Perceptual Criteria on Image Compression

Universitat Autònoma de Barcelona

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** Associate Professor of the Department of Optics School of Physics, Universitat de València, Spain. **Michael W. Marcellin** Director of the Signal Processing and Coding Laboratory Department of Electrical and Computer Engineering University of Arizona, USA. **Christine Fernandez-Maloigne** Director of the 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.

Day of the defense: **July 1st. 2011**

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 mis 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 una 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**, que siempre tuvieron una palabra de apoyo y aliento para mí durante mis estudios. *Gracias por todo.*

Jesús Jaime Moreno Escobar
Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Zacatenco
Instituto Politécnico Nacional, México, 2011

Abstract

The compression of an image is posible assuming that a portion of its information is redundant. Hence, this Ph.D. thesis proposes to exploit visual redundancy, existing in an image, reducing unperceivable frequencies for Human Visual System.

First, we define an image quality assessment, which is highly correlated with the opinion of observers. The proposed metrics weights the well-known PSNR by means of a Chromatic Induction Model ($\mathcal{C}_w$PSNR). Second, we propose an image compression algorithm, which exploit the high correlation and self-similarity of pixels in a given area or neighborhood by means of a fractal function, called H*i*-SET. H*i*-SET possesses the main features that modern image compressors have, namely it is a embedded coder, which permits a progressive transmission. Also, we propose a modification to the uniform scalar quantizer, which is applied to a pixel set in a certain Wavelet sub-band. Unlike this, the proposed modification permits to perform a pixel-by-pixel forward and inverse quantization, introducing into the compression process a perceptual distortion. Finally, a coding method for Region of Interest areas is presented, ρGBbBShift, which perceptually weights pixels of these areas in addition to maintain only the more important perceivable frequencies in the rest of the image.

Results exposed in this thesis show that $\mathcal{C}_w$PSNR is the best-ranked image quality assessment, when it is compared to the most common image compression distortions such as JPEG and JPEG2000, since $\mathcal{C}_w$PSNR possesses the best correlation with the opinion of observers, which is based on the results of psychophysical experiments belonging to the most important image databases in this field of science such as TID2008, LIVE, CSIQ and IVC. Furthermore, H*i*-SET coder obtains best results than

the JPEG2000 coder and other coders that use a Hilbert Fractal for image compression. Hence, when the proposed perceptual quantization is introduced to H*i*-SET coder, called Φ_{SET}, our compressor increments its efficiency both objective and subjective. When ρGBbBShift method is compared against MaxShift method of the JPEG2000 standard Part II, images coded by our method get the best results, comparing the overall image quality.

Both the proposed perceptual quantization and ρGBbBShift method are general algorithms that can be applied to another Wavelet based image compression algorithms such as JPEG2000, SPIHT or SPECK, for instance.

Resumen

La compresión de imágenes es posible asumiendo que cierta parte de la información en ella es redundante. 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, el cual explota la alta correlación de un vecindario de pixeles por medio de una función Fractal, llamado Hi-SET, el cual posee las mismas características que tiene un compresor de imágenes moderno, como ser una algoritmo *embedded* que permite la transmisión progresiva. También se propone una modificación a la clásica cuantificación *Dead-zone*, la cual es aplicada a un grupo entero de pixeles en una sub-banda Wavelet dada. La modificación propuesta permite hacer una cuantificación directa e inversa pixel-por-pixel introduciendo una distorsión perceptual. Finalmente se establece 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 las frecuencias perceptualmente 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 H*i*-SET obtiene mejores resultados que los obtenidos por JPEG2000 o los de otros algoritmos que utilizan el fractal de Hilbert para comprimir. Así cuando a H*i*-SET se la aplica la cuantificación perceptual propuesta, llamándolo Φ_{SET}, se incrementa su eficiencia tanto objetiva como subjetiva. Cuando el método ρGBbBShift es comparado con el método MaxShift de JPEG2000 Parte II, se obtienen mejores resultados perceptuales comparando la calidad subjetiva de toda la imagen de dichos métodos.

Tanto la cuantificación perceptual propuesta como el método ρGBbBShift 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

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
e-CSF	extended Contrast Sensitivity Function
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
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
RCT	Reversible Component Transformation
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 in one hand to identify and to remove non-perceptual information of an image, maintaining as far as possible, the same entropy as the source image and the other hand to 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 better model some features of the Human Visual (HVS). Thus, an adequate model can be easily incorporated into many image processing applications such as Quality Assessment instuments 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 (56, 58) consider that MSE is a poor device to be used in quality assessment systems.

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 sys-

tem, 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 (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 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.

Therefore, this dissertation is centered in 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 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 in fewer bits. That is, it has more entropy per bit. 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(50, Sec. 1.2).

In general, a system is composed by four subsystems: an input, a precess, 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.

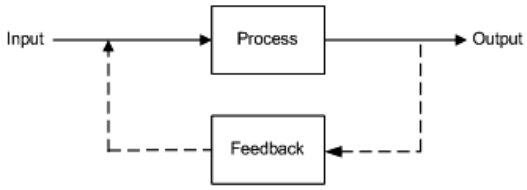


Figure 1.1: Description of *System* according to General System Theory.

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.

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, an alternative way of Entropy Coder, besides a perceptual image quality assessment, green blocks in Figure 1.3.

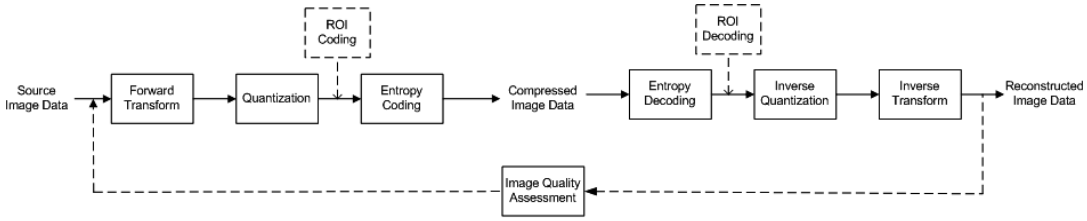


Figure 1.2: General Block Diagram for an image compression system.

Therefore the parts, that our systems 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), H*i*-SET Coding, H*i*-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, based on the interpretation of energy degradation.

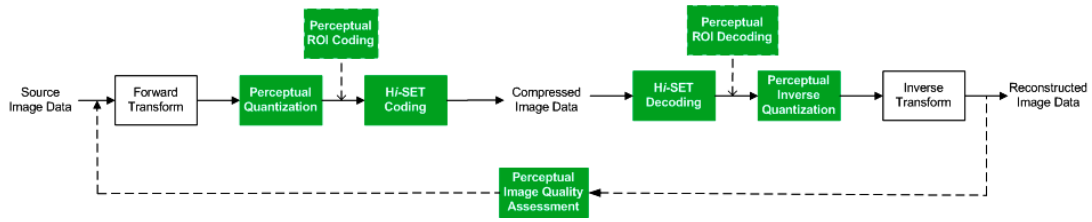


Figure 1.3: General Block Diagram for the proposed perceptual image compression system.

1.4 Thesis Outline

This dissertation consist of four chapters (4 to 5) that describe the new 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 watched at different distances. $\mathcal{C}_w$PSNR is the best-performing algorithm, 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 *H*ilbert *S*canning of *E*mbedded quad*T*rees (H*i*-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. H*i*-SET possesses also this latter feature. Furthermore, the H*i*-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 energy clustering both in frequency and space. The coding algorithm is composed of three general steps, using only one ordered list, the list of significant pixels.

The aim of Chapter 4 is to explain how to apply perceptual concepts in order to define a perceptual forward and inverse quantizer, which will be introduced to the H*i*-SET coder. The 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 H*i*-SET with the proposed perceptual quantizer (Φ_{SET}) shows that the latter is not favorable in PSNR, but the recovered image is more compressed 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 for codding of Region of Interest (ROI) areas. Introducing to the GBbBShift method perceptual criteria 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 and for not observing visual difference at ROI regarding MaxShift method, improving perceptual quality of the entire image.

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

Chapter 2

Full-Reference Quality Assessment using a Perceptual Chromatic Induction Model. Application to JPEG and JPEG2000 images

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 (56, 58) consider that MSE is a poor measure to be used in quality assessment systems. Therefore, it is important to know what are the MSE goodnesses and what are its problems in order to propose new image quality metrics that try to reproduce the properties of the Human Visual System (HVS).

In mathematical notation, let $f(i,j)$ and $\hat{f}(i,j)$ represent two images to be compared, 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 in comparison to $f(i,j)$

is being evaluated). The MSE and the PSNR are:

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

and

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

where $\mathcal{G}_{max}$ is the maximum possible intensity value in $f(i,j)$ ($M \times N$ pixels). That is, for 8 bit per pixel (bpp) gray-scale images we have $\mathcal{G}_{max} = 2^8 - 1 = 255$. For color images the PSNR is also defined by Eq. (2.2), where the color MSE is the mean MSE of the RGB components.

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

1. An efficient metrics for algorithm optimization. For example in JPEG2000, MSE is used both in Optimal Rate Allocation (5, 50) and Region of interest (6, 50).

2. MSE gives a clear meaning of the overall error signal energy because it is the energy of the difference signal between the two images.

However, the main problem of MSE is that it has a poor correlation with perceived image quality. An example is shown in Fig. 2.1, where both *Baboon*(a) and *Splash*(b) images are distorted by a JPEG2000 compression with PSNR=30 dB. These two noisy images have the same numerical image quality (e.g. PSNR) but they show dramatically different visual qualities. Thus, MSE and PSNR do not reproduces how the HVS perceives images.

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

Figure 2.1: 256×256 patches of Images *Baboon* and *Splash* compressed (i.e. distorted) by JPEG2000, both with PSNR=30dB. They are cropped for visibility.

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 used in section 2.3 in order to define the proposed $\mathcal{C}_w$PSNR image quality metrics. Section 2.4 shows experimental results, comparing $\mathcal{C}_w$PSNR with twelve image quality metrics such as MSSIM (53), SSIM (44) and VIF (57), among others. In these tests we use the perceptual image quality information supplied by the four image databases TID2008 (38, 39), LIVE (45), 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 $\mathcal{I}$ and an observation distance d, CIWaM obtains an estimation of the perceptual image $\mathcal{I}_\rho$ that the observer perceives when observing $\mathcal{I}$ 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).

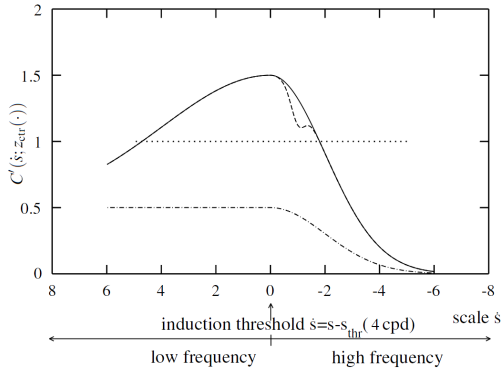


Figure 2.2: Continuous function: extended contrast sensitivity function. Dashed function: profile of e-CSF $C'(\dot{s}, z_{ctr}(s,o))$ with $z_{ctr}(x,y;s,o) = 0.75$. Dashed-dotted function: profile of $C_{min}(\dot{s})$. Dotted line: values above this value implies brightness contrast, and values below it implies brightness assimilation.

CIWaM takes an input image $\mathcal{I}$ 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 o. It is described as

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

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 *di*ag*on*a*l*.

The perceptual image $\mathcal{I}_\rho$ 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 $\mathcal{I}_\rho$ can be obtained by

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

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 $\mathcal{I}$ 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 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 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 CSF maximum value.

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

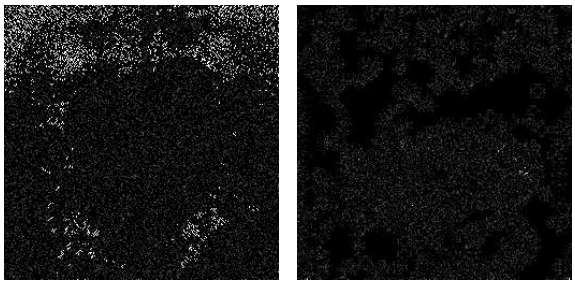
(a) Original image (b) d=30 cm. (c) d=100 cm. (d) d=200 cm.

Figure 2.3: (a) Original color image *Lenna* . (b)-(d) Perceptual images obtained by CIWaM at different observation distances d.

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

2.3 CIWaM weighted Peak Signal-to-Noise Ratio

In the referenced image quality problem, 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 characteristics of both the original image $f(i,j)$ (for example, spatial resolution, spatial frequencies, etc) and the distortion model Λ (for example, blurring, contrast change, noise, lossy compression, JPEG2000 compression, etc).



(a) (b)

Figure 2.4: Diagonal spatial orientation of the first wavelet plane, Images *Baboon*(a) and *Splash*(b) distorted by JPEG2000, PSNR=30dB.

Take the images *Babbon* and *Splash* compressed by JPEG2000 at PSNR=30 dB (Fig. 2.1). These two images have the same PSNR when compared to their corresponding original image, that is, they have the same numerical degree of distortion (i.e. the same numerical image quality PSNR). But, in contrast, their perceptual quality is clearly different, showing the image *Baboon* a better visual quality (i.e. it looks more similar to the original uncompressed image). Thus, PSNR numerical image quality and perceptual image quality do not correlate. On the image *Baboon* , high spatial frequencies are dominant. A modification of these high spatial frequencies by Λ induces a numerical distortion (i.e. a lower PSNR) but, in contrast, the modification of these high frequencies are not perceived by the HVS (i.e. we would expect a high PSNR). In contrast, on image *Splash*, mid and low frequencies are dominant. Modification of mid and low spatial frequencies also introduces a numerical distortion, but they are perceived by the HVS. This is the reason that, in this example, PSNR and perceptual image quality of these two images do not correlate. Fig. 2.4 shows the diagonal high

spatial frequencies of these two images, where they are more important frequencies in image *Baboon*.

Thus, given a particular distortion model (in the previous example, a JPEG2000 compression process at PSNR=30 dB) the visual perceptual quality of a distorted image $\hat{f}(i,j)$ depends on the characteristics of original image $f(i,j)$.

If we generate a set of distortions $\hat{f}_k(i,j) = \Lambda_k[f(i,j)]$ indexed by k (for example, let Λ be a blurring operator) we can study how the image quality of $\hat{f}_k(i,j)$ evolves while varying k, being k, for example, the degree of blurring. Depending on the characteristics of the original $f(i,j)$ image, the evolution of $\hat{f}_k(i,j)$ will be different. For example, if $f(i,j)$ contains many high spatial frequencies the PSNR will rapidly decrease when increasing k. In contrast, if $f(i,j)$ mainly contains low frequencies the PSNR will decrease slowly.

Similarly, the HVS is considered as a system that induces a distortion on the observed image $f(i,j)$. Hence, taking CIWaM as a mathematical model of the HVS, it is taken as a particular distortion model $\Lambda \equiv$ CIWaM that generates a perceptual image $\hat{f}(i,j) \equiv \mathcal{I}_\rho$ from an observed image $f(i,j) \equiv \mathcal{I}$, i.e $\mathcal{I}_\rho = CIWaM[\mathcal{I}]$. Similarly, a set of distortions is defined as $\Lambda_k \equiv \text{CIWaM}_d$, being d the observation distance. That is, we define a set of perceptual images $\mathcal{I}_{\rho,d} = \text{CIWaM}_d[\mathcal{I}]$ which are considered a set of perceptual distortions of image $\mathcal{I}$.

When images $f(i,j)$ and $\hat{f}(i,j)$ are simultaneously observed at distance d and this distance is reduced, we are able to better perceive the differences between them. In contrast, if they are observed from a far distance we are not able to perceive their differences, i.e. in this case the perceptual image quality of the distorted image is always high. The distance where the observer is not able to perceive any difference between this two images should be $d = \infty$. But, in practice, for similar enough images, the distance $d = \mathcal{D}$ where differences cannot be perceived ranges from some centimeters to few meters. Obviously, the more similar the two observed images, the smaller the distance D. In other words, the more similar $f(i,j)$ and $\hat{f}(i,j)$ (i.e. the highest the image quality of $\hat{f}(i,j)$), the smaller the distance D.

Since CIWaM is a model of the HVS, in the following section CIWaM is used to find 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. In addition, we want to estimate the perceptual image quality of $\hat{f}(i,j)$ compared to $f(i,j)$. The $\mathcal{C}_w$PSNR algorithm (graphically represented in Fig. 2.5) for assessing the perceptual image quality is defined by five general steps:

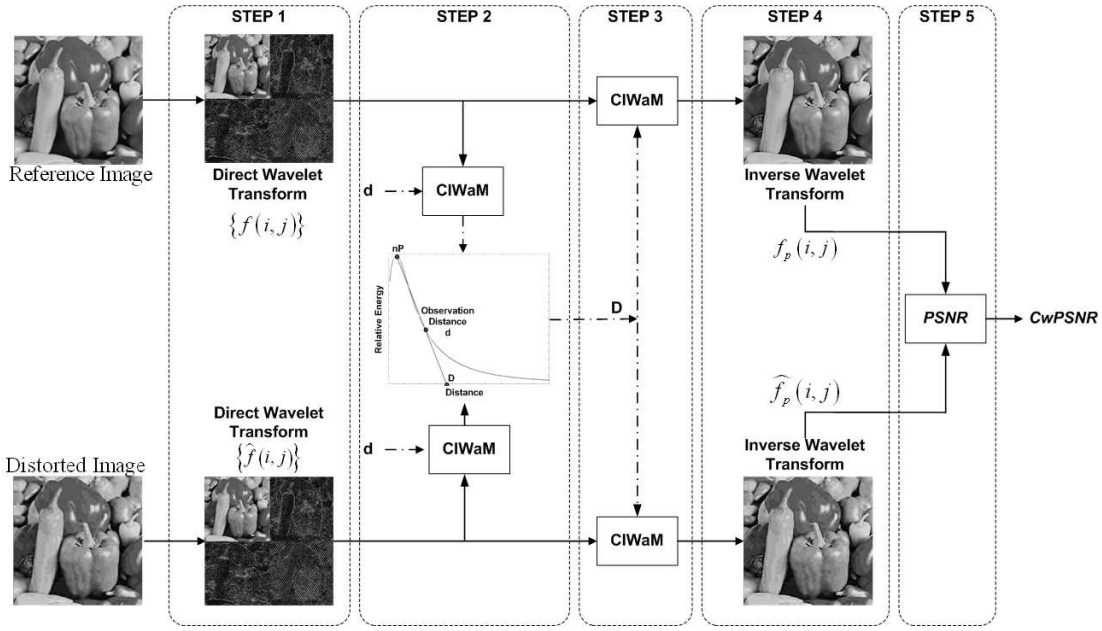


Figure 2.5: Methodology for PSNR weighting by means of CIWaM.

Step 1: Wavelet Transformation 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*(52) $\bar{\varepsilon}$ is the

absolute sum of the wavelet coefficient magnitudes, defined by (60)

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

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 means of the Equation 2.10, since a 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 $\{\omega_{s,o;\rho,d}\} = \alpha(\nu, r) \cdot \omega_{s,o}$ and $\{\hat{\omega}_{s,o;\rho,d}\} = \alpha(\nu, r) \cdot \hat{\omega}_{s,o}$ are obtained by applying CIWaM with an observation distance d. Their corresponding energies $\varepsilon_\rho(d)$ and $\widehat{\varepsilon}_\rho(d)$ are also obtained.

The relative energy ratio $\varepsilon\mathcal{R}(d)$ between these two wavelet coefficient sets are defined by

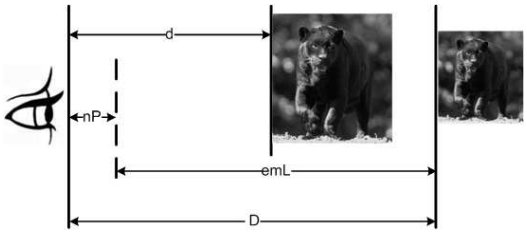
$$\varepsilon\mathcal{R}(d) = 10 \cdot \left| \log_{10} \frac{\varepsilon_\rho(d)}{\widehat{\varepsilon}_\rho(d)} \right| . \tag{2.11}$$

Fig. 2.6(b) shows an example of function $\varepsilon\mathcal{R}$ when d varies from 0 to ∞ centimeters.

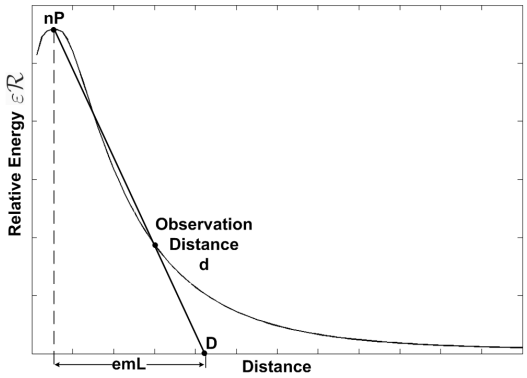
The peak $n\mathcal{P}$ at the $\varepsilon\mathcal{R}$ function is the distance $n\mathcal{P}$ where the relative ratio of the two energies is maximum. $\varepsilon\mathcal{R}(d)$ means measuring the relative energy ratio when both images are observed at distance d. At increasing distances the ratio is lower, until in $d = \infty$ the energies are equal. Our metrics is based on finding a lineal approximation to the distance D where the wavelet energies of the two images should be the same, that is, $\varepsilon\mathcal{R}(D) \approx 0$.

This is achieved by projecting the point $(n\mathcal{P}, \varepsilon\mathcal{R}(n\mathcal{P}))$ along the line that joins it with the point $(d, \varepsilon\mathcal{R}(d))$ to the point $(D, 0)$. Thus, distance D is the sum of distances $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$

$$\mathcal{D} = n\mathcal{P} + \varepsilon m\mathcal{L} . \tag{2.12}$$



(a) Portrayal of distances employed by the $\mathcal{C}_w$PSNR algorithm.



(b) Relative Energy function $\varepsilon\mathcal{R}$.

Figure 2.6: Definition of distances D, $n\mathcal{P}$ and $\varepsilon m\mathcal{L}$ both graphically (a) and in the $\varepsilon\mathcal{R}$ function (b).

where $\varepsilon m\mathcal{L}$ is the needed distance to match the energies from the point where the observer has the best evaluation of the assessed images to D and is defined by

$$\varepsilon m\mathcal{L} = \frac{\varepsilon\mathcal{R}(d)}{d\varepsilon\mathcal{R} + \sigma} \,, \tag{2.13}$$

where $d\varepsilon\mathcal{R}$ is the energy loss rate (wCal/cm) between $(n\mathcal{P}, \varepsilon\mathcal{R}\,(n\mathcal{P}))$ and $(d, \varepsilon\mathcal{R}\,(d))$, i.e. the negative slope of the line joining these points

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

When the two images are the same (for example, $\hat{f}(i,j)$ is a lossless version of $f(i,j)$), $f(i,j) = \hat{f}(i,j)$, 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 $d\varepsilon\mathcal{R} \neq 0$ (particularly, in our MatLab implementation $\sigma = realmin$).

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 (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: Obtain $\mathcal{C}_w$PSNR 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

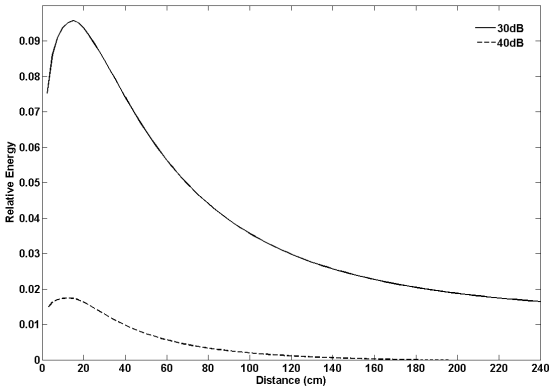
In this section we analyze the implications of three concepts of the $\mathcal{C}_w$PSNR algorithm; i.e. the $\varepsilon\mathcal{R}(n\mathcal{P})$ value, the distance D and the relation between these two points and the observation distance d. In short; first, the shape of the $\varepsilon\mathcal{R}$ function (and its value at $\varepsilon\mathcal{R}(n\mathcal{P})$) gives a first assessment of the image quality, second, 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.

For visually illustrating some of these characteristics, some images from the *Miscellaneous volume* of the University of Southern California, Signal and Image Processing

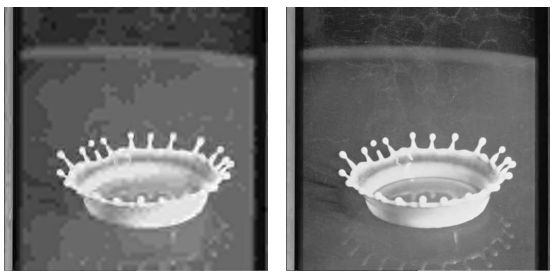
Institute image database (USC-SIPI image database, Figures A.5 and A.6) are used(2). All the distortions are implemented using JPEG2000 compression.

