWAS. One-dimensional B-B polynomial and hilbert scan for graylevel image coding. *Pattern Recognition*, **37**(4):789 – 800, 2004. Agent Based Computer Vision. 46

[10] MARTIN BOLIEK, CHARILAOS CHRISTOPOULOS, AND ERIC MAJANI. *Information Technology: JPEG2000 Image Coding System.* ISO/IEC JTC1/SC29 WG1, JPEG 2000, JPEG 2000 Part I final committee draft version 1.0 edition, April 2000. 2, 31, 34, 45, 59, 60, 61, 77

[11] MARTIN BOLIEK, ERIC MAJANI, J. SCOTT HOUCHIN, JAMES KASNER, AND MATHIASLARSSON CARLANDER. *Information Technology: JPEG2000 Image Coding System (Extensions).* ISO/IEC JTC 1/SC 29/WG 1, JPEG 2000 Part II final committee draft edition, Dec. 2000. 77, 85

[12] DAMON CHANDLER AND SHEILA HEMAMI. Vsnr: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, **16**(9):2284 –2298, 2007. 25, 57, 72

[13] C. CHRISTOPOULOS, J. ASKELOF, AND M. LARSSON. Efficient methods for encoding regions of interest in the upcoming jpeg2000 still image coding standard. *IEEE Signal Processing Letters*, **7**(9):247 –249, sep 2000. 77

[14] N. DAMERA-VENKATA, T. KITE, W. GEISLER, B. EVANS, AND A. BOVIK. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, **9**:636–650, 2000. 25, 57, 72

[15] J. GONZALEZ-CONEJERO, J. SERRA-SAGRISTA, C. RUBIES-FEIJOO, AND L. DONOSO-BACH. Encoding of images containing no-data regions within JPEG2000 framework. In *15th IEEE International Conference on Image Processing*, pages 1057 –1060, 12-15 2008. 87

[16] DAVID HILBERT. Über die stetige Abbildung einer Linie auf ein Flächenstück. *Mathematische Annalen*, **38**(3):459–460, Sept. 1891. xiii, 35

[17] M. HOLLANDER AND D.A. WOLFE. *Non-parametric Statistical Methods*. Wiley, 2nd edition, 1999. 25

[18] Q. HUYNH-THU AND M. GHANBARI. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, **44**(13):800–801, 2008. 25, 57, 72

[19] ISO/IEC 12640-1. Graphic technology - prepress digital data exchange - CMYK standard colour image data (CMYK/SCID), 1997. 55

[20] HYUNGJUN KIM AND C.C. LI. Lossless and lossy image compression using biorthogonal wavelet transforms with multiplierless operations. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, **45**(8):1113–1118, 1998. 46

[21] CORNELL UNIVERSITY VISUAL COMMUNICATIONS LABORATORY. MeTriX MuX visual quality assessment package , available at *http* : *//foulard.ece.cornell.edu/gaubatz/metrix_mux/*, 2010. 25

[22] ERIC C. LARSON AND DAMON M. CHANDLER. Most apparent distortion: a dual strategy for full-reference image quality assessment. In *Proc. SPIE*, **742**, 2009. 9, 23, 102

[23] PATRICK LE CALLET AND FLORENT AUTRUSSEAU. Subjective quality assessment IRCCyN/IVC database, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/. 9, 23, 99

[24] MICHAEL W. MARCELLIN, MARGARET A. LEPLEY, ALI BILGIN, THOMAS J. FLOHR, TROY T. CHINEN, AND JAMES H. KASNER. An overview of quantizartion of JPEG2000. *Signal Processing: Image Communication*, **17**(1):73–84, Jan. 2002. 33

[25] T. MITSA AND K. VARKUR. Evaluation of contrast sensitivity functions for formulation of quality measures incorporated in halftoning algorithms. *IEEE International Conference on Acustics, Speech and Signal Processing*, **5**:301–304, 1993. 25, 57, 64, 72

REFERENCES

[26] JAIME MORENO AND XAVIER OTAZU. Image coder based on H*i*lbert Scaning of Embedded quadTrees. *IEEE Data Compression Conference*, page 470, March 2011. 98

[27] K. T. MULLEN. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, **359**:381–400, February 1985. 11

[28] K.T. MULLEN. The contrast sensitivity of human color vision to red-green and blue-yellow chromatic gratings. *Journal of Physiology*, **359**:381–400, 1985. 10