2.3.2.1 First Sub-indicator: $\varepsilon\mathcal{R}(n\mathcal{P})$

When several distorted versions of the same original image are compared with the latter, the shape of the $\varepsilon\mathcal{R}$ functions of these images gives an approximation of the perceived quality of them. When the $\varepsilon\mathcal{R}$ function is close to the horizontal axis (e.g. it is close to zero), the distorted image have a high image quality because there are few differences, at any distance, between it and the original image. For example, Figures 2.7 and 2.8 show that the images with PSNR=40 dB have a $\varepsilon\mathcal{R}\,(n\mathcal{P})$ lower than the ones with PSNR=30 dB.



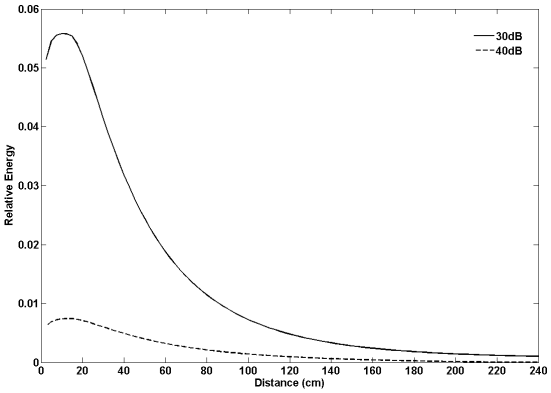
(a)



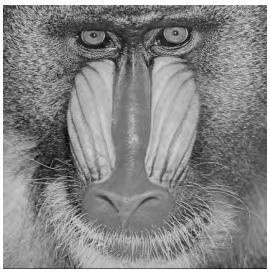
(b) PSNR=30dB (c) PSNR=40dB

Figure 2.7: (a) Relative Energy function $\varepsilon\mathcal{R}$ of Image *Splash*, distorted by JPEG2000 with (b) PSNR=30dB and (c) PSNR=40dB.

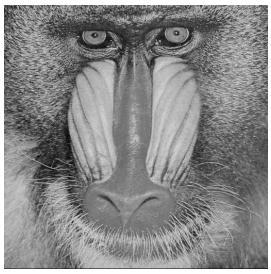
Thus, in the particular case where different distorted versions of the same original image are analyzed, the $\varepsilon\mathcal{R}(n\mathcal{P})$ value can be considered a first approximation to a



(a)



(b) PSNR=30dB



(c) PSNR=40dB

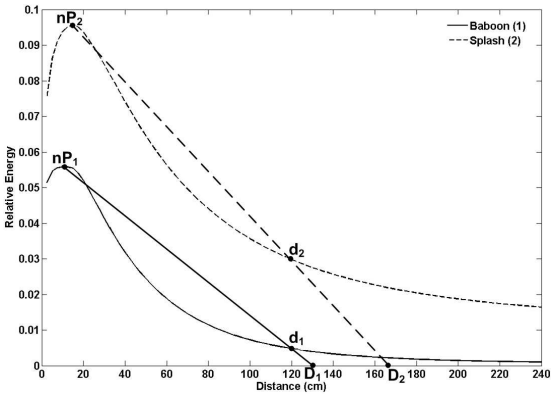
Figure 2.8: Relative Energy function $\varepsilon\mathcal{R}$ of Image *Baboon* (a), distorted by JPEG2000 with (b) PSNR=30dB and (c) PSNR=40dB.

perceptual image quality metrics.

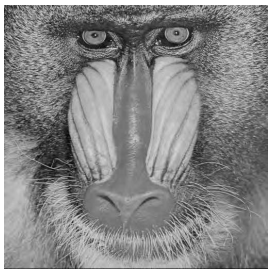
For example, in Figure 2.9 images *Baboon* and *Splash*, indexed by 1 and 2 respectively, are distorted with PSNR=30 dB. We can see that $\varepsilon\mathcal{R}(n\mathcal{P}_1) < \varepsilon\mathcal{R}(n\mathcal{P}_2)$. This clearly shows that the distorted image $Baboon_1$ has better perceptual image quality the one of $Splash_2$ and 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 $\varepsilon\mathcal{R}(n\mathcal{P})$ is not an accurate quality estimation. For example, in Fig. 2.10, the relative energy ratio $\varepsilon\mathcal{R}(n\mathcal{P}_2)$ of image *Sailboat on Lake* (Fig. 2.10(c))



(a) Relative Energy Chart



(b) D_1=130.36cm
$\mathcal{C}_w$PSNR=36.60dB

(c) D_2=167.46cm
$\mathcal{C}_w$PSNR=32.21dB

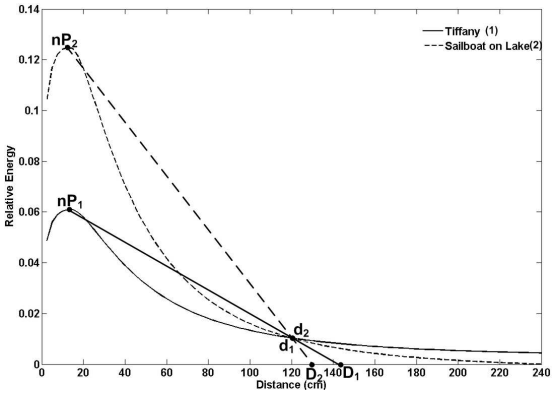
Figure 2.9: Relative Energy Chart of Images *Baboon* and *Splash* (a), which are distorted by means of JPEG2000 PSNR=30dB and Observation Distance d=120cm. Perceptual quality $\mathcal{C}_w$PSNR is equal to 36.60dB for (b) and 32.21dB for (c).

is twice the $\varepsilon\mathcal{R}(n\mathcal{P}_1)$ of image *Tiffany* (Fig. 2.10(b)), but image *Sailboat on Lake* has a better perceptual quality. Moreover, the PSNR of these two images is 31 dB. However, when $n\mathcal{P}_1$ is projected along d_1 to D_1, $D_1 > D_2$, that is, the distorted version of $Tiffany_1$ needs to be observer further than the $Sailboat\ on\ Lake_2$ image in order to perceive the same energy of its original image. It means that $Tiffany_1$ has a lower image quality.

Thus, distance D is a good approximation to an image quality estimator when the PSNR of the two images is the same.

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

Although, in general, D is a good approximation to image quality, there are situations where D is not a definitive image quality estimator. For example, in Fig. 2.10(c)



(a)

(b) D_1=141.45cm (c) D_2=129.67cm
$\mathcal{C}_w$PSNR=34.82dB $\mathcal{C}_w$PSNR=36.77dB

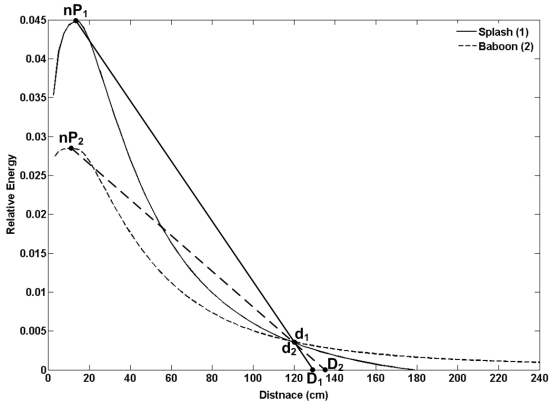
Figure 2.10: Relative Energy function $\varepsilon\mathcal{R}$ of Images *Tiffany* and *Sailboat on Lake* (a), distorted by JPEG2000 PSNR=31dB and observation distance d=120cm. Perceptual quality $\mathcal{C}_w$PSNR is (b) 34.82dB and (c) 36.77dB.

Sailboat on Lake$_2$ and Fig. 2.11(b) *Splash*$_1$ have practically the same D distance, e.g. $D_2 = D_1 = 129cm$, but subjective quality of *Splash*$_1$ is clearly better than the one of *Sailboat on Lake*$_2$.

The $\mathcal{C}_w$PSNR solution we present here to compare the perceptual quality of two images is to calculate the numerical quality (i.e. the PSNR) of their perceptual versions at distances D_1 and D_2. This way, $\mathcal{C}_w$PSNR predicts the image quality of an image at the distance D where the perceptual energy of the distorted image is linearly estimated to be equal to the perceptual energy of the original image. As shown in following sections, this approach allows us to estimate and compare the image quality of different images with different PSNR.

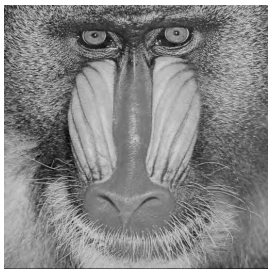
In the previous example, D is not a good image quality estimator. Now, we can see that the perceptual quality of Fig. 2.11(b) is correctly predicted by $\mathcal{C}_w$PSNR to be

twice less ($\sim 3dB$) than the perceptual image quality of Fig. 2.10(c).



(a)

(b) D_1=129.10cm PSNR=35.88dB



(c) D_2=135.89cm PSNR=31.74dB

Figure 2.11: Relative Energy function $\varepsilon\mathcal{R}$ of Images *Splash* and *Baboon*, distorted by JPEG2000 with $\mathcal{C}_w$PSNR=39.69 dB and observation distance d=120cm. Objective quality PSNR (b) 35.88dB and (c) 31.74dB.

2.4 Experimental Results

In this section, $\mathcal{C}_w$PSNR performance is compared with the psychophysical results obtained by human observers when judging the visual quality of some image databases. The results obtained by human observers are expressed in Mean Opinion Scores either diferencial (DMOS) or not (MOS). Hence, $\mathcal{C}_w$PSNR is tested across four image databases that supply the perceived image quality:

1. Tampere Image Database (TID2008) of the Tampere University of Technology, 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 (45).

3. Categorical Subjective Image Quality Image Database (CSIQ) of the Oklahoma State University(22).

4. Image and Video-Communication Image Database (IVC) of the Université de Nantes(23).

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. 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. The rating of the perceptual image quality is presented using the MOS index.

LIVE Database contains 29 original images (Fig. A.3), with 26 to 29 altered versions for every original image. In addition, rating of perceptual quality for distorted images is given in DMOS values. Concretely, LIVE includes 234 and 228 distorted images for JPEG and JPEG2000 compression degradation, respectively.

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.

CSIQ Database includes 30 original images (Fig. A.4), which are distorted by 6 different types of distortions at 4 or 5 degrees. This way, for JPEG and JPEG2000 compression distortions, CSIQ Database contains 150 distorted versions of these two degradations. CSIQ Database has 5000 perceptual evaluations of 25 observers and its assessments are reported in DMOS values.

In this work,we use the JPEG and JPEG2000 distortions.

2.4.1 Performance Measures

Strength of Relationship (SR) indicates how strong is the correlation between two variables. The SR analyzes 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 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) . Since image quality metrics obtained from human observers have no lineal relationship, it is not convenient to employ PCC, since even PSNR and MSE are the same metrics, PCC calculates different values. SROCC is a better choice for measuring SR between the opinion of observers and the results of a given metrics. However, when SROCC is used for testing a null hypothesis that is finally rejected, it is difficult to interpret (17). The Kendall Rank-Order Correlation Coefficient (KROCC) corrects this problem by reflecting SR between compared variables. Furthermore KROCC and it allows to estimate how similar are two rank-sets against the same object set. KROCC is interpreted as the probability to rank two sets in the same order taking into account the number of inversions of pairs of objects for transforming one rank into the other(1). Thus, KROCC shows an accurate assessment of SR between a metric and the opinion of an human observer. One of the limitations of KROCC is its complexity, which is higher than PCC and SROCC.

Furthermore, $\mathcal{C}_w$PSNR is compared with some state of the art numerical image quality estimators. Concretely, $\mathcal{C}_w$PSNR is compared to MSE(18), PSNR(18), SSIM(44), MSSIM(53), VSNR(12), VIF(57), VIFP(44), UQI(54), IFC(46), NQM(14), WSNR(25) and SNR.

We chose for each 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 considering:

- Gamma correction: Before CIWaM is applied to images, we apply a gamma correction $\gamma = 2.2$.
- Observation distance: Taking into account the observation distances of the several databases, we take d=8H, 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. For the experiments in this Section: $M_{size} = 19''$, $h_{res} = 1280$ and $v_{res} = 1280$.
- Wavelet Transform: Images are decomposed into three wavelet planes, that is $n = 3$, (Eq. 3.5).

2.4.2 Overall Performance

Table 2.3: KROCC of $\mathcal{C}_w$PSNR and other quality assessment algorithms on multiple image databases using JPEG distortion. 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.

	Image Database				
Metrics	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 shows the KROCC values for $\mathcal{C}_w$PSNR and the other twelve image quality estimators, for all the databases and using a JPEG compression distortion. Last column of Table 2.3 shows the performance obtained taking all the images (e.g. 534) from the cited image databases. Bold and italicized values represent the best and the second best performance assessments, respectively. $\mathcal{C}_w$PSNR is the best performer in all the databases and also on the union of them (last column). MSSIM is the second best-ranked metrics not only in all databases but also on their union, except for the CSIQ database where VIF is the second best. On the union of all the databases, $\mathcal{C}_w$PSNR correlation is 0.0243 higher than MSSIM and improves the performance of PSNR or MSE by 0.1402 for JPEG compression degradation.

Table 2.4 shows the performance of $\mathcal{C}_w$PSNR and the same twelve metrics for JPEG2000 compression distortion. In this case, $\mathcal{C}_w$PSNR is also the best metrics for all the databases. On the union of all the databases, our algorithm is again the best one. For this distortion, MSSIM is also the second best indicator for TID2008, LIVE and

Table 2.4: KROCC of $\mathcal{C}_w$PSNR and other quality assessment algorithms on multiple image databases using JPEG2000 distortion. 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.

	Image Database				
Metrics	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**

IVC image databases in addition to the databases union. For CSIQ image database, VIFP is the second best. We can see that global $\mathcal{C}_w$PSNR correlation is 0.0143 higher than MSSIM. Furthermore, $\mathcal{C}_w$PSNR improves the correlation of PSNR by 0.1155.

In summary, $\mathcal{C}_w$PSNR is the best image quality estimation algorithm for JPEG and JPEG2000 compression distortions, that is, for image compression algorithms that use either Discrete Cosine Transform or Wavelet Transform as a method of pixel transformation in samples for the quantization process (50, pg. 14).

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 (43), SPIHT (41) and SPECK (34, 35, 36), the execution path defines the correct order of the coefficients by comparison of its branching points (51). 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 H*i*-SET image compressor for the encoding and decoding processes. 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. 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

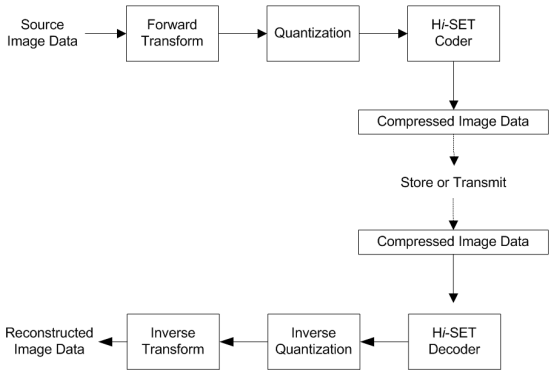


Figure 3.1: General block diagram of a generic compressor that uses H*i*-SET for encoding and decoding.

Quantization block introduces distortion and is only employed for lossy compression. In the following step, H*i*-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 H*i*-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.

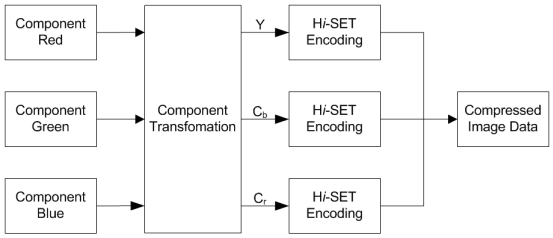


Figure 3.2: H*i*-SET multiple component encoder.

In this way, an RGB color image is decomposed into three components, namely Red, Green, and Blue color components. RGB color components are statistically more dependent between them that than YC_rC_b color components. Figure 3.2 depicts that

when H*i*-SET performs a 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 (58).

H*i*-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 (47, 50).

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

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

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 H*i*-SET is separated into different spatial frequency and orientation components using a multiresolution discrete wavelet decomposition (DWT) either reversible or irreversible (3, 48). Thus $\mathcal{I}$ 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:

$$DWT\left\{\mathcal{I}\right\} = \sum_{s=1}^{n} \sum_{o=v,h,d} \omega_s^o + c_n \tag{3.5}$$

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 $\mathbb{J}$ 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

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

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

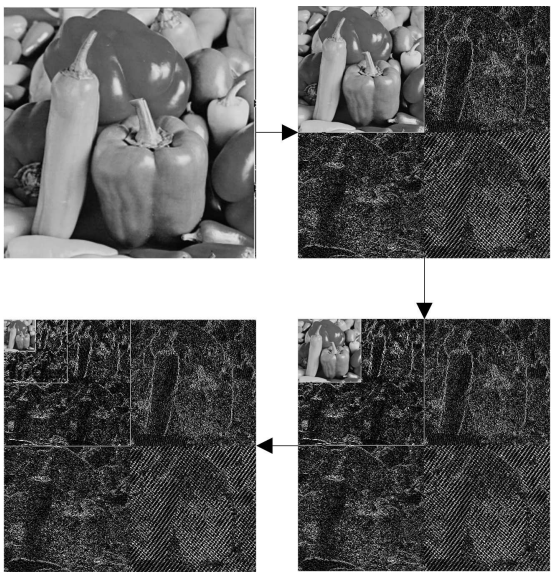


Figure 3.3: Three-level wavelet decomposition of the *Peppers* image.

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 3.4 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.

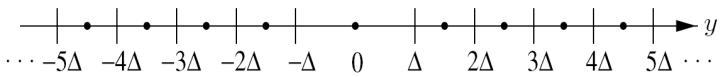


Figure 3.4: Dead-zone uniform scalar quantizer with step size Δ: vertical lines indicate the endpoints of the quantization intervals and heavy dots represent reconstruction values.

The International Organization for Standardization recommends to adopt the mid-point 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 is going to set them to q=0.

Thus taking a wavelet plane Δ_s^o, an array of quantized coefficients is obtained for further losslessly encoding, since the image degradations are induced only by the Quantization process.

3.5 The H*i*-SET Algorithm

3.5.1 Startup Considerations

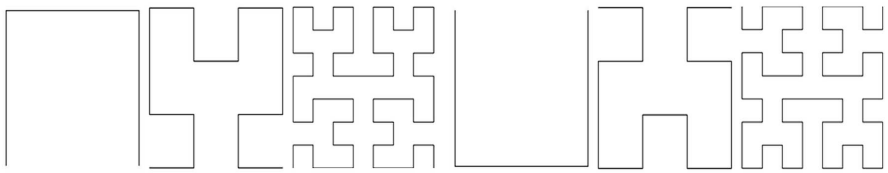
3.5.1.1 Hilbert space-filling Curve

The Hilbert curve is an iterated function that can be 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*, $\mathcal{W} \rightarrow \mathcal{X} \rightarrow \mathcal{Y} \rightarrow \mathcal{Z}$), $\mathcal{L}$ (*left*, $\mathcal{W} \rightarrow \mathcal{Z} \rightarrow \mathcal{Y} \rightarrow \mathcal{X}$), $\mathcal{R}$ (*right*, $\mathcal{Z} \rightarrow \mathcal{W} \rightarrow \mathcal{X} \rightarrow \mathcal{Y}$), and $\mathcal{D}$ (*down*, $\mathcal{X} \rightarrow \mathcal{W} \rightarrow \mathcal{Z} \rightarrow \mathcal{Y}$). Where $\rightarrow$ 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 3.5: First three levels of a Hilbert Curve. (a) Axiom = $\mathcal{D}$ proposed by David Hilbert in (16). (b) Axiom = $\mathcal{U}$ employed in this work.

The original work by David Hilbert (16) proposes an axiom with a $\mathcal{D}$ trajectory (Figure 3.5(a)), while we propose to start with an $\mathcal{U}$ trajectory (Figure 3.5(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 3.5.

The production rule of the Hilbert Curve is 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 3.6(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
if $\gamma = 1$ **then**
  $\theta = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$
else
  $\beta =$ **Algorithm 1** $(\gamma - 1)$
  $\theta = \begin{bmatrix} \beta^T & (\widetilde{\beta})^T + (3 \times 4^{\gamma-1}) \\ \beta + 4^{\gamma-1} & \beta + (2 \times 4^{\gamma-1}) \end{bmatrix}$

also defines how these subsets are formed and what are the coefficients 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 can be 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), H*i*-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: $\mathbb{S}_o$, thr
Result: $\mathbb{S}_1$, $\mathbb{S}_2$, $\mathbb{S}_3$, $\mathbb{S}_4$ and $\widehat{\mathcal{H}_{sig}}$
$\gamma = \log_4(length\ of\ \mathbb{S}_o)$
The cell 1 of the subsets $\mathbb{S}_1$, $\mathbb{S}_2$, $\mathbb{S}_3$ and $\mathbb{S}_4$ is declared with $4^{\gamma-1}$ elements, while the cell 2 with just one element.
$i = 1$
$\widehat{\mathcal{H}_{sig}}$ is emptied.
for j=*1* **to** 4^{γ} **do**
    Store $\mathbb{S}_o\left[from\ j\ to\ \left(i \times 4^{\gamma-1}\right)\right]$ into $\mathbb{S}_i(1)$.
    if $2^{thr} \leq \max|\mathbb{S}_i(1)| < 2^{thr+1}$ **then**
        $\mathbb{S}_i(2) = 1$
        Add **1** at the end of the $\widehat{\mathcal{H}_{sig}}$.
    else
        $\mathbb{S}_i(2) = 0$
        Add **0** at the end of the $\widehat{\mathcal{H}_{sig}}$.
    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, H*i*-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, H*i*-SET uses not only spare lists, but also spare cell arrays, both are denoted by an apostrophe, LSP', $\widehat{\mathcal{H}}'$ or $\mathbb{S}'_1$, 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: 𝕊1, 𝕊2, 𝕊3, 𝕊4, thr and γ
   Result: LSP and Ĥ
 1 LSP and Ĥ are emptied.
 2 if γ = 0 then
 3     for i = 4 to 1 do
 4         if 𝕊i(2) is significant then
 5             Add 𝕊i(1) at the beginning of the LSP.
 6             if 𝕊i(1) is positive then
 7                 Add 0 at the beginning of the Ĥ.
 8             else
 9                 Add 1 at the beginning of the Ĥ.
10 else
11     for i=1 to 4 do
12         if 𝕊i(2) is significant then
13             Call Algorithm 2 with 𝕊i(1) and thr as input data and Store the results into 𝕊'1, 𝕊'2, 𝕊'3,
               𝕊'4 and Ĥ'.
14             Add Ĥ' at the end of the Ĥ.
15             Call Algorithm 3 with 𝕊'1, 𝕊'2, 𝕊'3, 𝕊'4, thr and γ − 1 as input data and Store the results
               into Ĥ' and LSP'.
16             Add Ĥ' at the end of the Ĥ.
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 bit 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 H*i*-SET, a simple example is shown.

The wavelet transform coefficient matrix $\mathcal{H}$ of an 8×8 pixels image is depicted in Figure 3.6(a), which is a three scale ($n = 3$) transformation, which implies $\gamma = 3$. The indexed vector $\overrightarrow{\mathcal{H}}$ (Figure 3.6(c)) is acquired interleaving $\mathcal{H}$ with a three-level matrix θ (Figure 3.6(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 3.7 depicts the fractal partitioning diagram of the first bit-plane encoding.

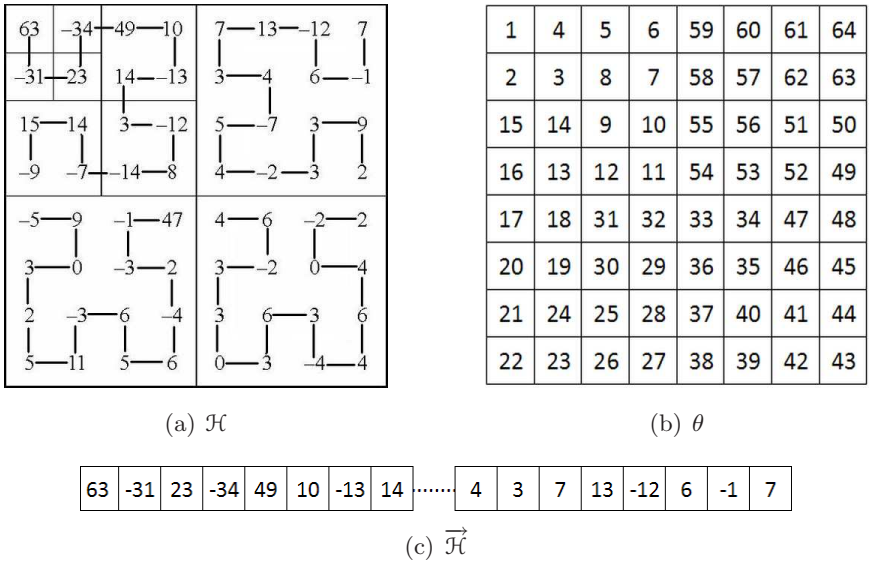
The following remarks refer to steps of the Table 3.3:

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 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) $\mathcal{H}$ (b) θ

(c) $\overrightarrow{\mathcal{H}}$

Figure 3.6: Example of Hilbert indexing of an 8×8 pixels image. (a) Three-scale wavelet transform matrix $\mathcal{H}$. (b) Hilbert Indexing matrix θ when $\gamma = 3$. (c) Interleaved resultant vector $\overrightarrow{\mathcal{H}}$.

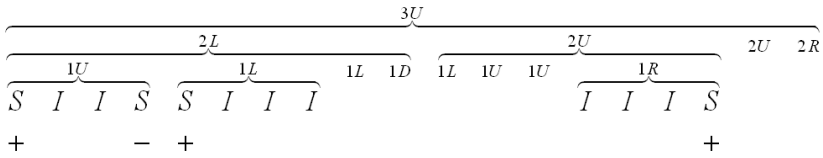


Figure 3.7: Fractal partitioning diagram of the first bit-plane encoding, using H*i*-SET scheme.

Table 3.3: The First bit-plane encoding using H*i*-SET scheme. $\mathcal{H}$, θ and $\overrightarrow{\mathcal{H}}$ are taken from Figure 3.6, 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	$1\mathcal{U}$	$\mathcal{SIIS}$	1001	
5	*sign*	$+-$	01	+32 −32
6	$1\mathcal{L}$	$\mathcal{SIII}$	1000	
7	*sign*	$+$	0	+32 −32 +32
8	$2\mathcal{U}$	$1\mathcal{L}\mathcal{U}\mathcal{U}\mathcal{R}$	0001	
9	$1\mathcal{R}$	$\mathcal{IIIS}$	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 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 3.8).

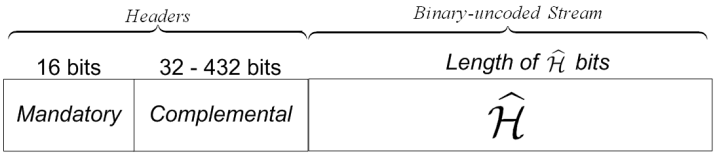


Figure 3.8: H*i*-SET Codestream Syntax.

Headers are subdivided in groups of Markers. We consider two types: Mandatory and Complemental Headers. Figure 3.9(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.

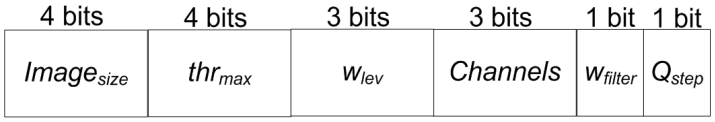
thr_{max} (4 bits). It stores the maximum threshold $thr - 1$ defined in eq (3.9), hence, its value varies from 1 to 16. Thus, H*i*-SET can process an image up to 16 bit-planes.

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.

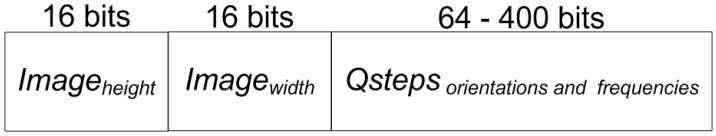
$Channels$ (3 bits). The number of image (color) components minus one is stored in this marker, thus managing up to eight components.

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 3.9: H*i*-SET Headers with their Markers.

Figure 3.9(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 3.10).

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

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

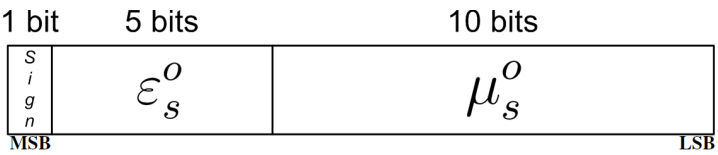


Figure 3.10: Structure of the Δ_s^o Sub-marker.

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. Quality of the recovered image is obtained by comparing it to 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 3.11 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.

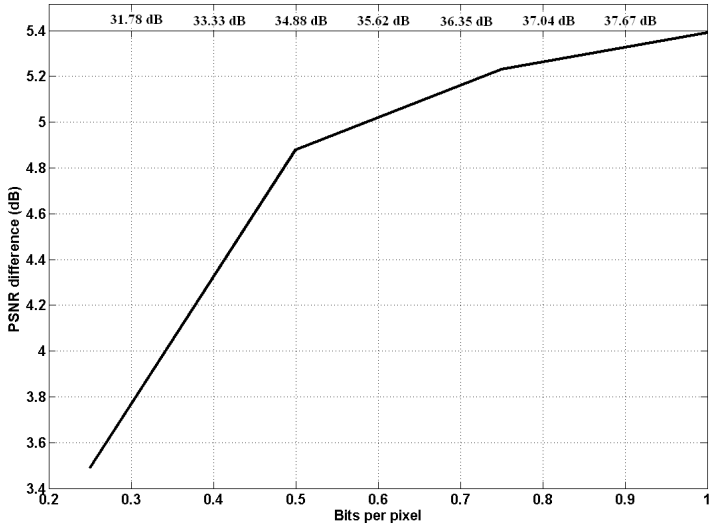
Figure 3.11(a) shows the PSNR difference between H*i*-SET and Kim and Li algorithm as a function of the bpp. On the upper horizontal axis we show the PSNR obtained at the bpp shown on the lower horizontal axis. On average, H*i*-SET reduces PSNR in 4.75 dB (i.e. reduces the Mean Square Error around 63.07 %).

Similarly, Figure 3.11(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.

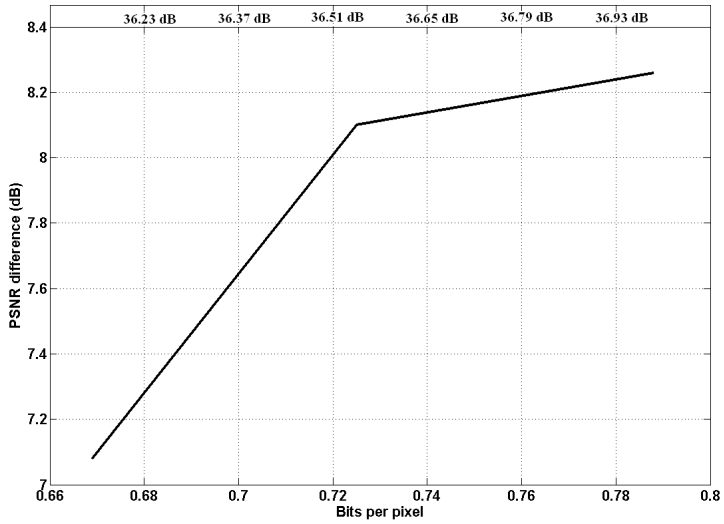
Thus, on average our method improves the image quality of these two Hilbert fractal based methods in approximately 6.20 dB.

3.7.2 Comparing H*i*-SET and JPEG2000 coders

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 (50) and H*i*-SET



(a) Kim and Li algorithm



(b) Biswas algorithm

Figure 3.11: Performance comparison (PSNR difference) between H*i*-SET and the algorithms proposed by and , for a gray-scale image *Lenna*. On the upper part of the figures we show the PSNR obtained at the bpp shown on the lower part.

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 $\mathcal{I}_{org}$ such that $\mathcal{I}_{org} \geq 2^{thr-bpl+1}$, being *bpl* the desired bit-plane and

thr the maximum threshold

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

These selected coefficients are inverse wavelet transformed in order to create a new source of image data, i.e. $\mathcal{I}'_{org}$, which are losslessly compressed, that is until the last bit-plane, by each coder. Figure 3.12 depicts this process. The software used to perform JPEG2000 compression is *Kakadu* (49) and *JJ2000* (40). The irreversible component transformation (ICT, YC_bC_r) is used in addition to the 9/7 irreversible wavelet transform.

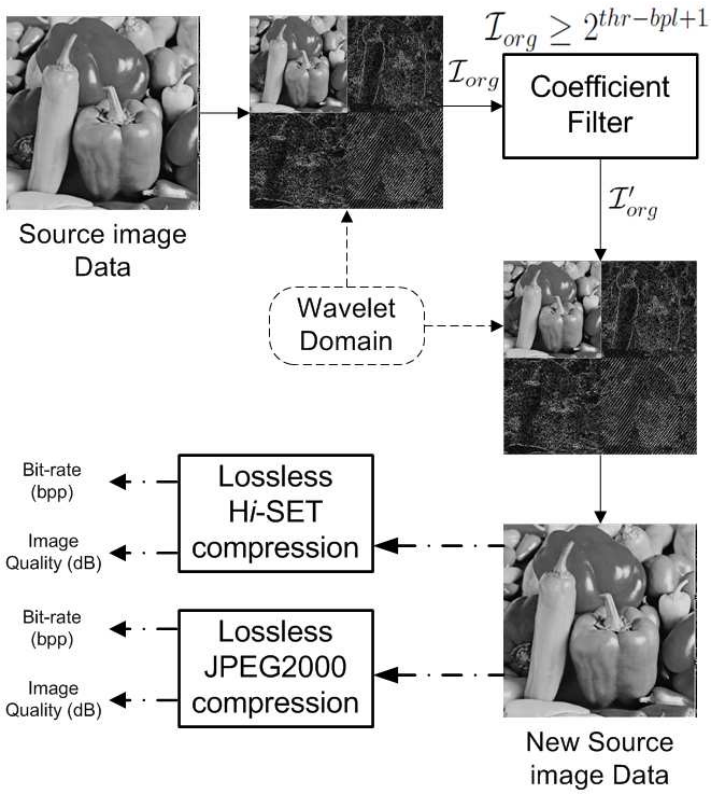


Figure 3.12: Bit-plane selection. Some coefficients are selected provided that they fulfil the current threshold.

H*i*-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.

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 3.13 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 (dashed function) at 1.59 bpp (1:5.04 compression ratio) in 2.38 KBytes and by H*i*-SET (continuous function) at 1.10 bpp (1:7.3 ratio) in 1.64 KBytes. In Figure 3.14 we can see this differences when the image *kodim18* is compressed at 0.8 bpp by JPEG2000 and H*i*-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 H*i*-SET with the same objective visual quality. At the same compression rate H*i*-SET is 1.84 dB better.

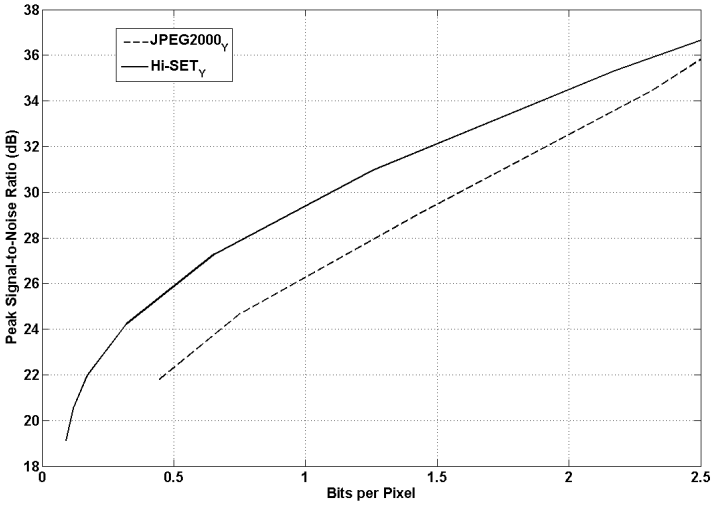
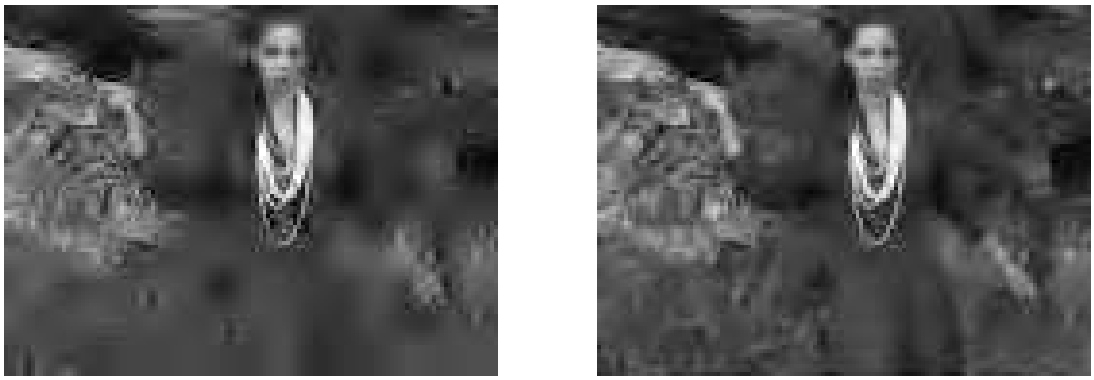


Figure 3.13: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 1: Compression rate vs image quality of the 128×96 gray-scale image database.

Experiment 2 Medium resolution gray-scale images. In this experiment, the source image data both for the JPEG2000 standard coder and H*i*-SET algorithms are the selected images from the TID2008 (Figure A.2) transformed into



(a) JPEG2000 23.99 dB (b) H*i*-SET 26.35 dB

Figure 3.14: Experiment 1. Example of 128×96 reconstructed image *kodim18* compressed at 0.8 bpp (Y Component).

gray-scale images (Y component). Figure 3.15 shows the average quality of the recovered images as a function of compression rate, for both JPEG2000 (dashed 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. The difference in visual quality between JPEG2000 and H*i*-SET when the image *kodim23* is compressed at 0.2 bpp can be seen in Figure 3.16. The image quality of the recovered image coded by JPEG2000 (a) is 2.74 dB lower than the one obtained by H*i*-SET(b).

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 3.17 shows the PSNR of recovered images as a function of compression rate. On the average, an image compressed by H*i*-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 3.18 we can see these differences when image *kodim06* is compressed at 1.4 bpp by JPEG2000 standard (a) and H*i*-SET(b). Thus, at the same compression rate, H*i*-SET obtains a better

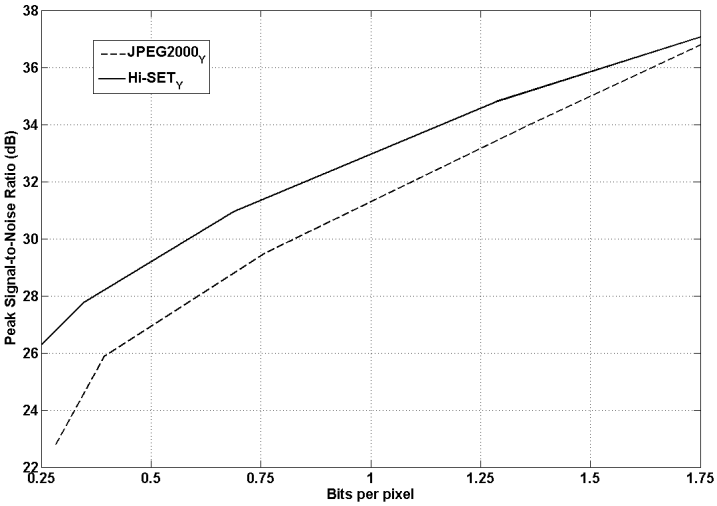


Figure 3.15: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 2: Compression rate vs image quality of the original image database in gray-scale.

(a) JPEG2000 27.05 dB

(b) H*i*-SET 29.79 dB

Figure 3.16: Experiment 2. Example of 512×384 reconstructed image *kodim23* compressed at 0.2 bpp (Y Component).

image quality (up to 2.26 dB better) than JPEG2000 coder. On avergage, H*i*-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.

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 3.19 shows the rela-

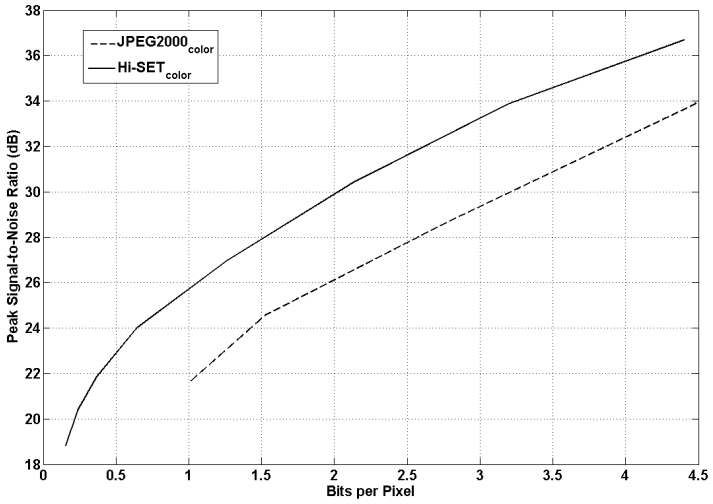
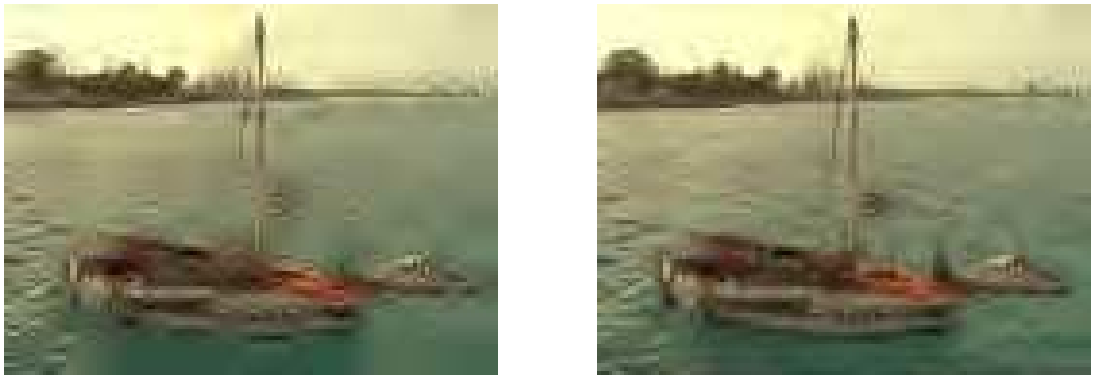


Figure 3.17: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 3: Compression rate vs image quality of the 128×96 color image data base.



(a) JPEG2000 25.99 dB

(b) H*i*-SET 28.25 dB

Figure 3.18: Experiment 3. Example of 128×96 reconstructed image *kodim06* compressed at 1.4 bpp (Y, C_b and C_r Components).

tion between compression rate and average quality. On average, a 512×384 image compressed by H*i*-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 3.20 we can see the difference when the image *kodim04* is compressed at 0.4 bpp by JPEG2000 (a) and H*i*-SET(b). At the same compression 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.

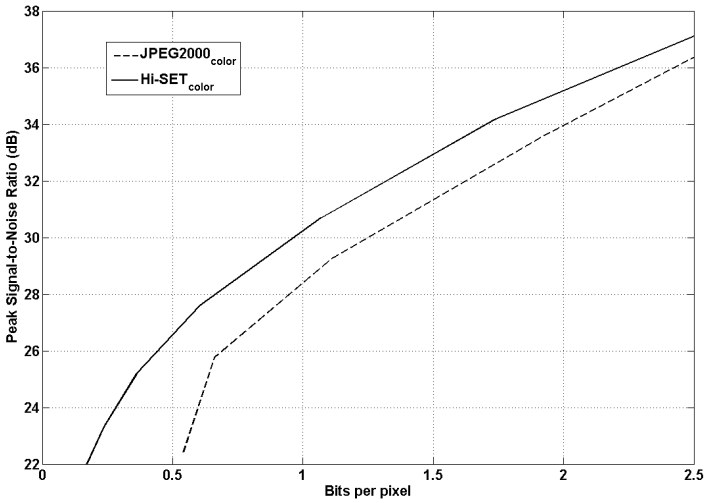
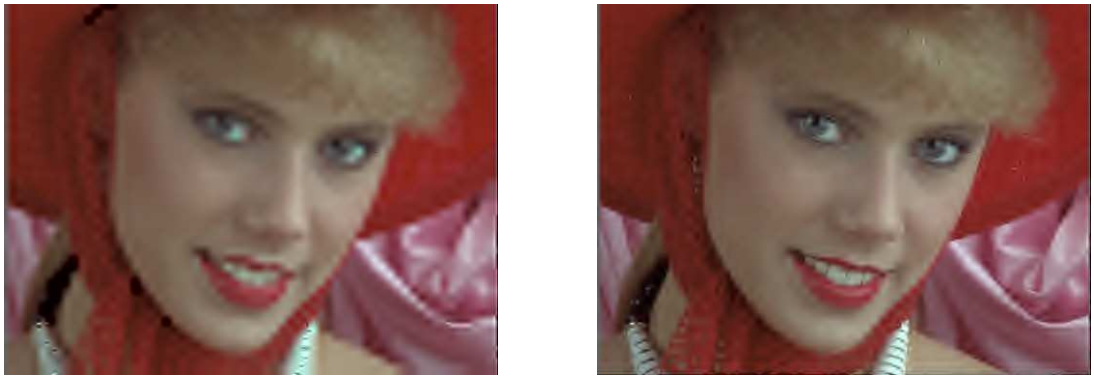


Figure 3.19: Comparison between H*i*-SET and JPEG2000 image coders. Experiment 4: Compression rate vs image quality of the original color image data base.



(a) JPEG2000 28.53 dB

(b) H*i*-SET 30.36 dB

Figure 3.20: Experiment 4. Example of 512 × 384 reconstructed image *kodim04* compressed at 0.4 bpp (Y, C_b and C_r Components).

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. 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 3.21 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 3.21 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*.

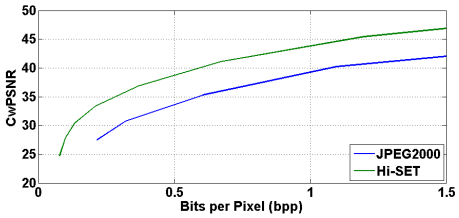
bpp (rate)	JPEG2000 PSNR in dB's	H*i*-SET PSNR in dB's
0.25 (32:1)	19.08	23.82
0.50 (16:1)	24.91	28.00
0.75 (10.67:1)	29.65	31.30

H*i*-SET and JPEG2000 are also 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(44), MSSIM(53), VSNR(12), VIF(57), VIFP(44), UQI(54), IFC(46), 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), CSIQ(Sec. A.4), IVC(Sec. A.1), LIVE(Sec. A.3) and TID2008(Sec. A.2), for color and gray-scale (Y Channel) compression. But in this Section, only $\mathcal{C}_w$PSNR results are exposed, Fig. 3.22, the rest of metrics can be shown in Annex B. Thus, Fig. 3.22 also demonstrates that H*i*-SET significantly improves the results of JPEG2000 coder.

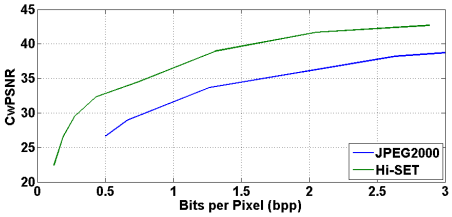
(a) JPEG2000 19.48 dB

(b) H*i*-SET 26.56 dB

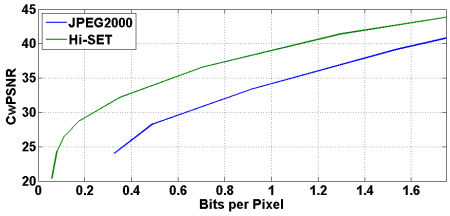
Figure 3.21: Experiment 5. Examples of 2048×2560 reconstructed image *Bicycle* compressed at 0.38 bpp (Y Component).



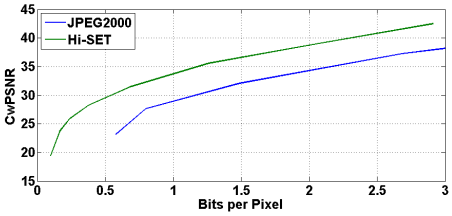
(a) CMU gray-scale



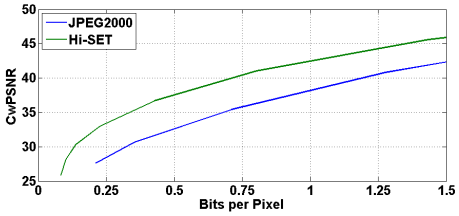
(b) CMU color



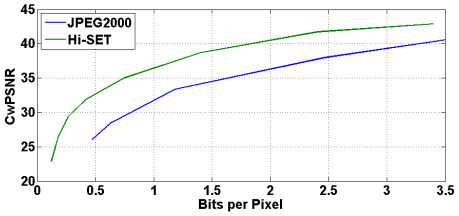
(c) CSIQ gray-scale



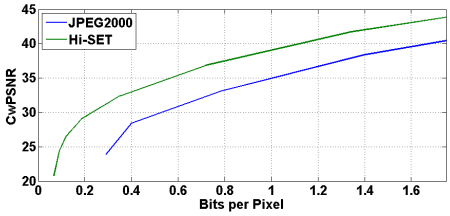
(d) CSIQ color



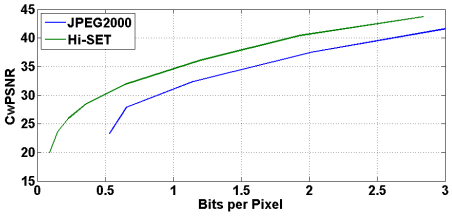
(e) IVC gray-scale



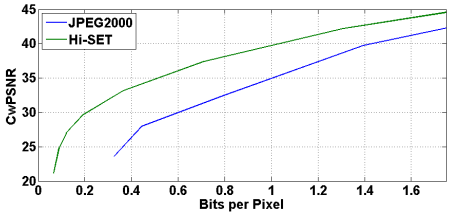
(f) IVC color



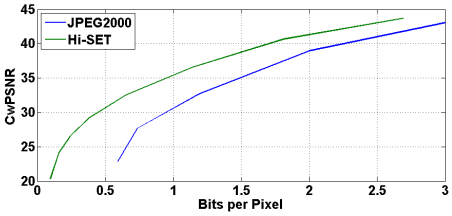
(g) LIVE gray-scale



(h) LIVE color



(i) TID2008 gray-scale



(j) TID2008 color

Figure 3.22: Comparison between JPEG2000 vs H*i*-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 weighting for 400 dpi's

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:

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

Unlike the classical technics of Visual Frequency Weighting (VFW) on JPEG2000, which apply one CSF weight per sub-band (10, Annex J.8), Perceptual Quantization through CIWaM applies one CSF weight per coefficient over all wavelet planes $\omega_{s,o}$. Thus Equation 4.1 introduces the perceptual criteria of Equation 2.4 to each quantized coefficient of Equation 3.6. 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 Local VFW 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 images were calculated 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 was *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 (Section 3.4, continuous function with heavy dots), and adding to it a previous quantization step developed by CIWaM (function with heavy stars).