[29] NAILA MURRAY, MARIA VANRELL, XAVIER OTAZU, AND ALEJANDRO PARRAGA. Saliency estimation using a non-parametric low-level vision model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011)*, pages 433 –440, 2010. 9

[30] D. NISTER AND C. CHRISTOPOULOS. Lossless region of interest with a naturally progressive still image coding algorithm. In *International Conference on Image Processing*, pages 856 –860 vol.3, oct 1998. 77

[31] DAVID NISTER AND CHARILAOS CHRISTOPOULOS. Lossless region of interest coding. *Signal Processing*, **78**(1):1 – 17, 1999. 77

[32] X OTAZU, C.A PÁRRAGA, AND M VANRELL. Toward a unified chromatic induction model. *Journal of Vision*, **10(12)**(6), 2010. 2, 9

[33] X. OTAZU, M. VANRELL, AND C.A. PARRAGA. Multiresolution wavelet framework models brightness induction effects. *Vision Research*, **48**:733–751, 2007. 2

[34] W. A. PEARLMAN AND A. SAID. Image wavelet coding systems: Part II of set partition coding and image wavelet coding systems. *Foundations and Trends in Signal Processing*, **2**(3):181–246, 2008. 29, 34, 37

[35] W. A. PEARLMAN AND A. SAID. Set partition coding: Part I of set partition coding and image wavelet coding systems. *Foundations and Trends in Signal Processing*, **2**(2):95–180, 2008. 29, 37

[36] W.A. Pearlman, A. Islam, N. Nagaraj, and A. Said. Efficient, low-complexity image coding with a Set-Partitioning Embedded bloCK coder. *IEEE Transactions on Circuits and Systems for Video Technology*, **14**(11):1219 – 1235, Nov. 2004. 29

[37] PEIPA. Pilot european image processing archive, available at http://peipa.essex.ac.uk/. 89

[38] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, and V. Lukin. Metrics performance comparison for color image database. *Fourth international workshop on video processing and quality metrics for consumer electronics*, page 6 p., 2009. 9, 23, 100

[39] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti. TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, **10**:30–45, 2009. 9, 23, 49, 100

[40] Cannon Research, École Polytechnique Fédérale de Lausanne, and Ericsson. JJ2000 implementation in Java, available at http://jj2000.epfl.ch/, 2001. 49, 62, 73, 83, 108, 110, 111, 113, 117, 119, 125

[41] A. Said and W.A. Pearlman. A new, fast, and efficient image codec based on Set Partitioning In Hierarchical Trees. *IEEE Transactions on Circuits and Systems for Video Technology*, **6**(3):243 – 250, June 1996. 29, 32, 98

[42] David Salomon. *Data Compression: The Complete Reference.* ISBN-13: 978-1-84628-602-5. Springer-Verlag London Limited, fourth edition, 2007. 5

[43] Peter Schelkens, Athanassios Skodras, and Touradj Ebrahimi. *The JPEG 2000 Suite.* The Wiley-IS&T Series in Imaging Science and Technology, 2009. xiii, 48

[44] J.M Shapiro. Embedded image coding using Zerotrees of wavelet coefficients. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **41**(12):3445 – 3462, Dec. 1993. 29, 98

REFERENCES

[45] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, **15**(2):430 –444, feb. 2006. 9, 25, 57, 72

[46] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, **15**(11):3440 –3451, 2006. 9, 23, 101

[47] R. Sheikh, A. Bovik, and G. de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, **14**:2117–2128, 2005. 25, 57, 72

[48] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, **18**(5):36–58, September 2001. 31, 77, 98

[49] Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, **3**(2):186 – 200, 1996. 32

[50] David Taubman. Kakadu software, available at http://www.kakadusoftware.com/, July 2010. 49, 57, 73, 105, 107, 114, 116, 123

[51] David S. Taubman and Michel W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. ISBN: 0-7923-7519-X. Kluwer Academic Publishers, 2002. 3, 8, 28, 31, 48, 49, 72, 77

[52] Bryan E. Usevitch. A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000. *IEEE Signal Processing Magazine*, **18**(5):22–35, 2001. 29

[53] G. van de Wouwer, P. Scheunders, and D. van Dyck. Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing*, **8**(4):592 –598, April 1999. 15

[54] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *Conference Record of the Thirty-Seventh Asilomar*

Conference on Signals, Systems and Computers., **2**, pages 1398 – 1402 Vol.2, 2003. 9, 25, 57, 72