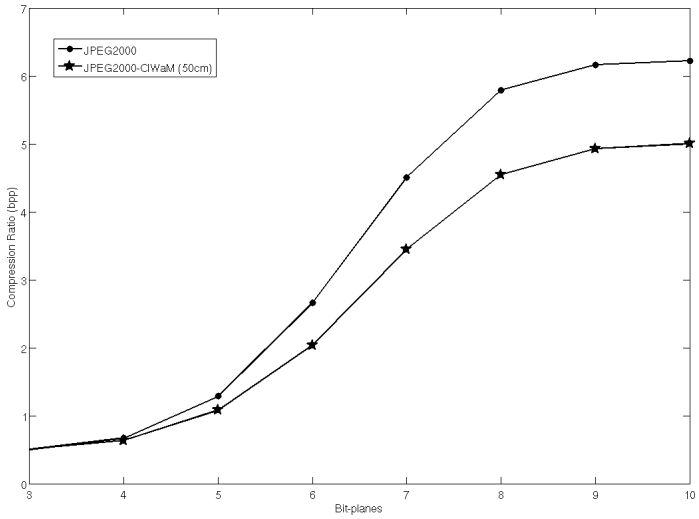


Figure 4.1: JPEG2000 Compression ratio by Bit-plane. Function with heavy dots: JPEG2000 only quantized by the dead-zone uniform scalar manner. Function with heavy stars: JPEG2000 perceptually pre-quantized by the chromatic induction wavelet model, in addition to a dead-zone uniform scalar quantification.

CIWaM used as a method of forward quantization, achieves better compression ratios with the same threshold, reaching better results at the highest bit-planes, since CIWaM reduces unperceivable coefficients. Figure 4.2 shows the contribution of CIWaM in the JPEG2000 compression ratio, for example at the eighth bit-plane, CIWaM diminishes 1.2423 bits per pixel less than without it, 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 reconstructed images compressed at 0.9 and 0.4 bits per pixel, respectively, by means of JPEG2000 without (a) and with perceptual pre-quantization (b). Also this figures demonstrate that the CIWaM subjective quality is higher than the objective one.

The Figure 4.5 shows examples of reconstructed images of *Baboon* compressed at 0.59, 0.54 and 0.45 bits per pixel by means of JPEG2000 without (a) and with perceptual pre-quantization (b and c). PSNR in Fig. 4.5(a) is 26.18 dB and in Fig. 4.5(b)

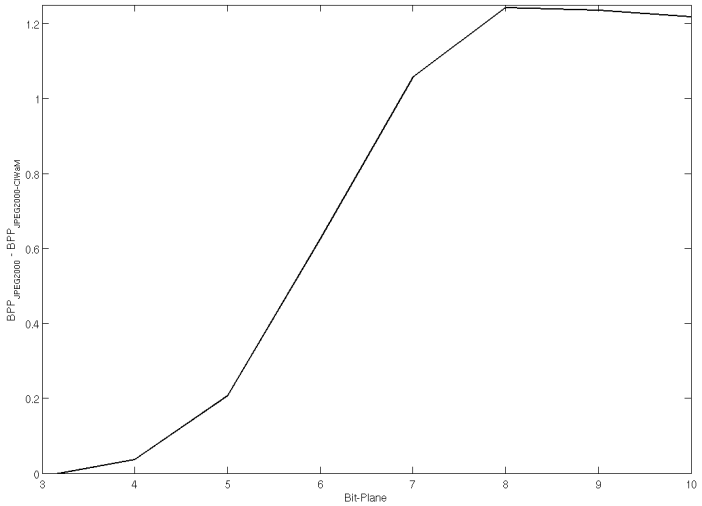
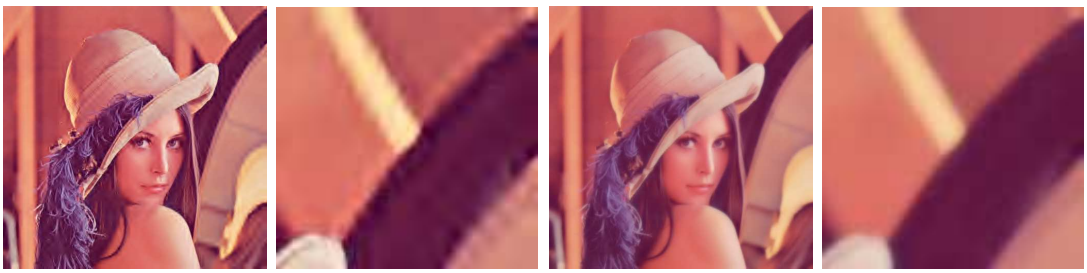


Figure 4.2: Contribution of a CIWaM pre-quantification over the JPEG2000 compression ratio by each Bit-plane.



(a) JPEG2000 31.19 dB. (b) JPEG2000-CIWaM 27.57 dB.

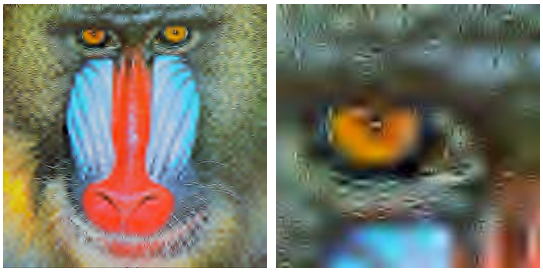
Figure 4.3: Examples of reconstructed images of Lena compressed at 0.9 bpp.

(a) JPEG2000 25.12 dB. (b) JPEG2000-CIWaM 24.57 dB.

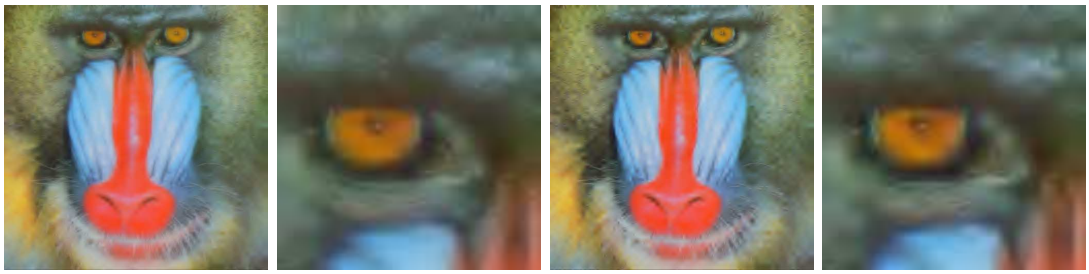
Figure 4.4: Examples of reconstructed images of F-16 compressed at 0.4 bpp.

is 26.15 dB but a perceptual metric like WSNR (25), for example, assesses that it is equal to 34.08 dB. Therefore, the reconstructed image pre-quantized by CIWaM is per-

ceptually better than the one just quantized by a Scalar Quantizer, since the latter has more compression artifacts, even the result at 0.45 bpp (Fig. 4.5(c)) has less artifacts, showing for example that the *Baboon*'s eye is softer and better defined and saving additionally 4.48 KB of information.



(a) JPEG2000 compressed at 0.59 bpp.



(b) JPEG2000-CIWaM compressed at 0.54 bpp. (c) JPEG2000-CIWaM compressed at 0.45 bpp.

Figure 4.5: Examples of reconstructed images of Baboon.

4.4 Perceptual Inverse Quantization

The proposed perceptual Quantization is a generalized method, which can be apply to wavelet-transform-based image compression algorithm such as EZW, SPIHT, SPECK or JPEG2000. But we introduce both forward and inverse perceptual quantization into the H*i*-SET coder. This process is shown in green by Fig. 4.6, which maintains the embedded features not only of H*i*-SET algorithm but also of any wavelet-based image coder. Thus, Perceptual Quantization + H*i*-SET = PH*i*-SET or Φ_{SET}.

Figure 4.6: The Φ_{SET} image compression algorithm.

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, 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.

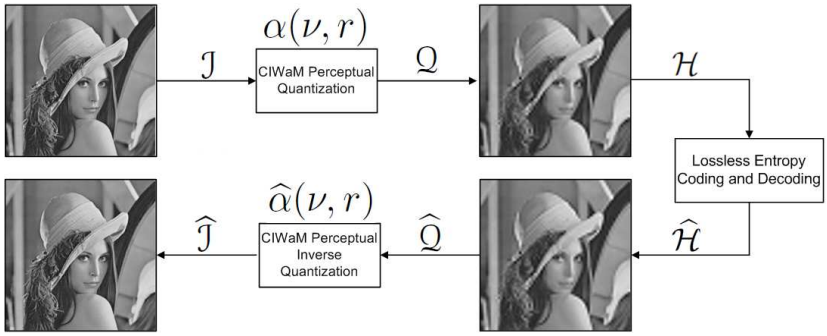
The main challenge underlies in to recover not only a good approximation of coefficients $\mathcal{Q}$ but also the visual weight $\alpha(\nu, r)$ that weighted them. A recovered approximation $\widehat{\mathcal{Q}}$ with a certain distortion Λ is decoded from the bitstream by the entropy decoding process.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

The reduction of the dynamic range is uniformly made by the perceptual quantizer, thus the statistical properties of $\mathcal{I}$ are maintained in $\widehat{\mathcal{Q}}$. Therefore our hypothesis is that to apply CIWaM to $\widehat{\mathcal{Q}}$, with the same viewing conditions applied to $\mathcal{I}$, could recover a decoded approach of the encoded visual weights, $\widehat{\alpha}(\nu, r)$. Thus, the perceptual inverse quantizer or the reconstructed $\widehat{\alpha}(\nu, r)$ introduces perceptual criteria to 3.7 and is given by:

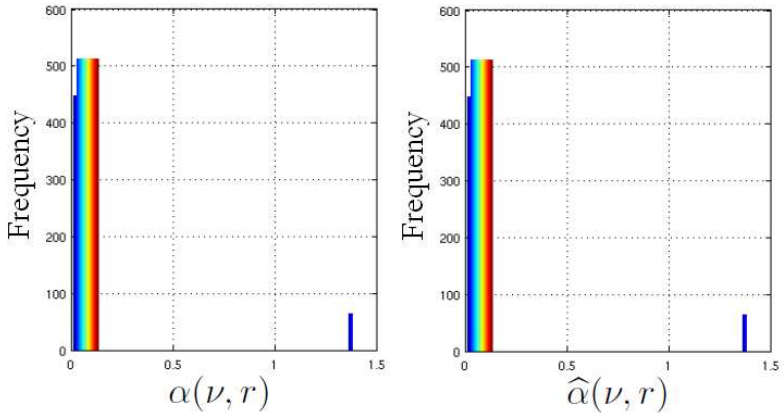
$$\widehat{\mathcal{I}} = \begin{cases} \displaystyle\sum_{s=1}^{n} \sum_{o=v,h,d} sign(\widehat{\omega_{s,o}}) \, \frac{\Delta_s \cdot (|\widehat{\omega_{s,o}}| + \delta)}{\widehat{\alpha}(\nu, r)} + (\widehat{c_n} + \delta) \cdot \Delta_n & |\widehat{\omega_{s,o}}| > 0 \\ 0, & \widehat{\omega_{s,o}}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

For the sake of demonstrating 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 in Figure 4.7. Figure 4.7a 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) Process



(b) Encoded VW (c) Decoded VW

Figure 4.7: (a) Process to find the histograms of Encoded (b) and Decoded (c) visual frequency weights for the 512×512 image *Lena*, channel Y at 10 meters.

Experiment 2: Correlation 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. Then, we measure the lineal correlation between $\alpha(\nu, r)$ and $\widehat{\alpha}(\nu, r)$. Table 4.2 shows that there is a hight 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.

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 of image quality after applying $\widehat{\alpha}(\nu, r)$ at $d = 2000$

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**

centimeters to the $\widehat{\mathcal{Q}}$ images. On average, this gain is about 13 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 before applying $\widehat{\alpha}(\nu, r)$.

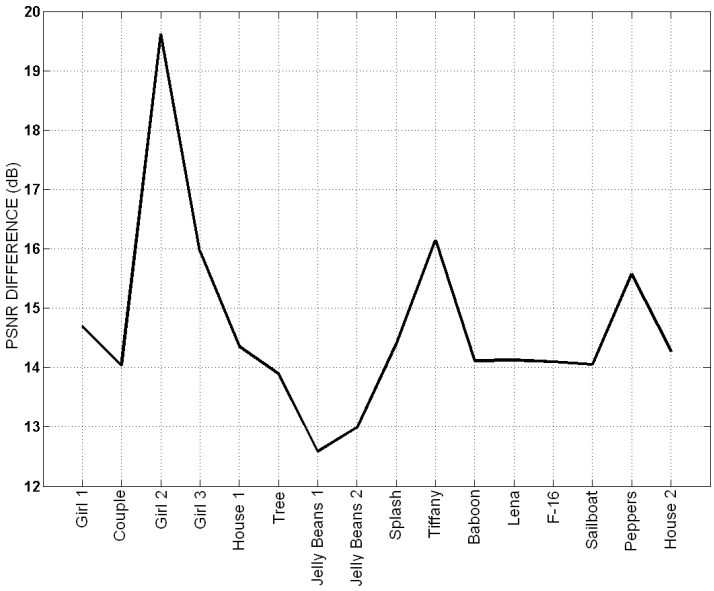
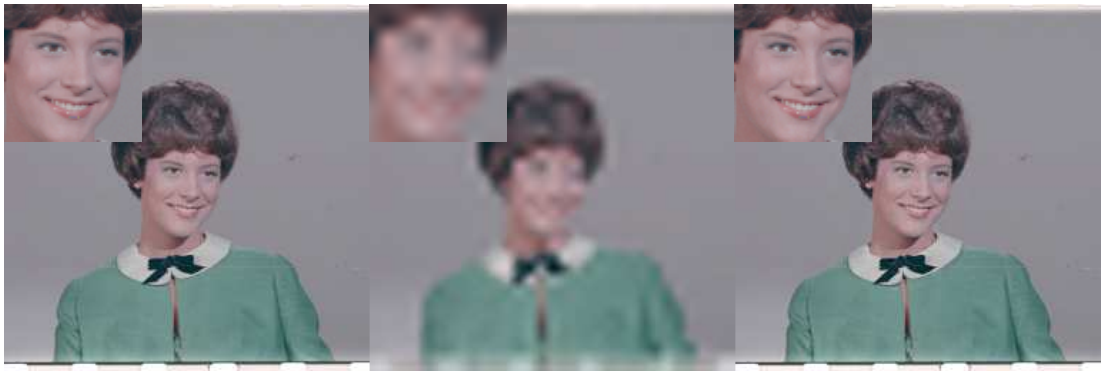
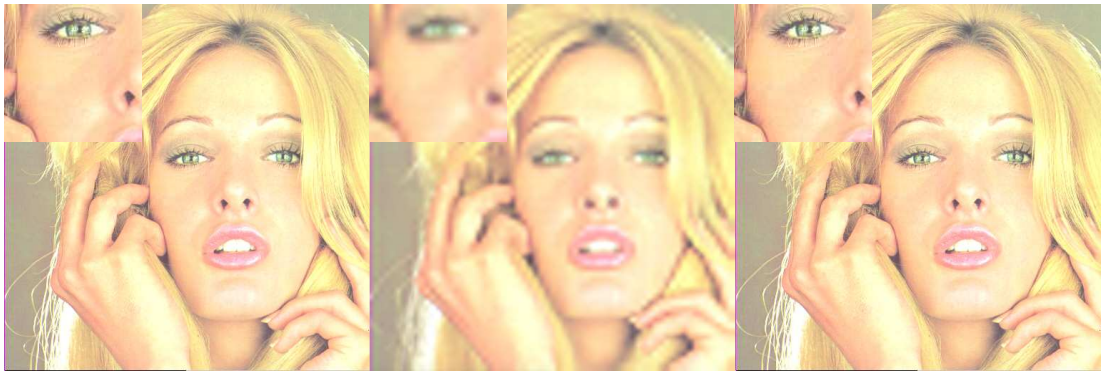


Figure 4.8: PSNR difference of each color image of the CMU database.

After applying $\widehat{\alpha}(\nu, r)$, these sixteen recovered images seem to have perceptually lossless quality. For the sake of prove with more results the latter affirmation, we perform this 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,



(a) Girl 2



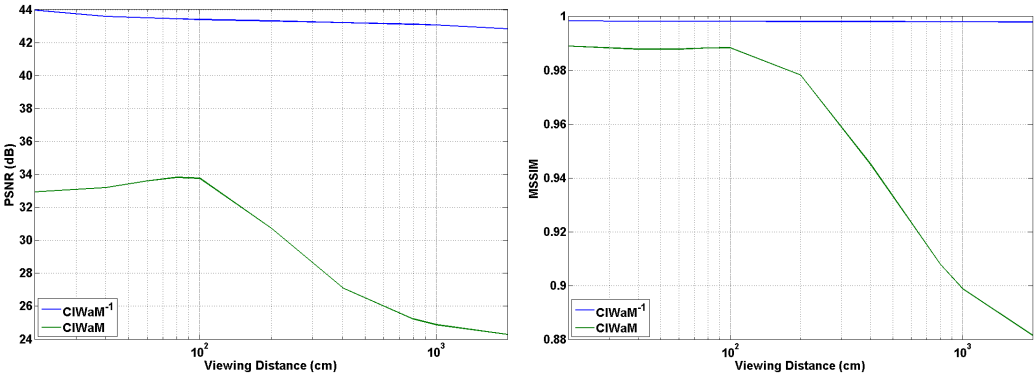
(b) Tiffany

(c) Peppers

Figure 4.9: Visual examples of Perceptual Quantization. Right images are the original images, central images are perceptual quantized images after applying $\alpha(\nu, r)$ at $d = 2000$ centimeters and left images are recovered images before applying $\widehat{\alpha}(\nu, r)$.

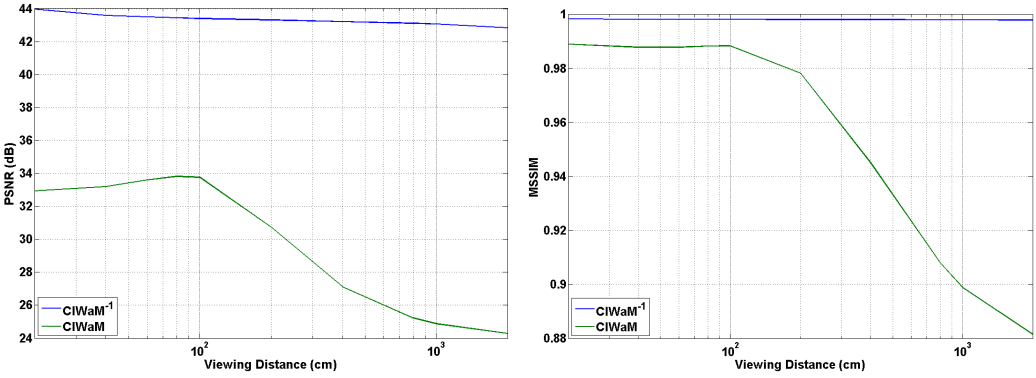
respectively.

In Figs. 4.10 and 4.11, green functions denoted as CIWaM are perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as CIWaM^{-1} are recovered images before 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 with respect to 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: Compression of Gray-scale Images (Y Channel) of the CMU image database. Green functions denoted as CIWaM are perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as CIWaM^{-1} are recovered images before applying $\widehat{\alpha}(\nu, r)$.



(a) PSNR (b) MSSIM

Figure 4.11: Compression of Color Images of the CMU image database. Green functions denoted as CIWaM are perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as CIWaM^{-1} are recovered images before applying $\widehat{\alpha}(\nu, r)$.

4.5 Φ_{SET} Codestream Syntax

Φ_{SET} Codestream Syntax is similar to the H*i*-SET one (Section 3.6), only two Markers are added inside Complemental Header (Fig. 3.9(b)), Perceptual Quantization Marker (PQ) and Observation Distance Marker (d).

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)).

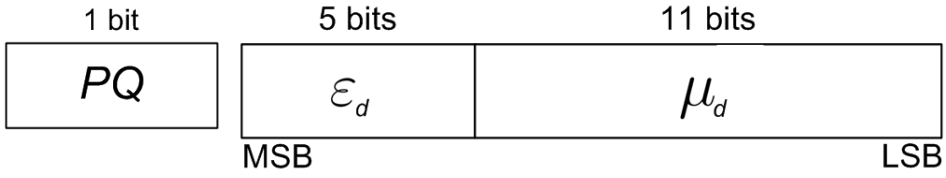
The eleven least significant bits are employed for the allocation of μ_d, which is defined as:

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

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, $R_{d\max} = 11$.



(a) PQ Marker (b) d Marker

Figure 4.12: Markers added to Complemental Header (Fig. 3.9(b)). (a) Perceptual Quantization Marker and (b) Structure of Observation Distance Marker

4.6 Experiments and Results

For the sake of comparing the performance between the JPEG2000(50) and Φ_{SET} coders, both algorithms are tested according to the process depicted in Fig. 4.13. Thus, first a Φ_{SET} compression with certain viewing conditions is performed, which gives a certain bit-rate (bpp). Then, a JPEG2000 compression is developed with the same bit-rate. Ones both algorithms recover their distorted images, they are compared with some numerical image quality estimators such as: MSE(18), PSNR(18), SSIM(44), MSSIM(53), VSNR(12), VIF(57), VIFP(44), UQI(54), IFC(46), NQM(14), WSNR(25), SNR and $\mathcal{C}_w$PSNR(Section 2.3.1).

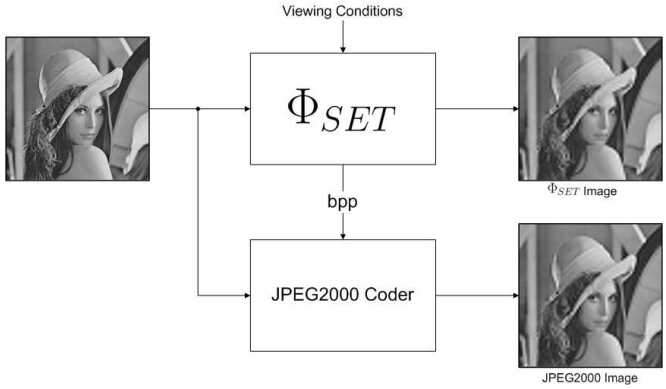
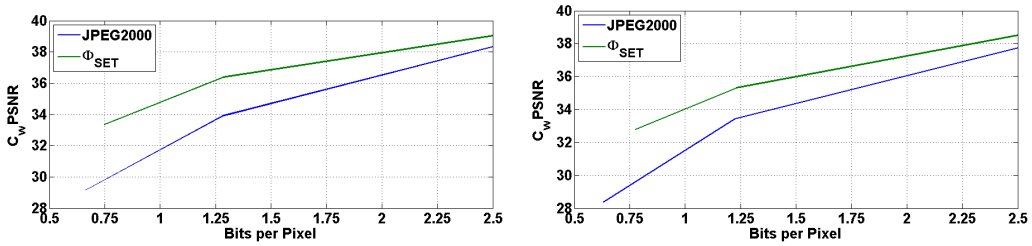


Figure 4.13: Process for comparing JPEG2000 and Φ_{SET}.



(a) CMU Image Database (b) IVC Image Database

Figure 4.14: Comparison between Φ_{SET} and JPEG2000 image coders. Compression rate vs perceptual image quality, performed by $\mathcal{C}_w$PSNR, of the CMU (a) and IVC (b) image databases.

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 exhibited in Sections C.2.1 and C.2.2 for the CMU and IVC Image Databases, respectively.

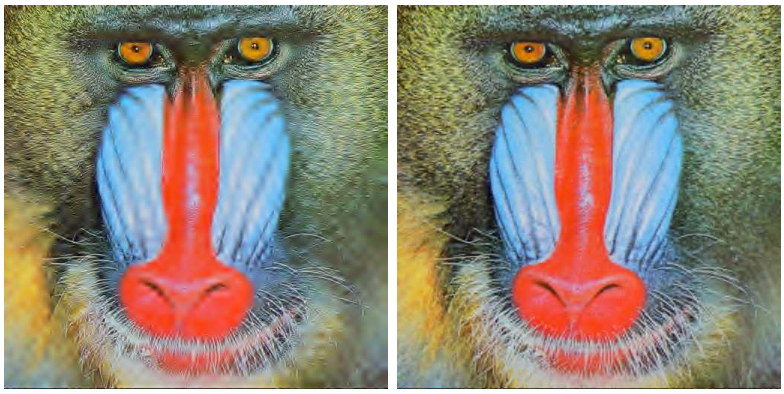
(a) JPEG2000, $\mathcal{C}_w$PSNR=34.40 dB (b) Φ_{SET}, $\mathcal{C}_w$PSNR=37.81 dB

(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 4.15: Example of reconstructed color images *Lenna*, *Girl2* and *Tiffany* of the CMU image database compressed at 0.92 bpp(a-b), 0.54 bpp(c-d) and 0.93 bpp(e-f), 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 4.16: Example of reconstructed color images *Barbara*, *Mandrill* and *Clown* of the IVC image database compressed at 0.76 bpp (a-b), 1.15 bpp (c-d) and 0.96 bpp (e-f), respectively.

The parameters for estimating the $\mathcal{C}_w$PSNR assessment are: $d = 8H$, $M_{size} = 19''$,

$h_{res} = 1280$ and $v_{res} = 1280$.

Fig. 4.14(a) shows the perceptual quality, performed by $\mathcal{C}_w$PSNR, of the recovered color images both 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(49). On the average, a color image with 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 4.15 we can see this differences when the 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.

Fig. 4.14(b) shows the perceptual quality, performed by $\mathcal{C}_w$PSNR, of the recovered color images both 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 34.70 dB of perceptual image quality and by Φ_{SET} (green function) has 36.86 dB. In Figure 4.16 we can see this differences when the 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.

Chapter 5

Perceptual Generalized Bitplane-by-Bitplane Shift

5.1 Introduction

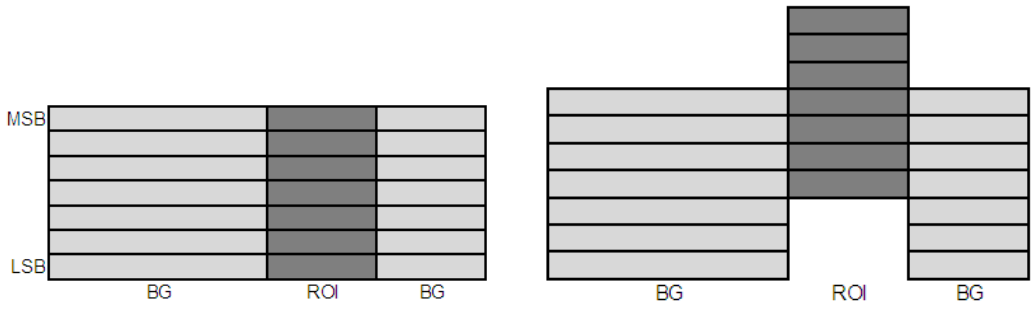
Region of interest (ROI) image coding is a feature that modern image coder possesses, which allows to encode an specific region in an image with better quality than the rest or background (BG). ROI coding is one of the requirements in the JPEG2000 image coding standard (10, 11, 47, 50) and defines two methods(4, 13, 30, 31, 50):

1. Based on general scaling
2. Maximum shift (MaxShift)

The general ROI scaling-based method scales (shift) coefficients in such a way that the bits associated with the ROI are placed in higher bit-planes than the bits associated with the background as shown in Figure 5.1(b). Thus, 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, the ROI is decoded and refined, 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:

1. Wavelet transform is calculated.



(a) No ROI coding

(b) Scaling Based Method, $\varphi = 3$

Figure 5.1: Scaling based ROI coding method. Background is denoted as BG and Region of Interest as ROI.

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

3. The wavelet coefficients are quantized. Then, the quantized coefficients are stored in a sign magnitude representation, using the most significant part of the precision, this 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 encode.

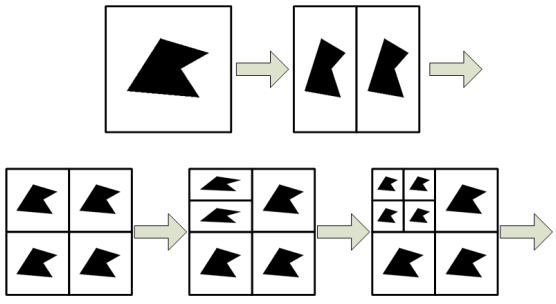


Figure 5.2: ROI mask generation, wavelet domain.

ROI scaling-based method needs as input the scaling value φ, while MaxShift method calculates it. Hence, the encoder defines from quantized coefficients this scaling value such that:

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

where $\max\{\mathcal{M}_{\mathcal{BG}}\}$ 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. 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 will be explained in section 5.2.2.

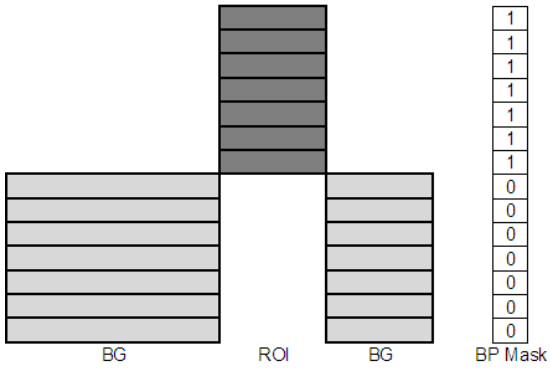


Figure 5.3: MaxShift method, $\varphi = 7$. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP mask.

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 (59). BbBShift shifts bitplanes on a bitplane-bybitplane 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.

During the encoding process:

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

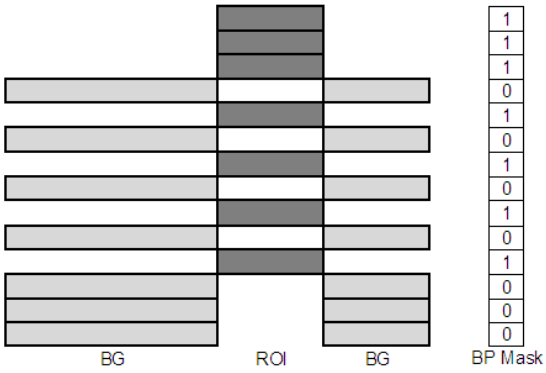


Figure 5.4: BbBShift ROI coding method, $\varphi_1 = 3$ and $\varphi_2 = 4$. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP mask.

- 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 $\varphi_1 + 2bpl - 1$
- Else 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.

5.2.2 GBbBShift

In practice, 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 modified BbBShift method and proposed the generalized bitplane-by-bitplane shift (GBbBShift) method in (55), which introduces the option to improve visual quality either of ROI or BG or both. Figure 5.5 shows that with GBbBShift method is posible to decode some bitplanes of BG, improving 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.

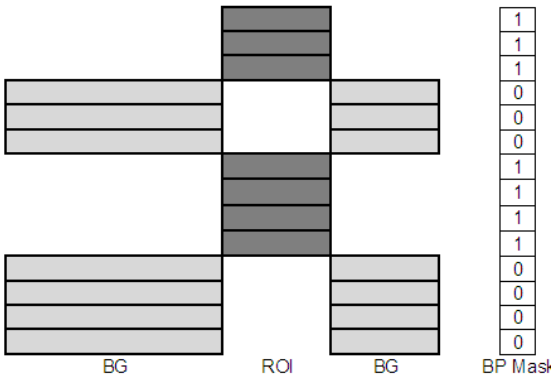


Figure 5.5: GBbBShift ROI coding method. Background is denoted as BG, Region of Interest as ROI and Bitplane mask as BP mask.

Therefore, GBbBShift gives 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 than original image possesses. A ROI bitplane is represented by 1, while a BG bitplane by 0. For instance, 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 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, should be transmitted.

5.3 ρGBbBShift Method

In order to have several kinds of options for bitplane scaling technics, a perceptual generalized bitplane-by-bitplane shift(ρGBbBShift) method is proposed. ρGBbBShift 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.

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:

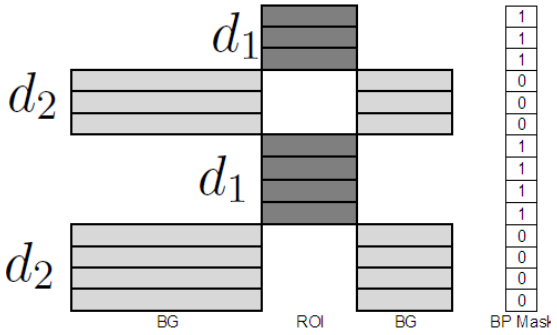


Figure 5.6: ρGBbBShift ROI coding method. Background is denoted as BG (Quantized at d_2), Region of Interest as ROI (Quantized at d_1)and Bitplane mask as BP mask.

1. Calculate φ using Equation 5.1.
2. Verify that the length of BP_{mask} is equal to 2φ.
3. - For all ROI Coefficients, quantize them using Equation **??** with d_1 as viewing distance.
   - For all BG Coefficients, quantize them using Equation **??** with d_2 as viewing distance, 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 τ bitplanes all ROI perceptual quantized coefficients of the $(\varphi - \eta)$-th bitplane and increment η.
   - Else: Shift up η bitplanes all BG perceptual quantized coefficients of the $(\varphi - \tau)$-th bitplane and increment τ.

At the decoder, shifting scheme is as follows:

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

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 τ bitplanes all perceptual quantized coefficients, which pertain to the $(2\varphi - (\tau + \eta))$-th bitplane of the recovered image and increment η.
   - Else: Shift down η bitplanes all perceptual quantized coefficients, 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 $\overline{c}_{i,j}$ as a shifted down c obtained in the previous step, with φ bitplanes.

   - If $(c_{i,j}$ & $BP_{mask}) > 0$, Dequantize $\overline{c}_{i,j}$ using Equation 4.2 with d_1 as viewing distance.
   - Else, Dequantize $\overline{c}_{i,j}$ using Equation 4.2 with d_2 as viewing distance.

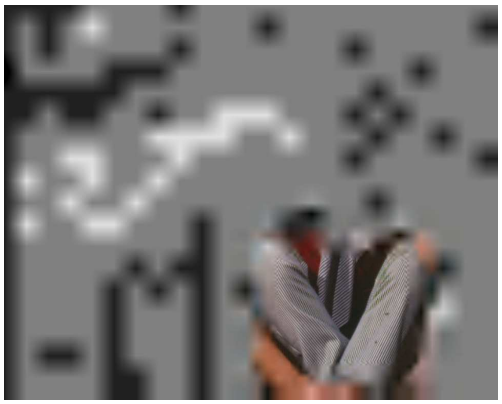
5.4 Experimental Results

ρGBbBShift method, as the algorithms presented here, can be applied to many image algorithms such as JPEG2000 or H*i*-SET. We test our method applied to H*i*-SET and the results are contrasted with MaxShift method in JPEG2000. The parameter setups 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. Also, we use the Kakadu implementation when an image is compressed by JPEG2000 standard(49).

5.4.1 Experiments

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

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). But the BG is perceptually better when ρGBbBShift is used.



(a) MaxShift in JPEG2000 coder, 0.5 bpp

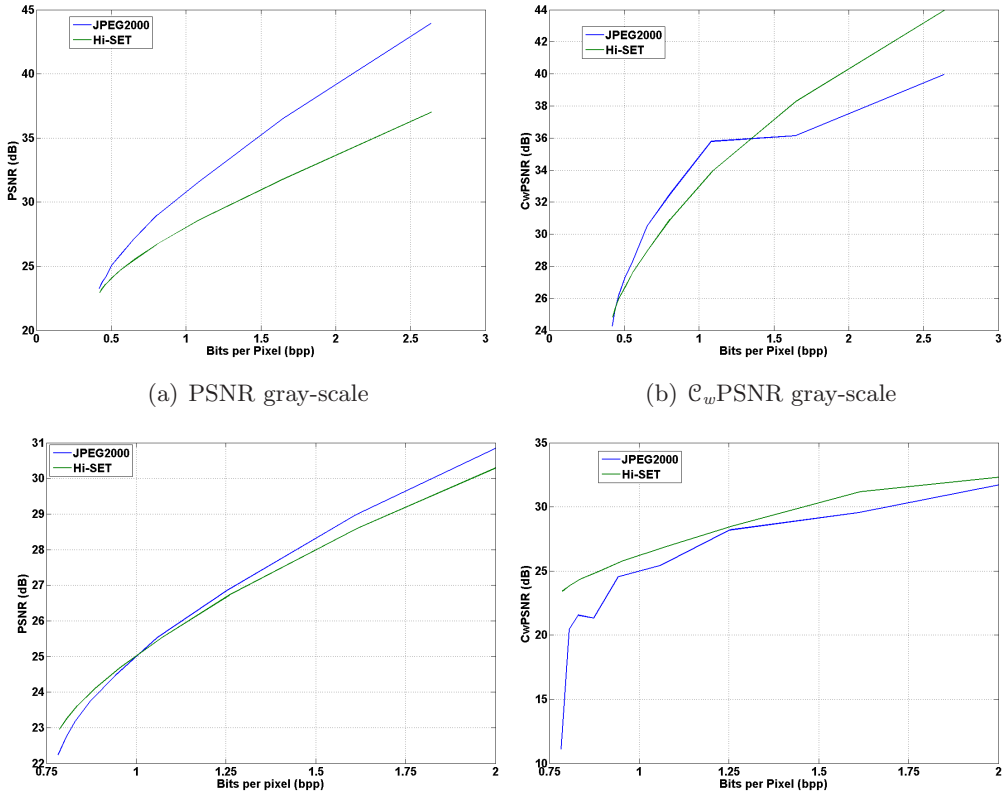
(b) GBbBShift in JPEG2000 coder, 0.5 bpp

(c) ρGBbBShift in H*i*-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 H*i*-SET coder ((c)$BP_{mask} = 1111000110110000$).

In order to better qualify the performance of MaxShift and ρGBbBShiftmethods, we compress two different gray-scale images *1600*, from CSIQ image database (Fig A.4), and *Lenna* at different bit-rates, from 0.5 to 2 bpp. ROI area is a patch at the center of these images, whose size is 1/16 of the image.

In the whole image quality assessment of image *1600*, JPEG2000 obtains better objective quality both for gray-scale and color images (Figures 5.8(a) and 5.8(c), respectively). But when the subjective quality is estimated ρGBbBShift coded images are perceptually better.



(a) PSNR gray-scale

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

(c) PSNR color

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

Figure 5.8: Comparison between MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to H*i*-SET coder. 512×512 pixel Image *1600* with 8 (a-b) and 24 (c-d) bits per pixel are employ 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 objectively (a and c) and subjectively (b and d).

A visual example is depicted by Figure 5.9, 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 H*i*-SET coder.

Similarly to the results obtained with the image *1600*, when a ROI area is defined in Image *Lenna*, ρGBbBShift obtains less objective quality (Figures 5.10(a) and 5.10(c)), but better subjective quality both for gray-scale and color images (Figures 5.10(b) and 5.10(d), respectively).

Figure 5.11 shows a visual example, when image *Lenna* is compressed at 0.34 bpp

(a) MaxShift method in JPEG2000 coder, 0.42 bpp

(b) ρGBbBShift method in Hi-SET coder, 0.42 bpp

Figure 5.9: 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 MaxShift method in JPEG2000 coder((a) $\varphi = 8$) and ρGBbBShift method in Hi-SET coder ((b)$BP_{mask} = 1111000110110000$).

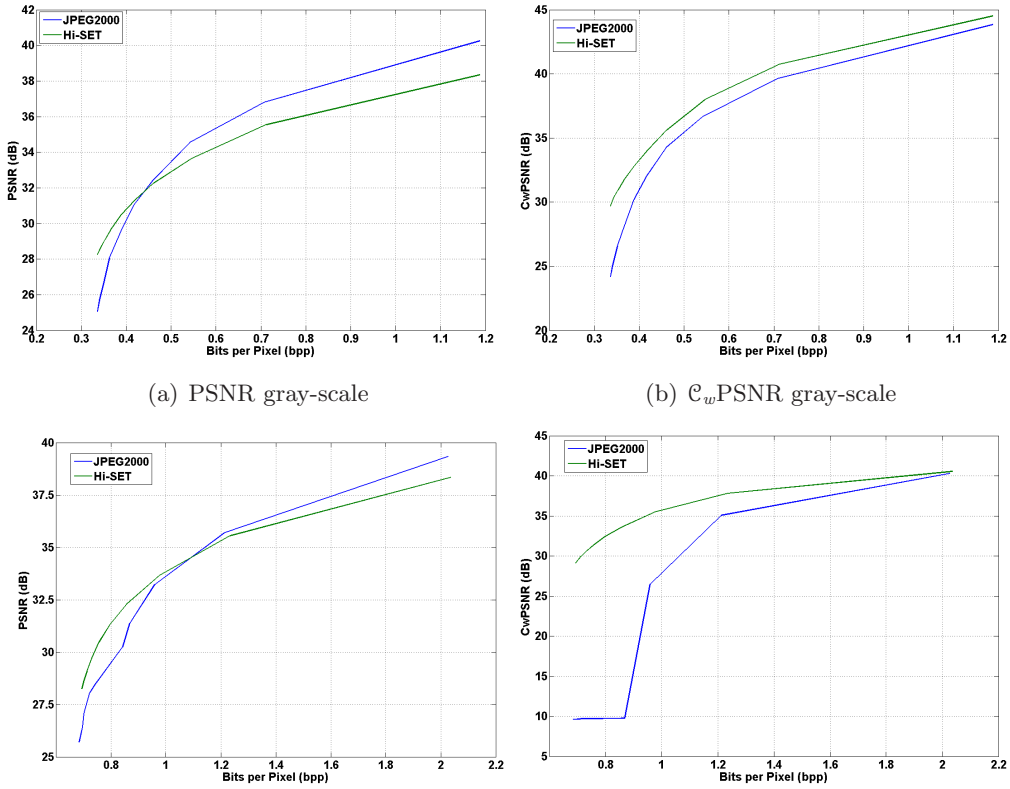
by JPEG2000 and Hi-SET. Thus, it can be observed that ρGBbBShift provides an important perceptual difference regarding the MaxShift method.

5.4.2 Application in other image compression fields

The usage of ROI coded images depend 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)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 located this ROI_ϕ, for the sake of better interpreting a given ROI coded image. In addition, according Federal laws in some countries, ROI areas must be without losses(61). ρGBbBShiftis able to accomplish these two features.

Figure 5.12 shows an example of medical application. A rectangular ROI of the Im-



(a) PSNR gray-scale

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

(c) PSNR color

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

Figure 5.10: Comparison between MaxShift method applied to JPEG2000 coder and ρGBbBShift applied to H*i*-SET coder. 512×512 pixel Image *Lenna* with 8 (a-b) and 24 (c-d) bits per pixel are employ 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 objectively (a and c) and subjectively (b and d).

age *mdb202* from PEIPA image database(37) , coordinates [120 440 376 696], is coded at 0.12 bpp by JPEG2000 and H*i*-SET, employing MaxShift and ρGBbBShiftmethods, respectively. The overall image quality measured by PSNR in Figure 5.12(a) is 37.21 dB, while in Figure 5.12(c) is 36.76 dB. Again, PSNR does not reflect perceptual differences between images (Figures 5.12(b) and 5.12(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.

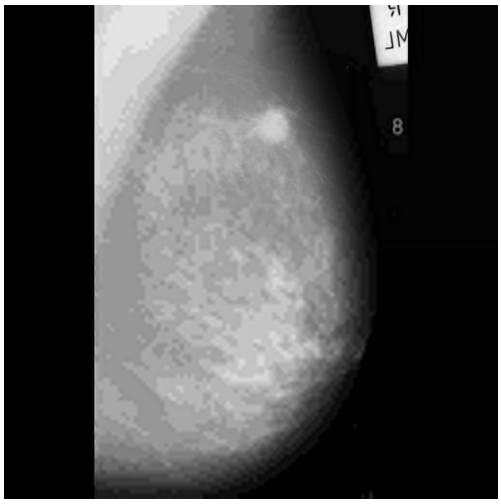
(a) MaxShift method in JPEG2000 coder, 0.34 bpp

(b) ρGBbBShift method in H*i*-SET coder, 0.34 bpp

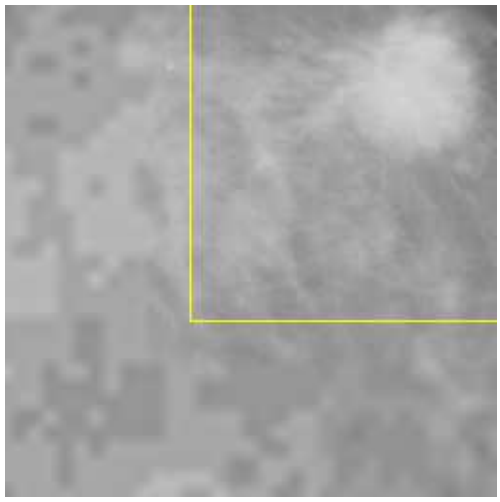
Figure 5.11: 512×512 pixel Image *Lenna* from CMU image database with 8 bits per pixel (a). ROI is a patch at the center of the image, whose size is 1/16 of the image. Decoded images at 0.34 bpp using MaxShift method in JPEG2000 coder((b) $\varphi = 8$) and ρGBbBShift method in H*i*-SET coder ((c)$BP_{mask} = 1111000110110000$).

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 in advance, in order to transfer and store them. However, only a small part of the image is useful and therefore some regions are sketched(62).

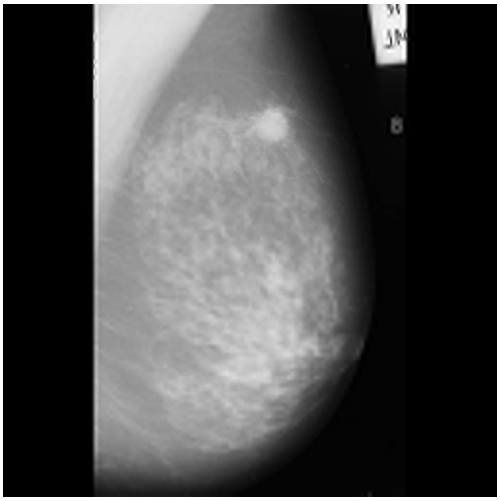
Figure 5.13 shows an example of the application of ROI in Remote Sensing. Image *2.1.05*, from *Volumen 2: aerials* of USC-SIPI image database 8 bits per pixel(2), is compressed at 0.42 bpp. MaxShift method spends all the bit-ratio for coding ROI, located at [159 260 384 460], while ρGBbBShift balances a perceptually lossless ROI area with an acceptable representation of the BG. Hence, the overall image quality measured by PSNR in Figure 5.13(a) is 16.06 dB, while in Figure 5.13(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.



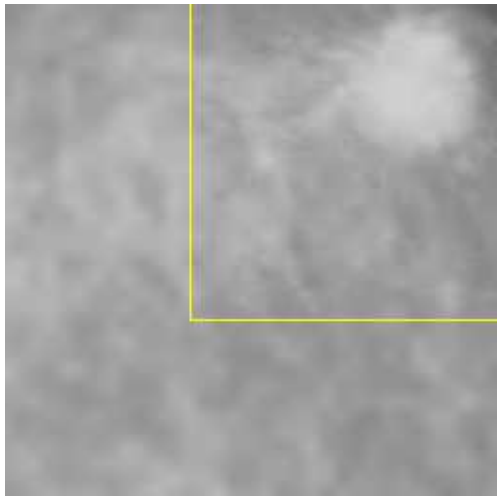
(a) MaxShift method in JPEG2000 coder, 0.12 bpp



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

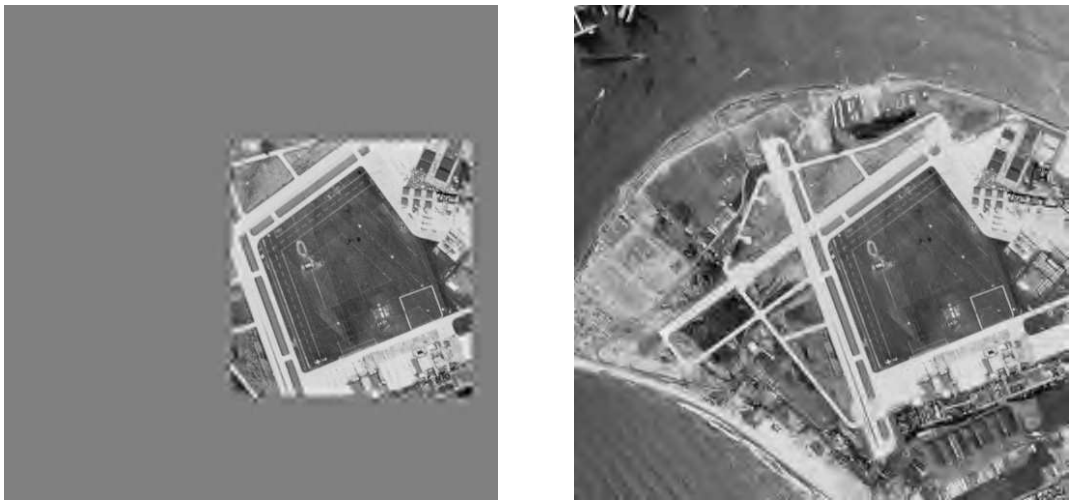


(c) ρGBbBShift method in Hi-SET coder, 0.12 bpp



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

Figure 5.12: Example a medial 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 in JPEG2000 coder((a-b) $\varphi = 8$) and ρGBbBShift method in Hi-SET coder ((c-d)$BP_{mask} = 1111000110110000$).