[55] ZHANG WANG AND ALAN BOVIK. A universal image quality index. *IEEE Signal Processing Letters*, **9**:81–84, 2002. 25, 57, 72

[56] ZHOU WANG, SERENE BANERJEE, BRIAN L. EVANS, AND ALAN C. BOVIK. Generalized bitplane-by-bitplane shift method for JPEG2000 ROI coding. *IEEE International Conference on Image Processing*, **3**:81–84, September 22-25 2002. 80

[57] ZHOU WANG AND A.C. BOVIK. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE*, **26**(1):98 –117, jan. 2009. 1, 7

[58] ZHOU WANG, A.C. BOVIK, H.R. SHEIKH, AND E.P. SIMONCELLI. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, **13**(4):600 –612, april 2004. 9, 25, 57, 72

[59] ZHOU WANG AND ALAN C. BOVIK. *Modern Image Quality Assessment.* Morgan & Claypool Publishers: Synthesis Lectures on Image, Video, & Multimedia Processing, 1 edition, February 2006. 1, 7, 31

[60] ZHOU WANG AND ALAN C.BOVIK. Bitplane-by-bitplane shift (Bb BShift) - a suggestion for JPEG2000 region of interest image coding. *IEEE Signal Processing Letters*, **9**(5):160 – 162, May 2002. 79

[61] BETH A. WILSON AND MAGDY A. BAYOUMI. A computational kernel for fast and efficient compressed-domain calculations of wavelet subband energies. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, **50**(7):389 – 392, July 2003. 15

[62] BETTYE WILSON. *Ethics and Basic Law for Medical Imaging Professionals.* F.A. Davis Co., 1997. 89

[63] L. ZHANG AND XIANCHUAN YU. New region of interest coding for remote sensing image based on multiple bitplanes up-down shift. In *Systems and Control in Aerospace and Astronautics, 2006. ISSCAA 2006. 1st International Symposium on*, pages 5 pp. –673, jan. 2006. 93

Publications

Journals

- Jaime Moreno and Xavier Otazu, Full-Reference Quality Assessment using a Chromatic Induction Model: JPEG and JPEG2000, *Journal of the Optical Society of America A*, Submitted.

- Jaime Moreno and Xavier Otazu, Image Coder Based on Hilbert Scanning of Embedded quadTrees, *Digital Signal Processing*, Submitted.

- Jaime Moreno and Xavier Otazu, Perceptual Quantization using a Chromatic Induction Model, *IEEE Transactions on Image Processing*, Submitted.

- Jaime Moreno and Xavier Otazu, Perceptual Generalized Bitplane-by-Bitplane Shift, *IEEE Signal Processing Letters*, Submitted.

Conferences and other Publications

- Jaime Moreno and Xavier Otazu, Image Coder Based on Hilbert Scanning of Embedded quadTrees: An Introduction of the Hi-SET Coder, *2011 IEEE International Conference on Multimedia and Expo (ICME 2011)* , Barcelona, Spain from July 11 to 15, 2011, Accepted.

- Jaime Moreno and Xavier Otazu, Image Coder Based on Hilbert Scanning of Embedded quadTrees, *Data Compression Conference (DCC) 2011*, abstract on page 470, Snowbird, USA, 29-31 March 2011.

- Jaime Moreno and Xavier Otazu, Full-Reference Perceptual Image Quality Assessment through the Chromatic Induction Wavelet Model, *Fifth CVC Workshop*

on the Progress of Research & Development, CVCRD'2010, Bellaterra, Spain, October 29th, 2010.

- Jaime Moreno , Xavier Otazu and Maria Vanrell, Local Perceptual Weighting in JPEG2000 for Color Images, *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science*, pages 255-260, Joensuu, Finland, June 2010.

- Jaime Moreno , Xavier Otazu and Maria Vanrell, Contribution of CIWaM in JPEG2000 Quantization for Color Images, *The CREATE Conference 2010*, pg. 132-136, Gøvik, Norway, June 2010.

- Jaime Moreno , Xavier Otazu and Maria Vanrell, Perceptual Criteria on JPEG2000 Quantization, *Fourth CVC Workshop on the Progress of Research & Development, CVCRD'2009*, Bellaterra, Spain, October 30th, 2009.

- Jaime Moreno , Xavier Otazu and Maria Vanrell, Perceptual Criteria on JPEG2000 Quantization,*The CREATE Conference 2009*, Gargnano, Italy, 19-24th October 2009.

Index