(a) MaxShift in JPEG2000 coder, 0.42 bpp

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

Figure 5.13: Example a remote sensing application. 512×512 pixel Image *2.1.05* from *Volumen 2: aerials* of USC-SIPI image database 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 in JPEG2000 coder((a) $\varphi = 8$) and ρGBbBShift method in H*i*-SET coder ((b)$BP_{mask} = 1111000110110000$).

Chapter 6

Conclusions and Future work

The main goal of this thesis was in one hand to identify and to remove non-perceptual information of an image, maintaining as far as possible, the same entropy as the source image and the other hand to introduce these perceptual criteria into a proposed image compression system. Hence, the main contributions this work are a proposal of a perceptual image compression system and a image quality assessment.

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. The $\mathcal{C}_w$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. $(\varepsilon\mathcal{R}(n\mathcal{P}))$. 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 can not 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 any kind of distortion is performed by measuring the objective numerical quality (i.e. the PSNR) of the perceptual images predicted by CIWaM at D cm.

$\mathcal{C}_w$PSNR was tested in four well-known image databases such as TID2008, LIVE, CSIQ and IVC. It is the best-ranked image quality method for JPEG and JPEG2000

distortions when compared to several existing 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 increases the correlation of PSNR with perceived image quality. Since when CIWaM weights PSNR, correlation of predicting subjective ratings either of PSNR or MSE improves the results by 14% and 11.5% for the same kind of distortions, on the average.

H*i*-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 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. Table 6.1 summarize the average improvements when compressing the TID2008 Image Database.

Table 6.1: Average improvement of H*i*-SET in front of JPEG2000

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

H*i*-SET coder improves the image quality of the JPEG2000 coder around 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 H*i*-SET improved the results not only objectively but also metrics like MSSIM, UQI or VIF, which are perceptual indicators.

In Chapter 4, we proposed the incorporation both of forward and inverse quantizer to H*i*-SET using CIWaM, 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 Forward Perceptual Quantization improves the JPEG2000 compression and image perceptual quality and impacts, on the average, with about 20 percent. In addition, both Forward and Inverse Quantization were applied to H*i*-SET, which significatively improved the results regarding the JPEG2000 compression.

ρGBbBShift, described in Chapter 5, is generalized method, which can be applied to any wavelet-base compressor. We introduced ρGBbBShift method to the H*i*-SET coder. Thus, our proposal visually improved the results obtained by previous method like BbBShift and GBbBShift. In our experiments showed that ρGBbBShift provides an important perceptual difference regarding the MaxShift method, when it is applied not only to conventional images like *Lenna* 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 tool of assessing the image quality. This indicator can be considered as a set of three gauges, which can be used for different purposes.

- Demonstration that CIWaM operates correctly on natural images and inside the field of image compression.

- 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 store the applied weighting for inverse quantizing, this is because CIWaM properties permits to predict perceptual weighting *a posteriori*.

- Development of a image coder, which is a serious alternative of JPEG2000 exploiting the recursion of fractal, avoiding the massive storage of pixel coordinates.
- Proposal of a new method for coding of Regio of Interest areas, which can be applied to any wavelet based compression scheme.

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 apply 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(43), SPIHT(41), JPEG2000 (47) and H*i*-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 H*i*-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).

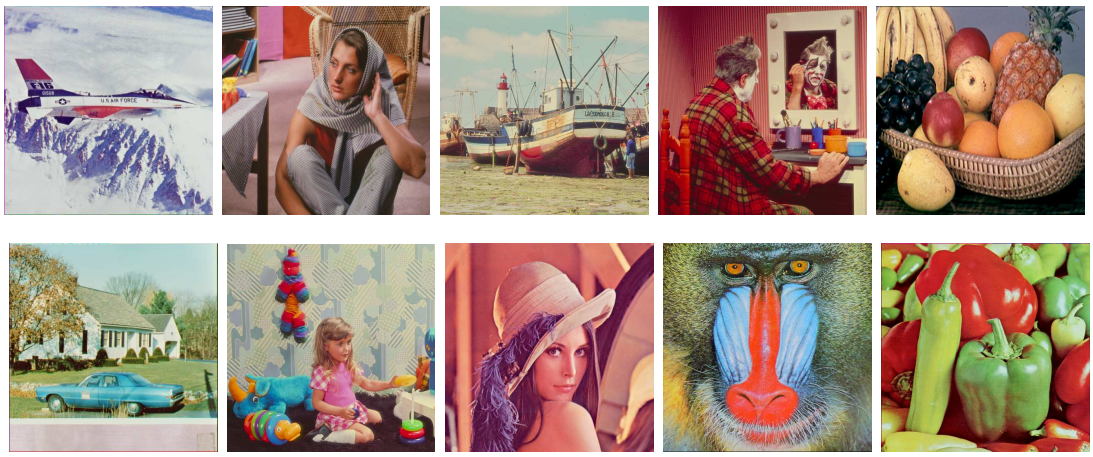


Figure A.1: Tested 512 × 512 pixel 24-bit color images, belonging to the IVC Image database.

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).

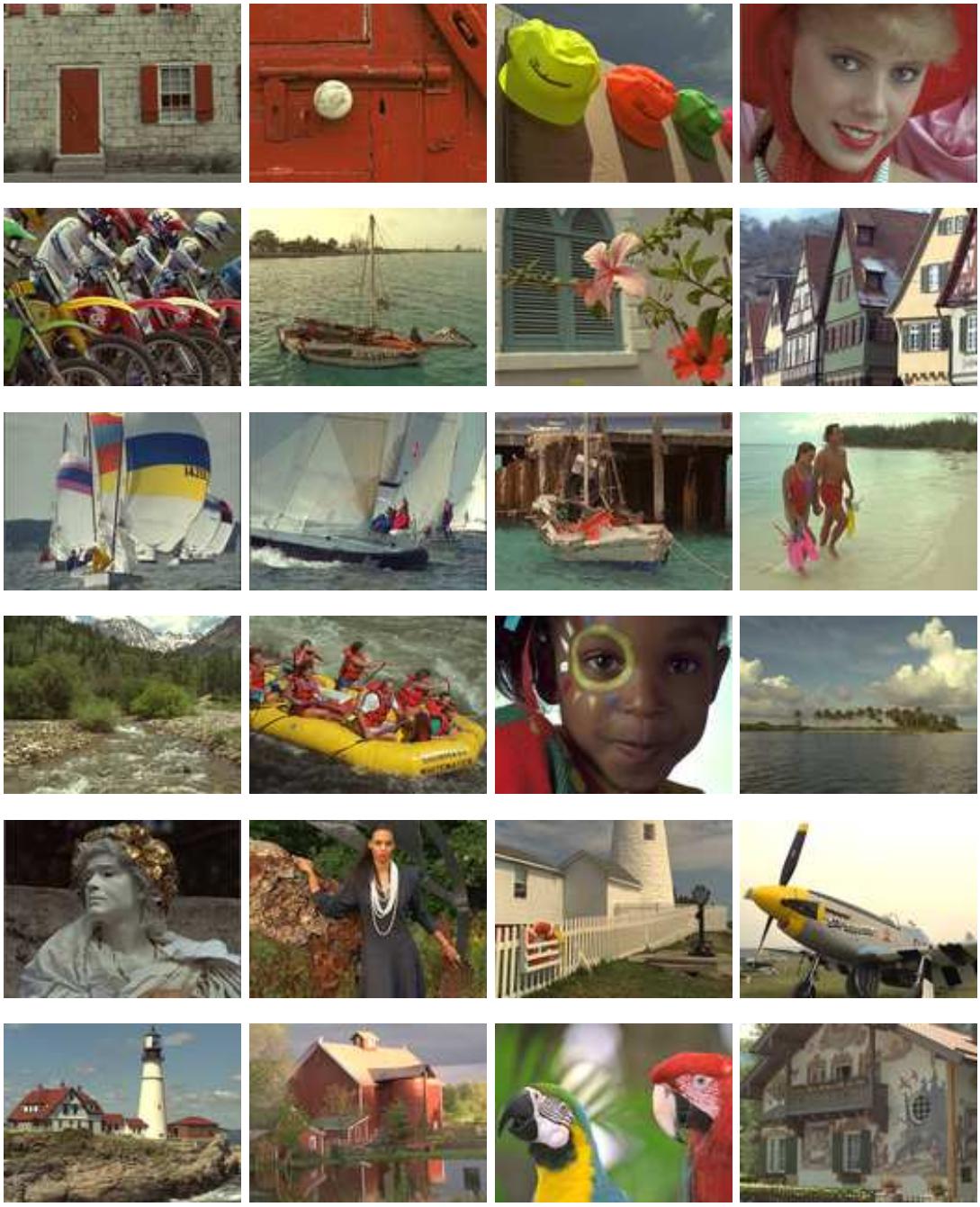


Figure A.2: Tested 512 × 384 pixel 24-bit color images, belonging to the Tampere test set.

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(45).

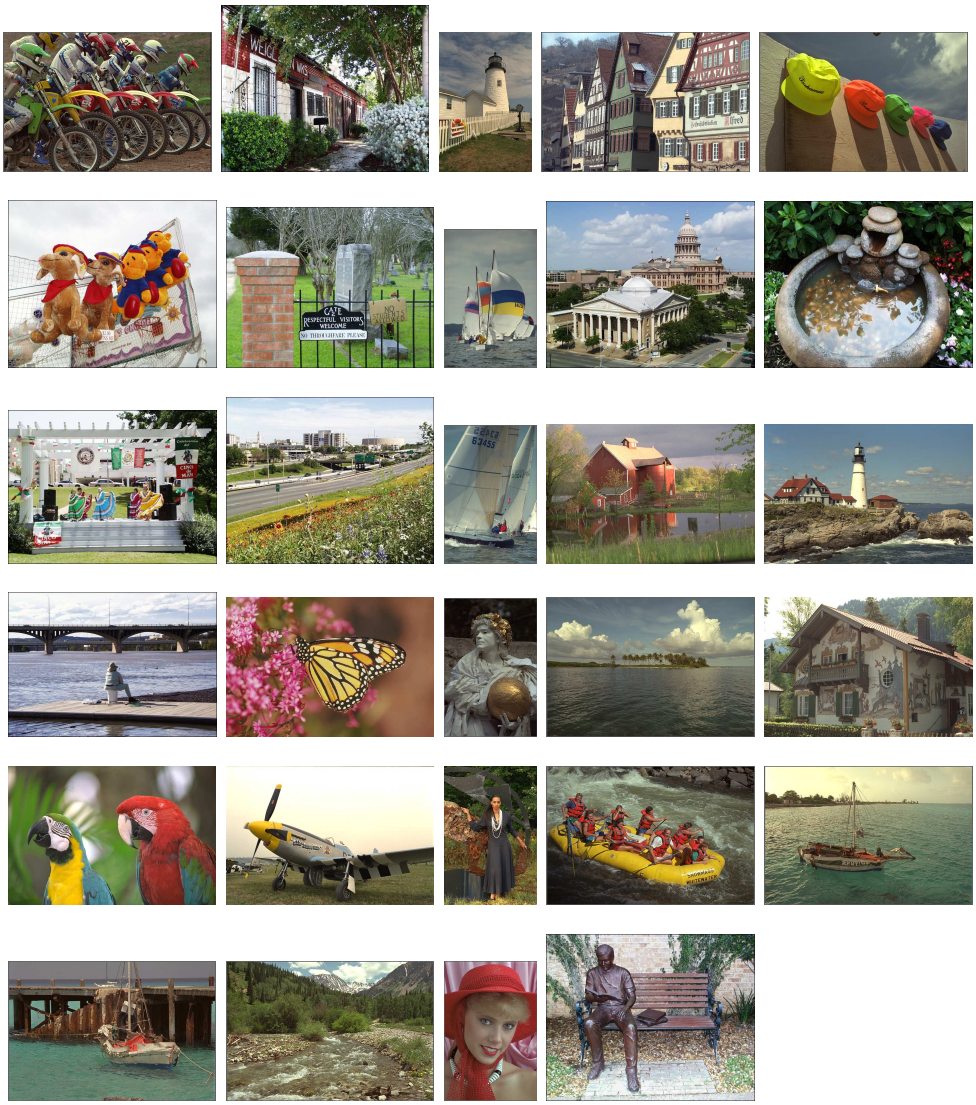


Figure A.3: Set of 29 tested images of 24-bit color, belonging to the LIVE Image database.

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).

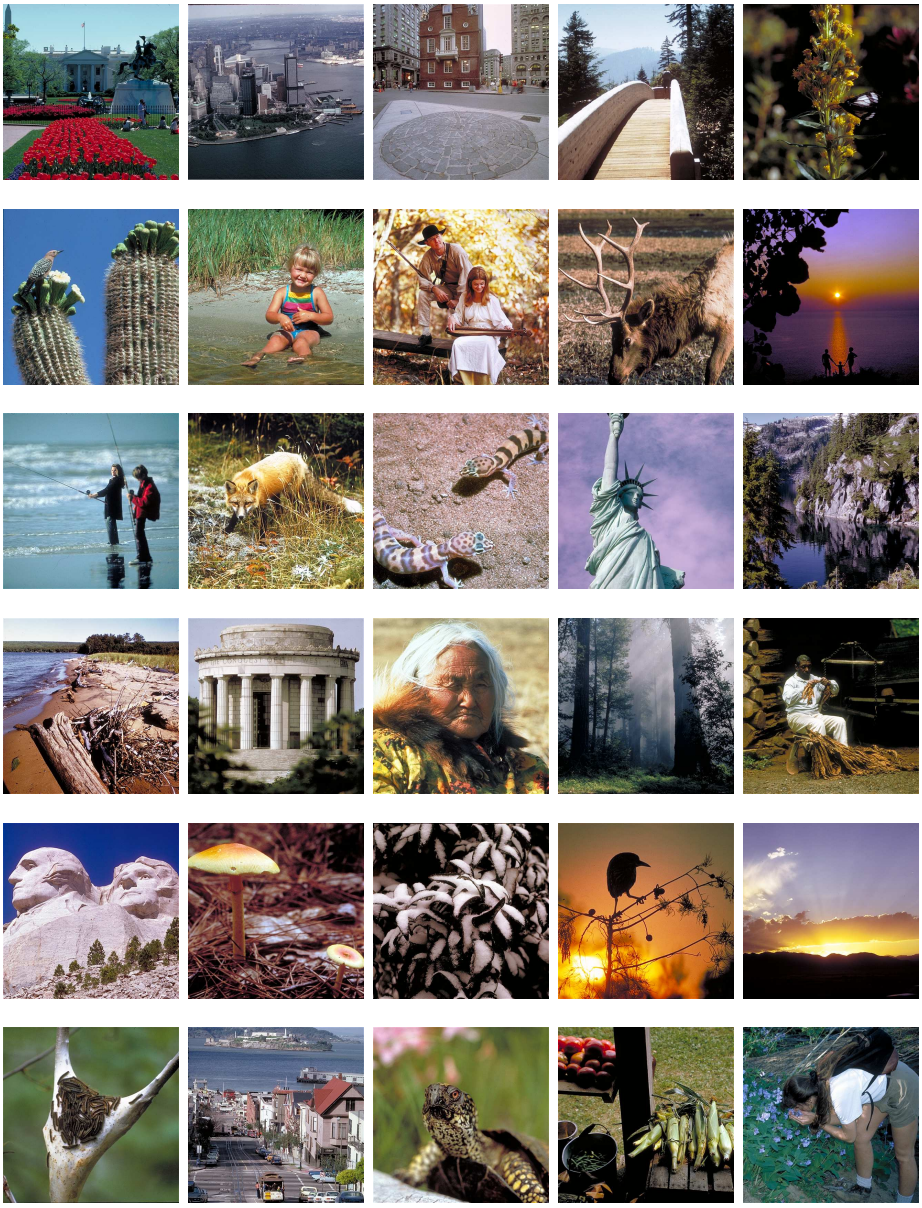


Figure A.4: Tested 512 × 512 pixel 24-bit color images, belonging to the CSIQ Image database.

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).

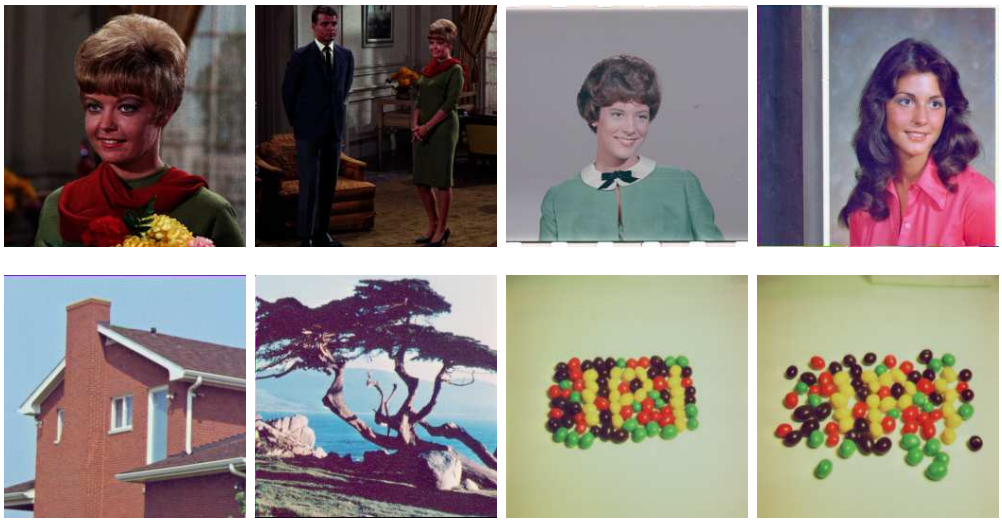


Figure A.5: Tested 256×256 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.

Figure A.6: Tested 512×512 pixel 24-bit Color Images, obtained from the University of Southern California Image Data Base.

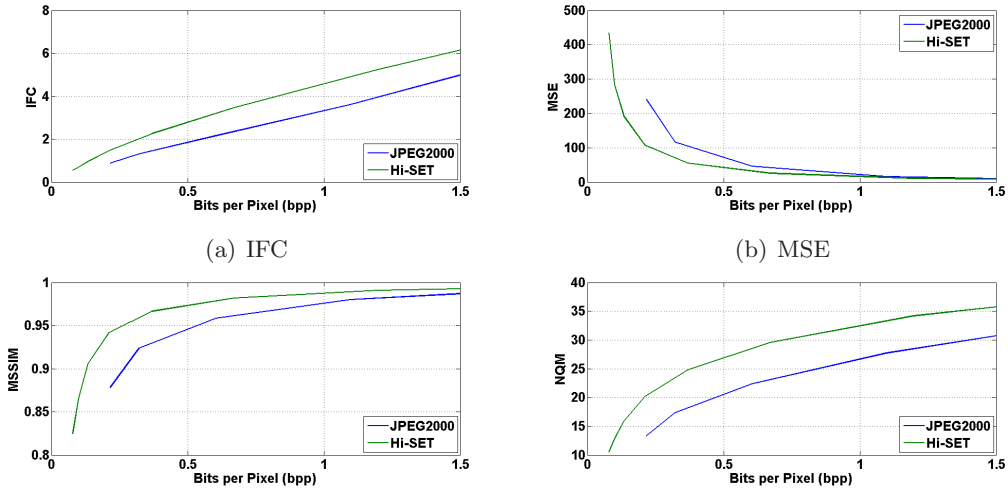
Appendix B

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.

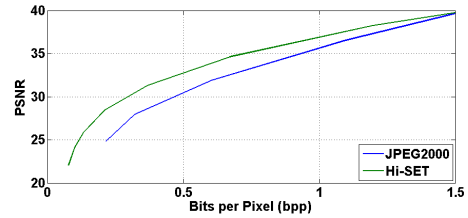


(a) IFC (b) MSE

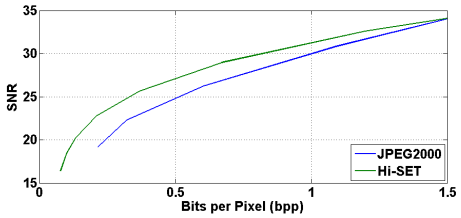
(c) MSSIM (d) NQM

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

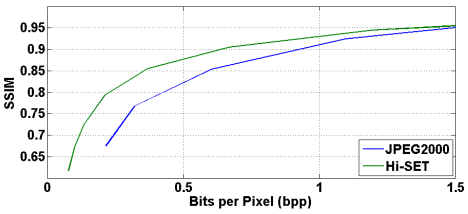
B. JPEG2000 VS H*i*-SET: COMPLEMENTARY RESULTS OF CHAPTER 3



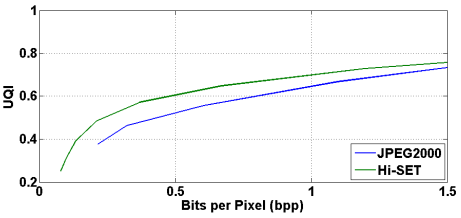
(a) PSNR



(b) SNR

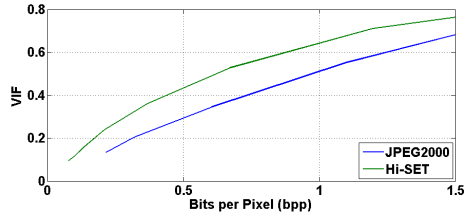


(c) SSIM

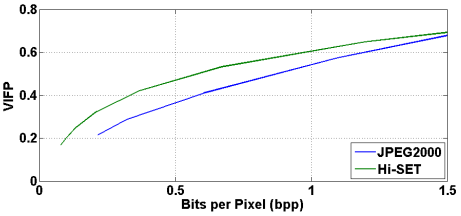


(d) UQI

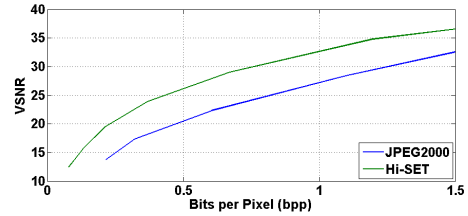
Figure B.2: Gray-Scale CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.



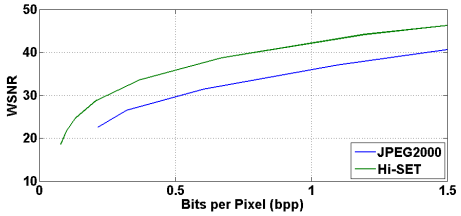
(a) VIF



(b) VIFP



(c) VSNR

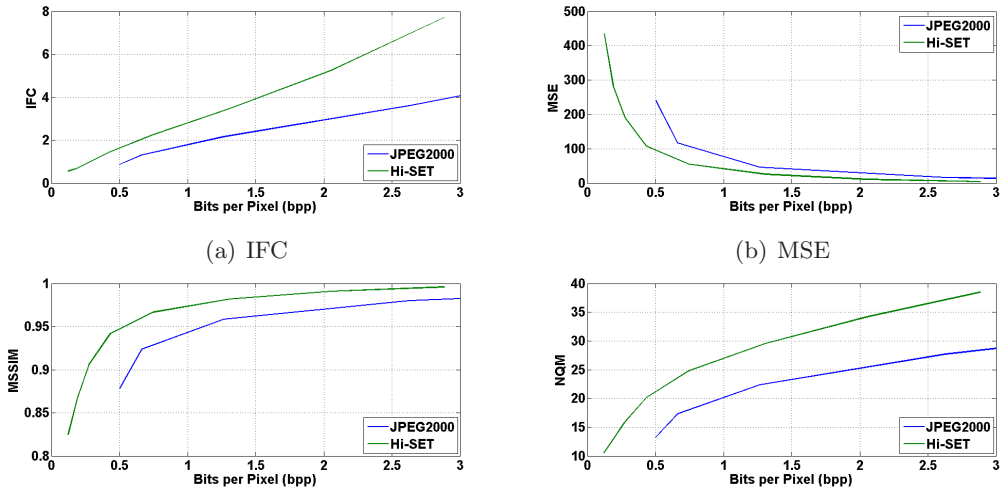


(d) WSNR

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

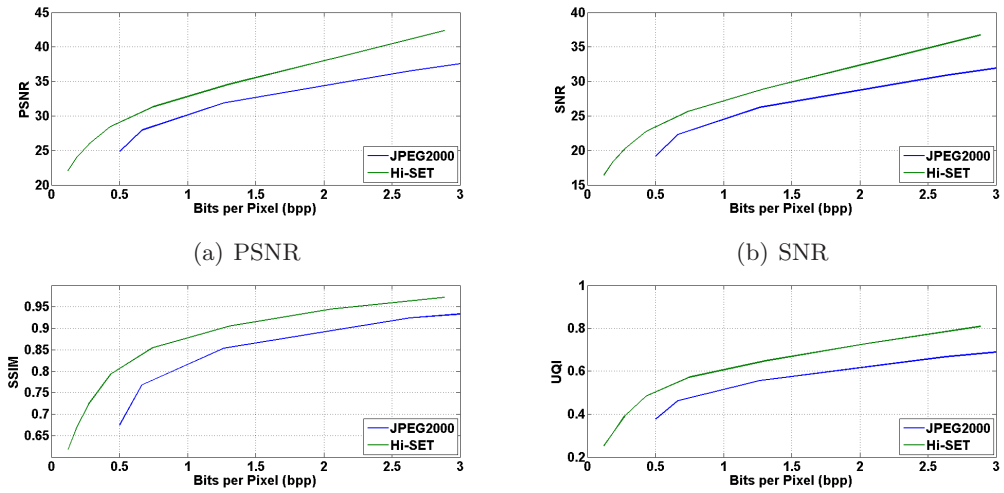
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

(c) MSSIM (d) NQM

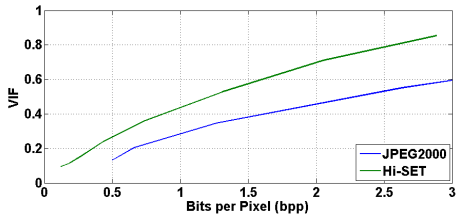
Figure B.4: Color CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



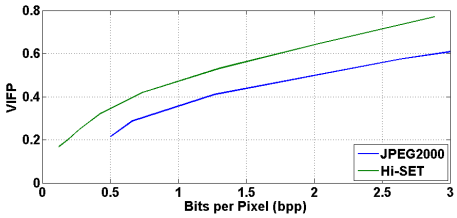
(a) PSNR (b) SNR

(c) SSIM (d) UQI

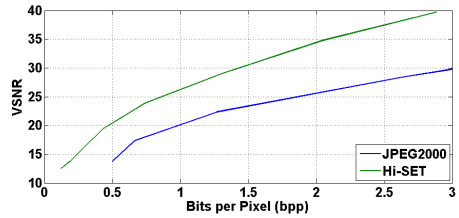
Figure B.5: Color CMU Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.



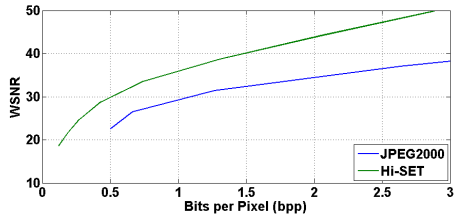
(a) VIF



(b) VIFP



(c) VSNR



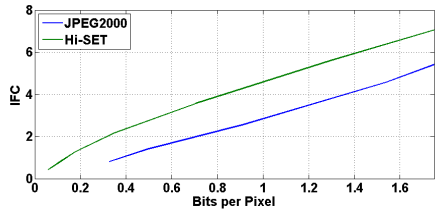
(d) WSNR

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

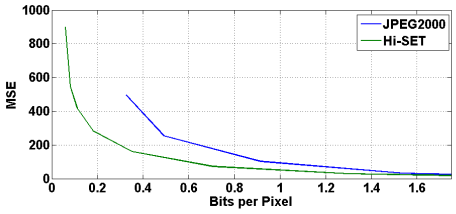
B.2 Categorical Subjective Image Quality Image Database

B.2.1 Gray-Scale (Y Channel)

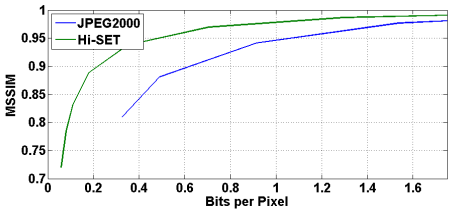
Compression of Gray-Scale Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



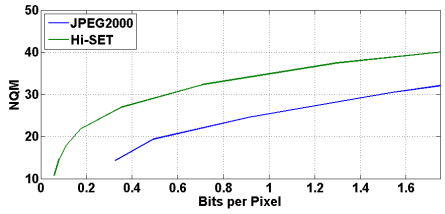
(a) IFC



(b) MSE

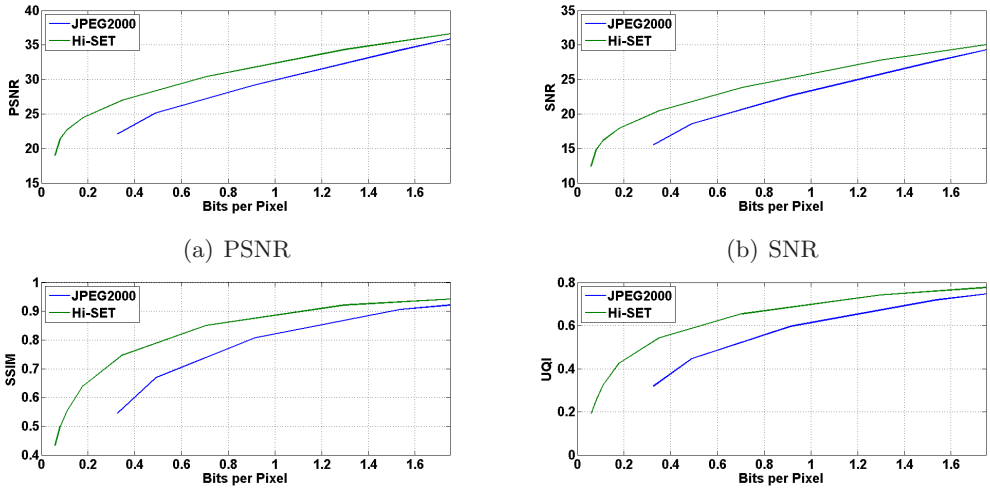


(c) MSSIM



(d) NQM

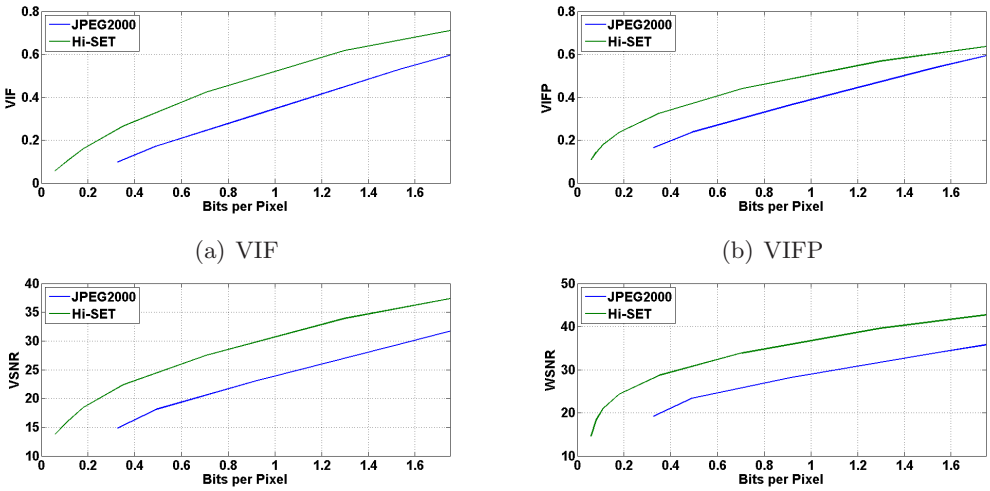
Figure B.7: Gray-Scale CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



(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.



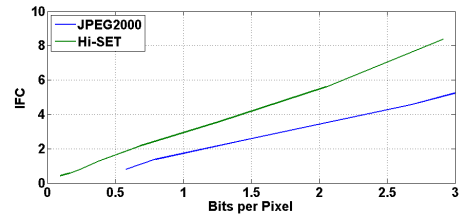
(a) VIF (b) VIFP

(c) VSNR (d) WSNR

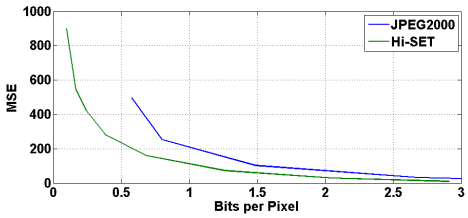
Figure B.9: Gray-Scale CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.2.2 Color Images

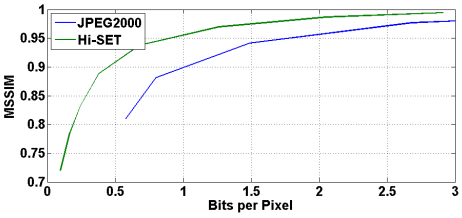
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



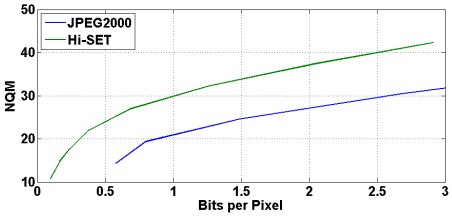
(a) IFC



(b) MSE

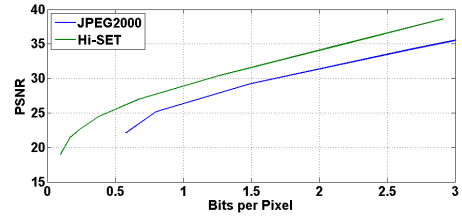


(c) MSSIM

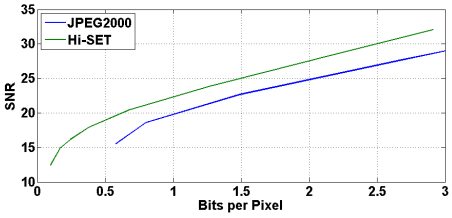


(d) NQM

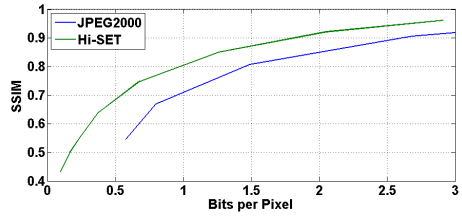
Figure B.10: Color CSIQ Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



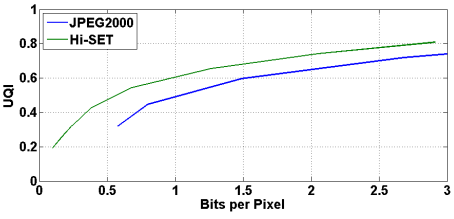
(a) PSNR



(b) SNR

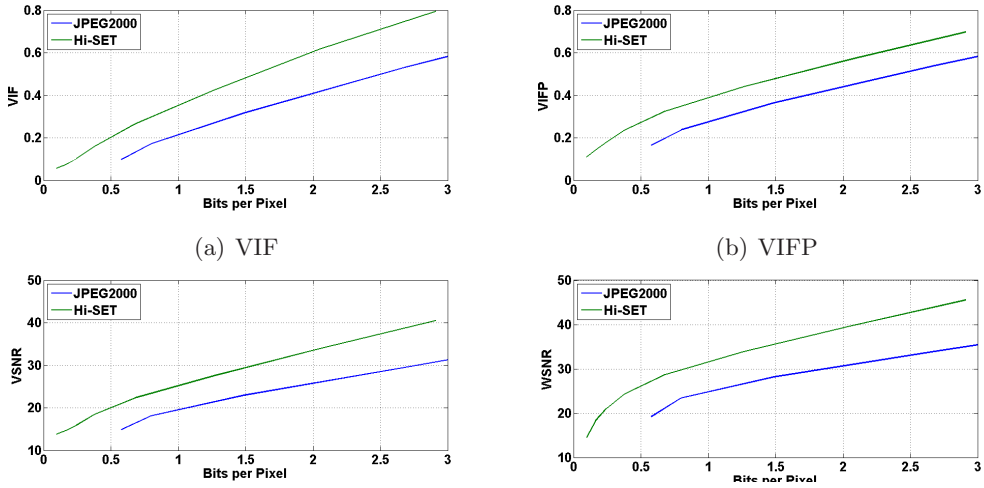


(c) SSIM



(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

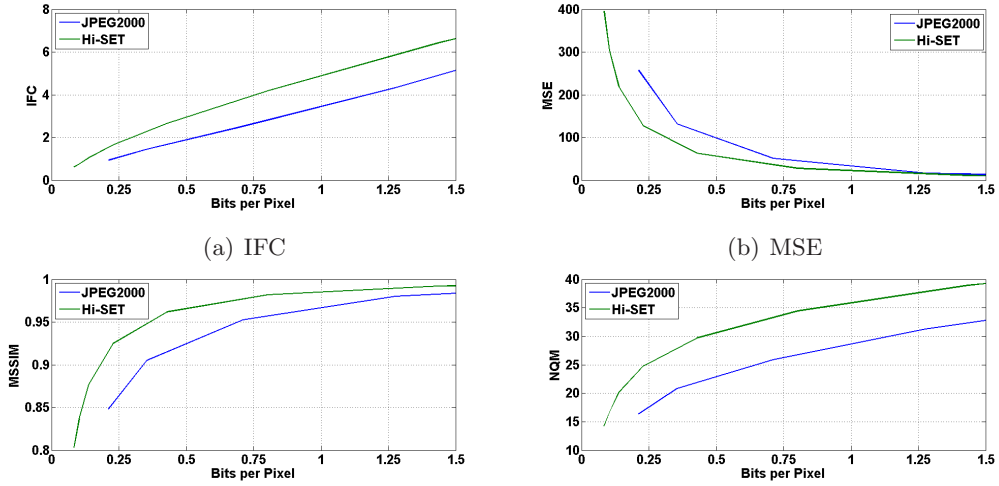
(c) VSNR (d) WSNR

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 vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.

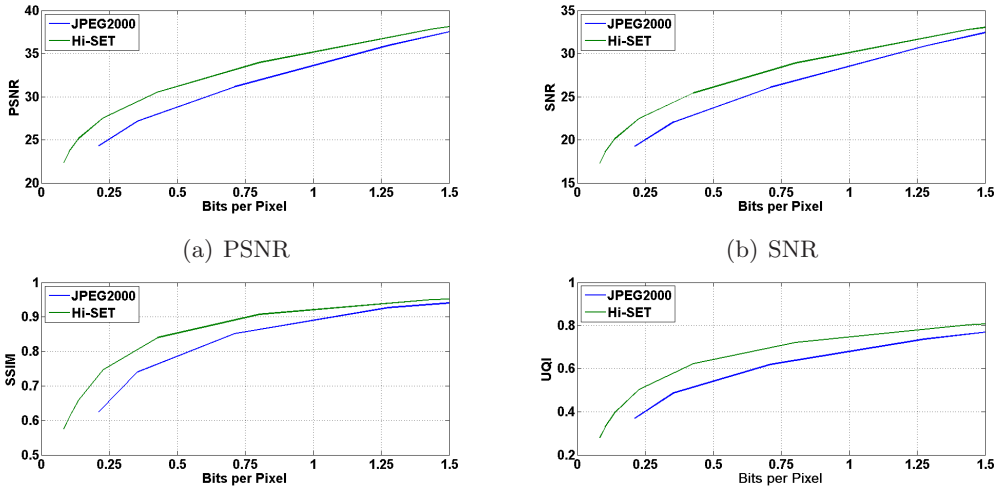


(a) IFC (b) MSE

(c) MSSIM (d) NQM

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

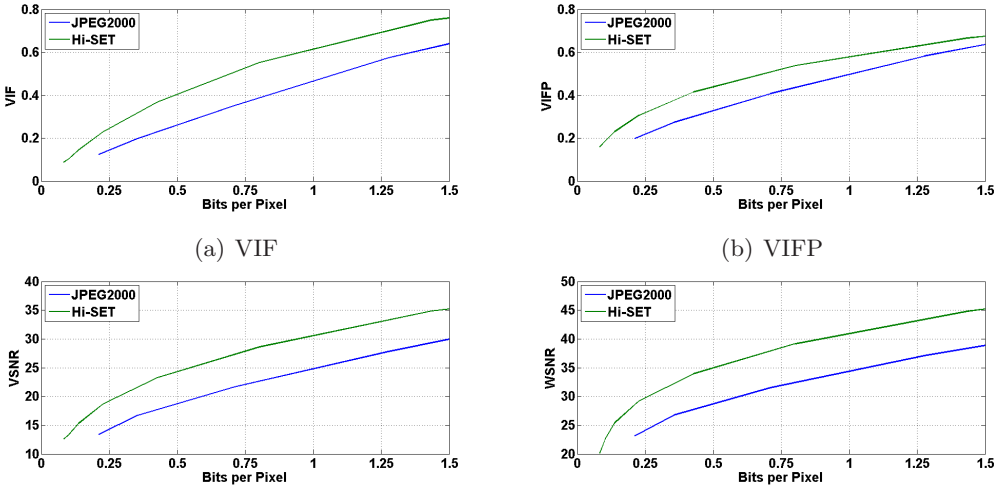
B. JPEG2000 VS H*i*-SET: COMPLEMENTARY RESULTS OF CHAPTER 3



(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.



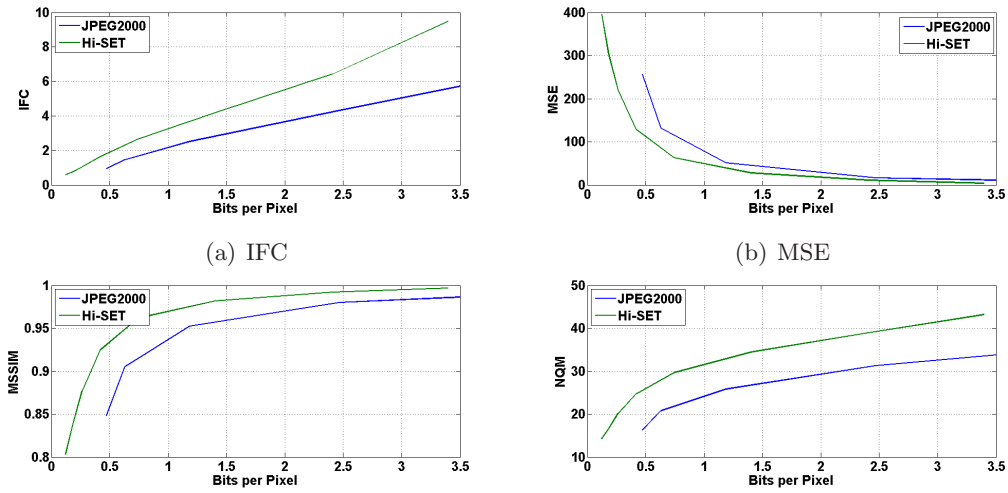
(a) VIF (b) VIFP

(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

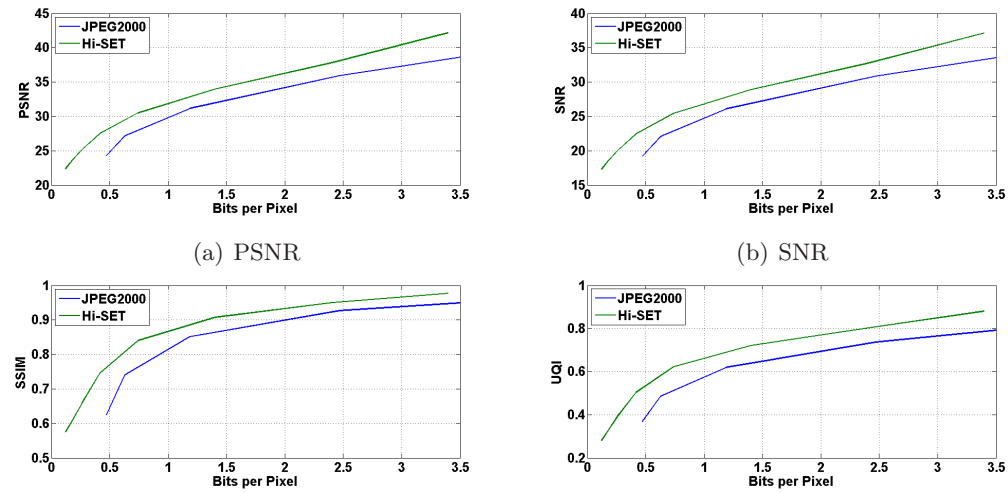
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

(c) MSSIM (d) NQM

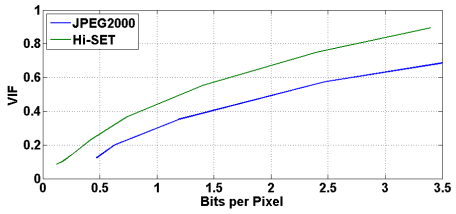
Figure B.16: Color IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



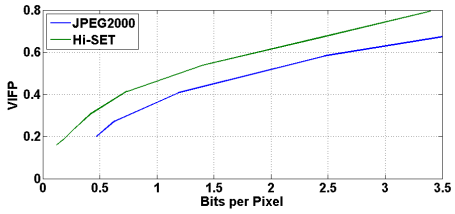
(a) PSNR (b) SNR

(c) SSIM (d) UQI

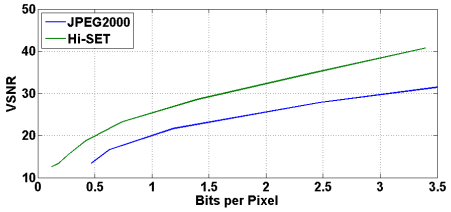
Figure B.17: Color IVC Image Database: JPEG2000 vs H*i*-SET. Metrics employed: PSNR, SNR, SSIM and UQI.



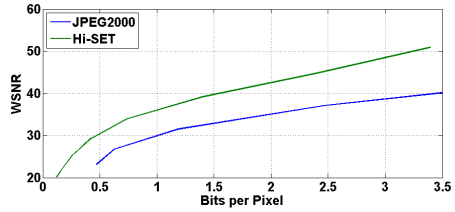
(a) VIF



(b) VIFP



(c) VSNR



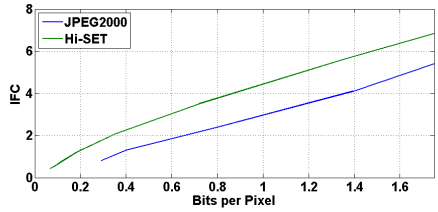
(d) WSNR

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

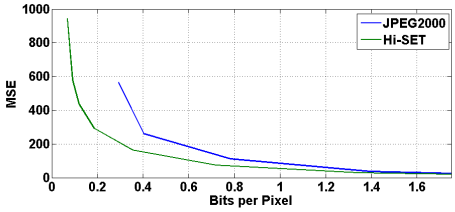
B.4 Image Database of the Laboratory for Image and Video Engineering

B.4.1 Gray-Scale (Y Channel)

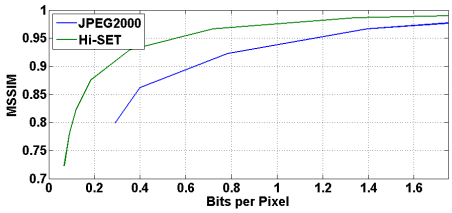
Compression of Gray-Scale Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



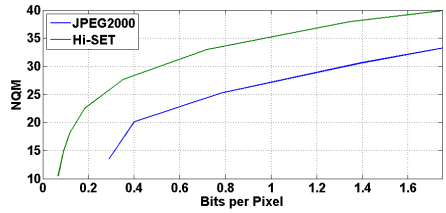
(a) IFC



(b) MSE

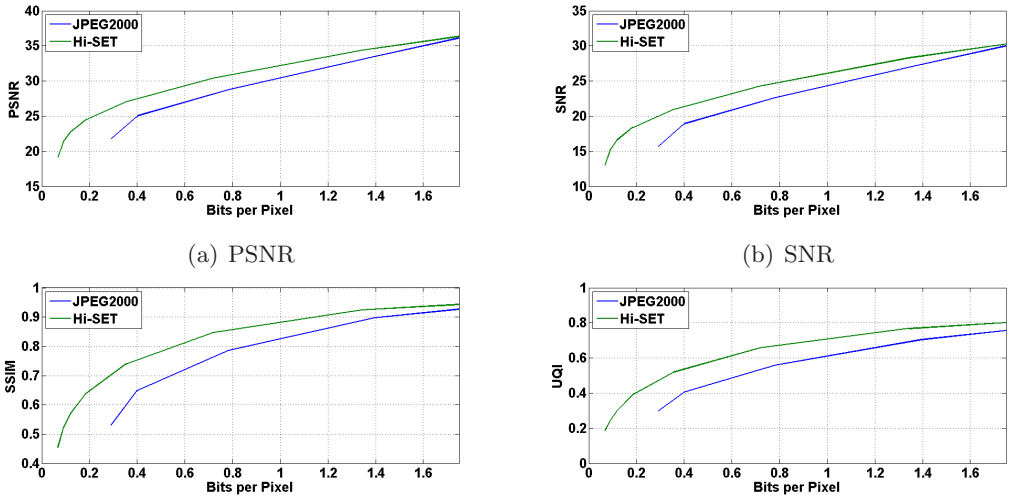


(c) MSSIM



(d) NQM

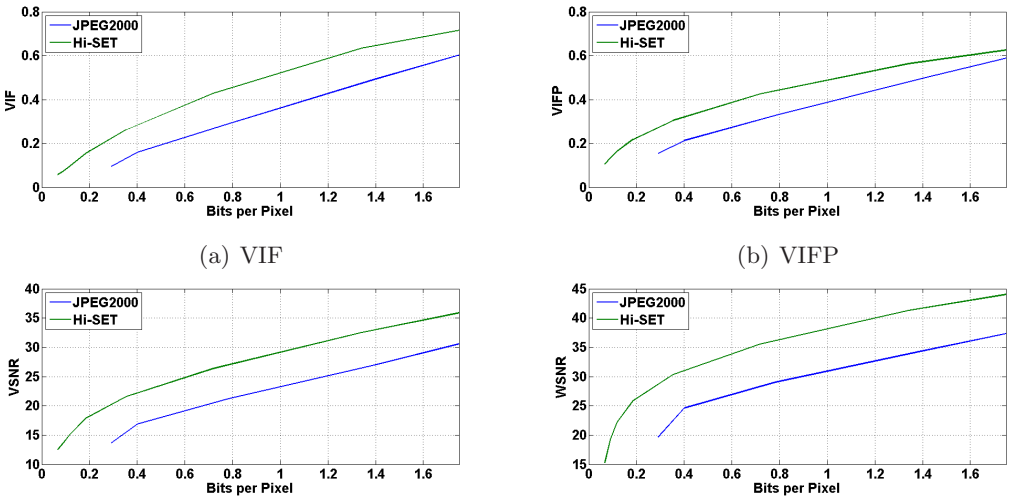
Figure B.19: Gray-Scale LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



(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.



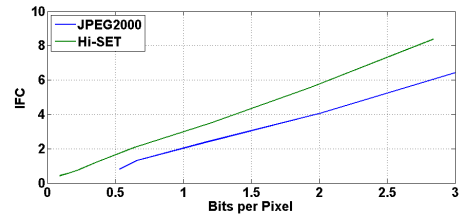
(a) VIF (b) VIFP

(c) VSNR (d) WSNR

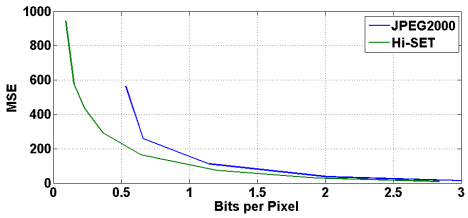
Figure B.21: Gray-Scale LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: VIF, VIFP, VSNR and WSNR.

B.4.2 Color Images

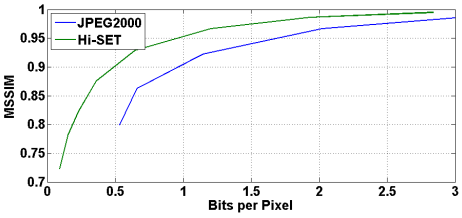
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



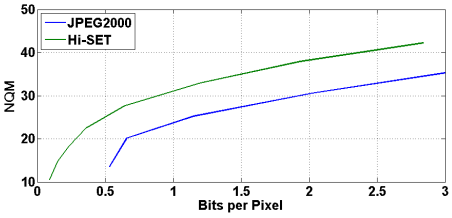
(a) IFC



(b) MSE

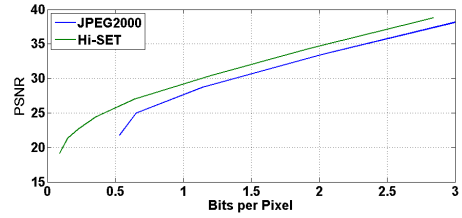


(c) MSSIM

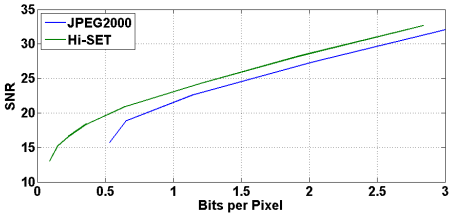


(d) NQM

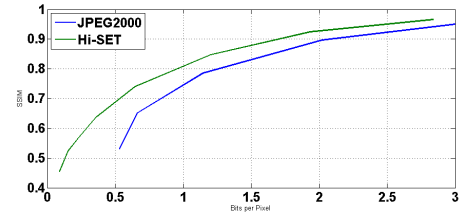
Figure B.22: Color LIVE Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



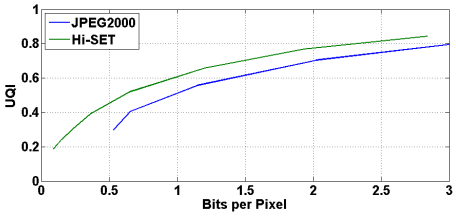
(a) PSNR



(b) SNR

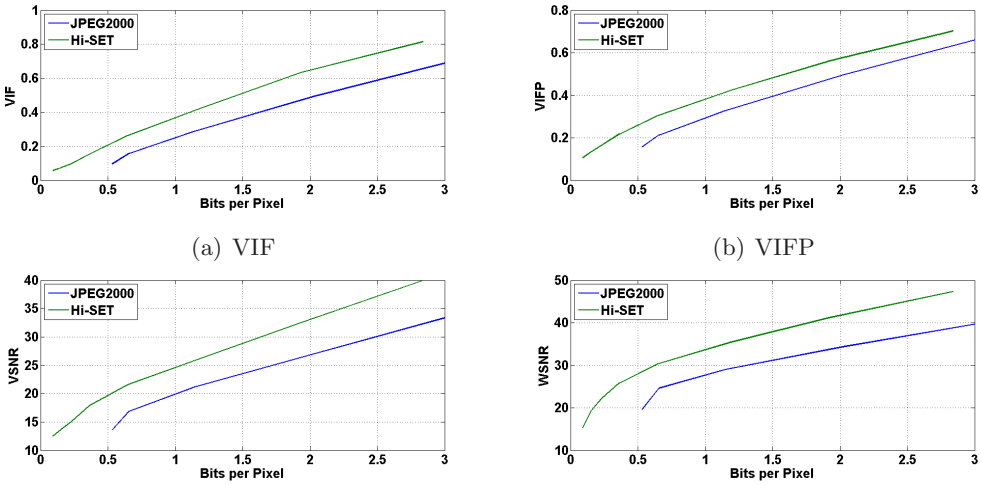


(c) SSIM



(d) UQI

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



(a) VIF (b) VIFP

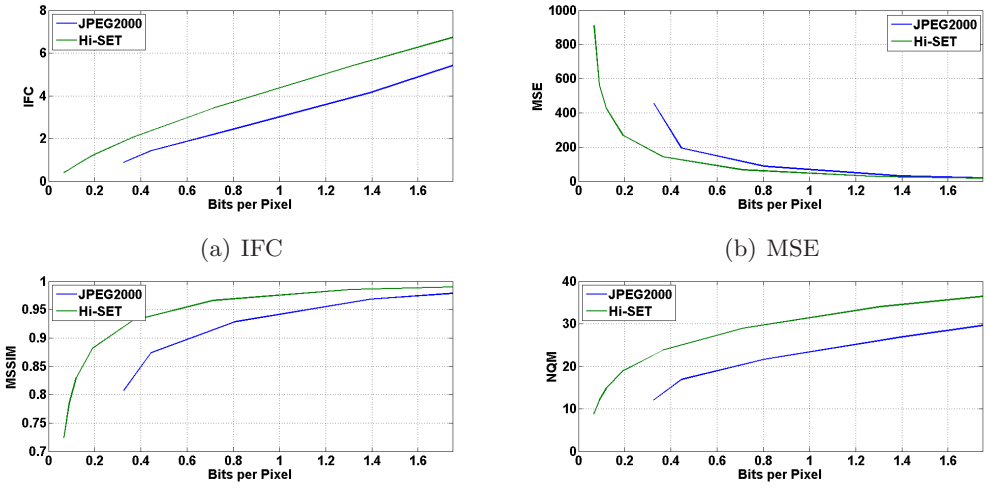
(c) VSNR (d) WSNR

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)

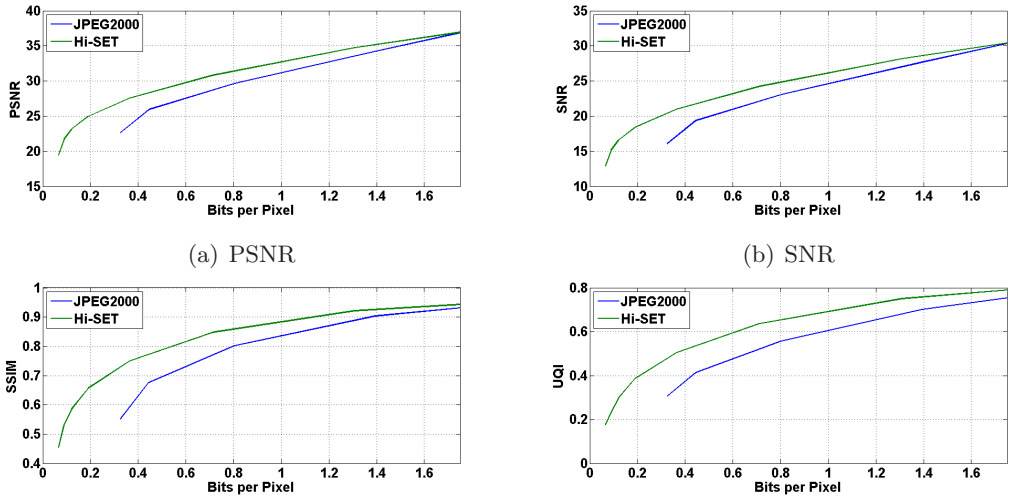
Compression of Gray-Scale Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

(c) MSSIM (d) NQM

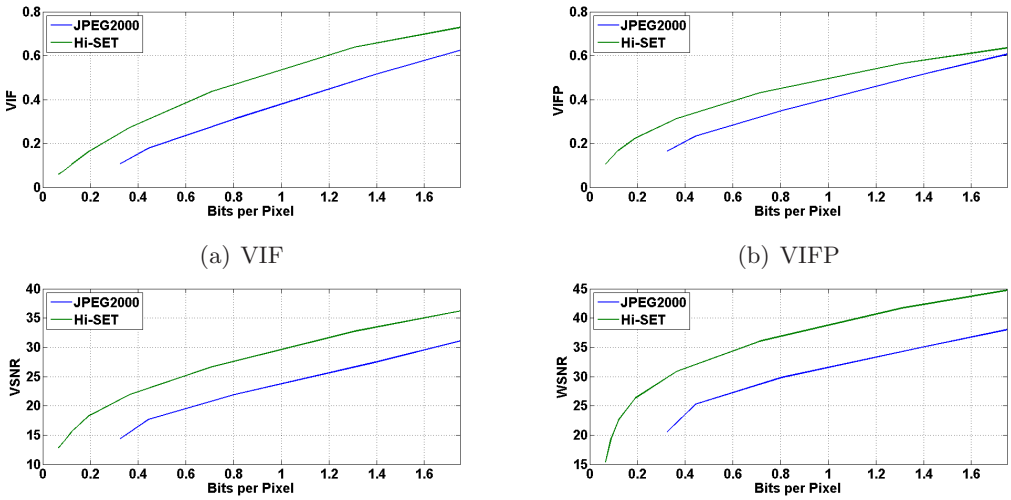
Figure B.25: Gray-Scale TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



(a) PSNR (b) SNR

(c) SSIM (d) UQI

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



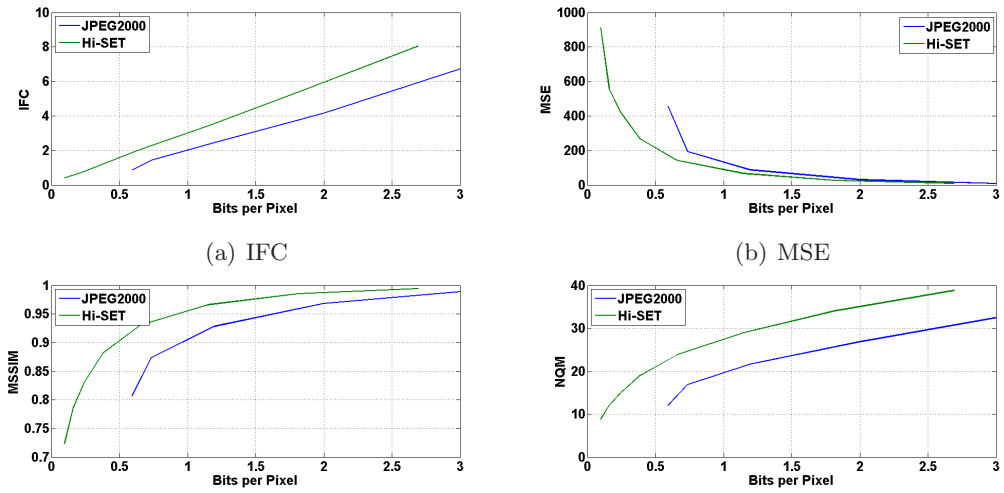
(a) VIF (b) VIFP

(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

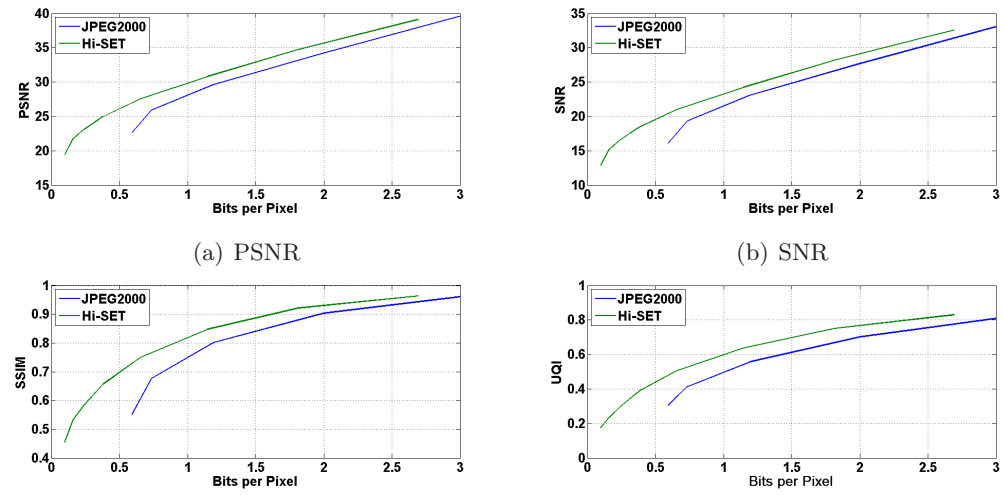
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by H*i*-SET coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

(c) MSSIM (d) NQM

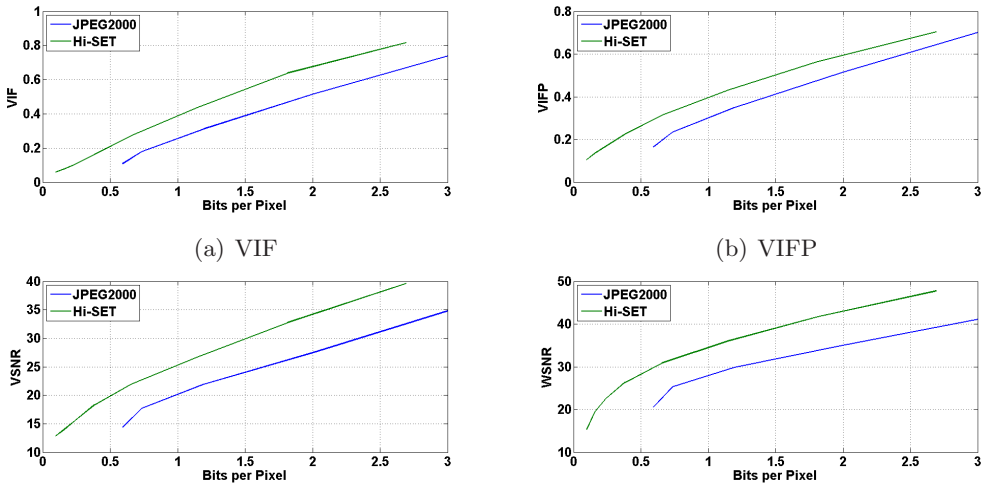
Figure B.28: Color TID2008 Image Database: JPEG2000 vs H*i*-SET. Metrics employed: IFC, MSE, MSSIM and NQM.



(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

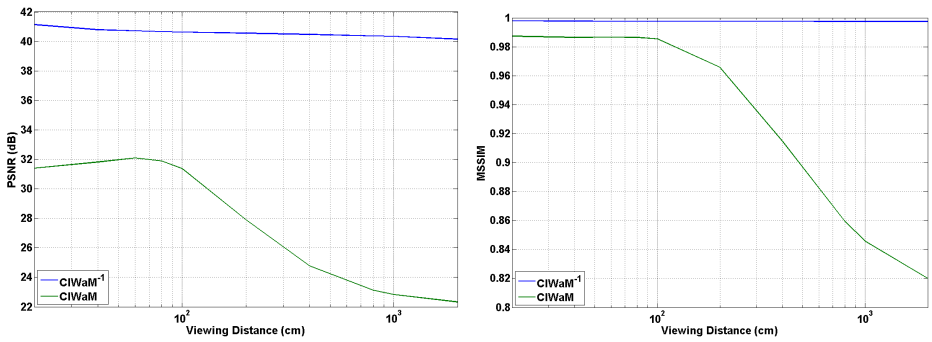
Complementary Results of Chapter 4

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

Green functions denoted as CIWaM are perceptual quantized images after applying $\alpha(\nu, r)$, while blue functions denoted as CIWaM^{-1} are recovered images before applying $\widehat{\alpha}(\nu, r)$.

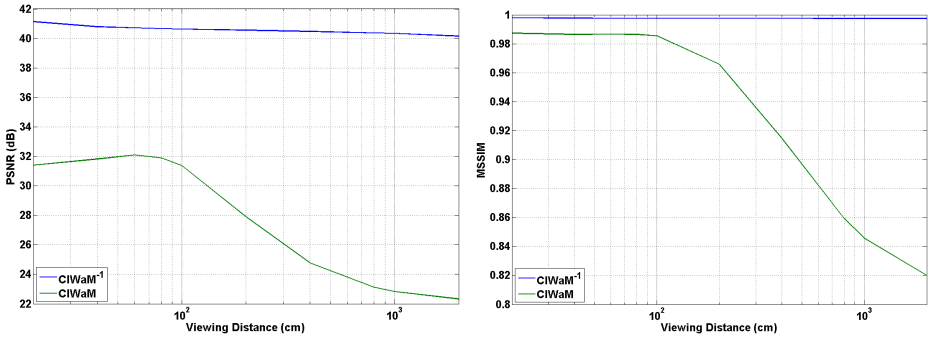
C.1.1 Categorical Subjective Image Quality Image Database

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.

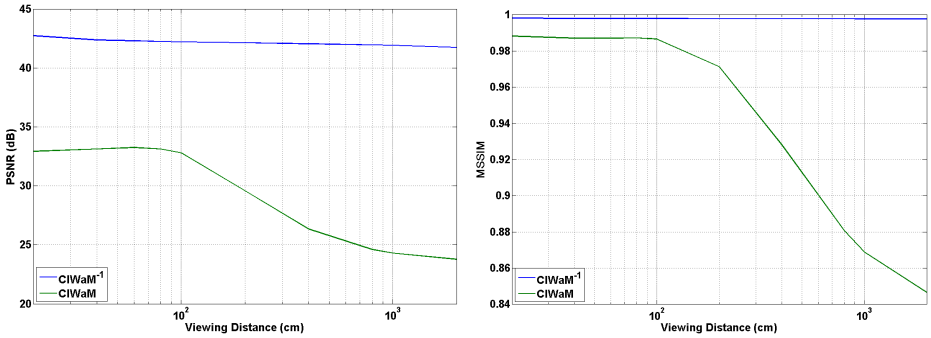


(a) PSNR (b) MSSIM

Figure C.2: Perceptual Quantization of Color Images of the CSIQ image database.

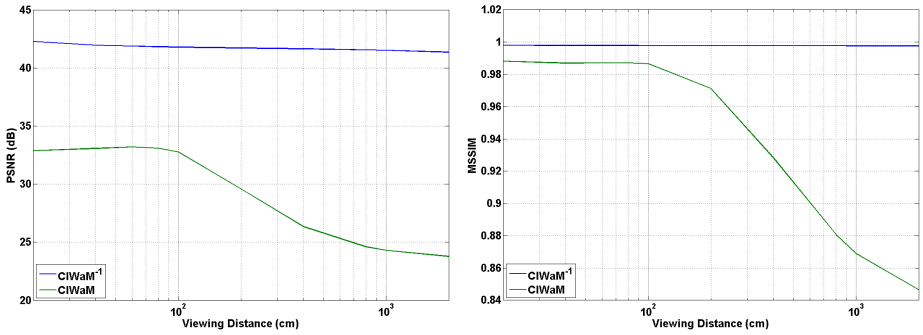
C.1.2 Image and Video-Communication Image Database

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.



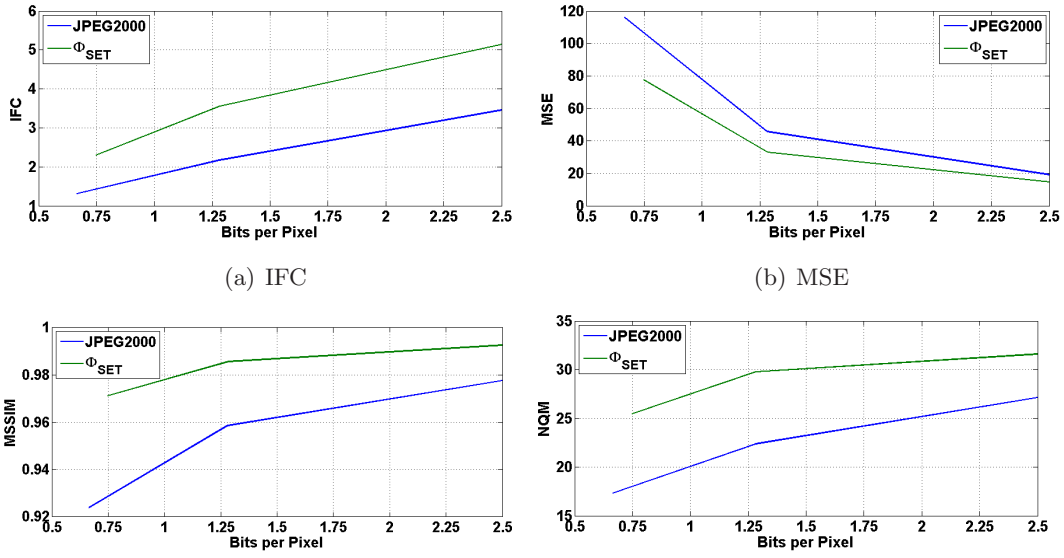
(a) PSNR (b) MSSIM

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

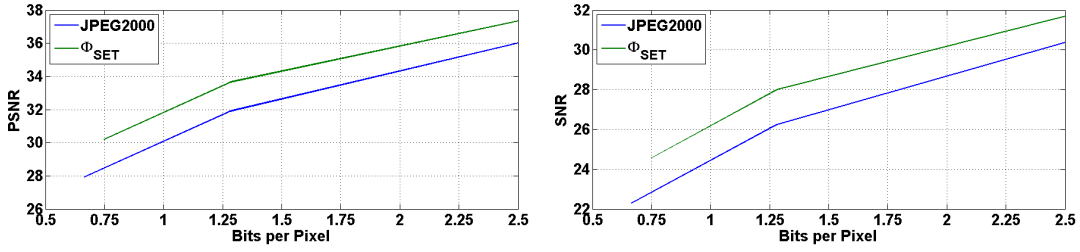
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by Φ_{SET} coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

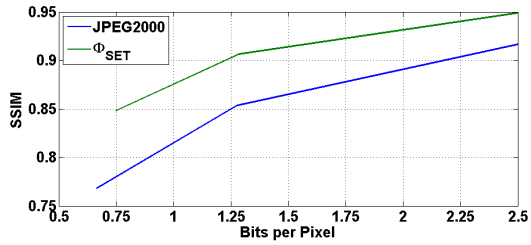
(c) MSSIM (d) NQM

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

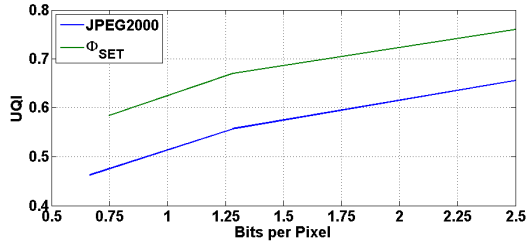


(a) PSNR

(b) SNR

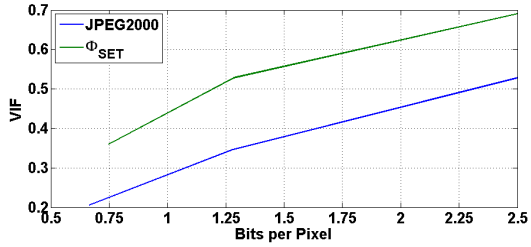


(c) SSIM

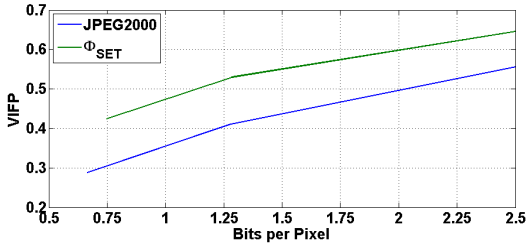


(d) UQI

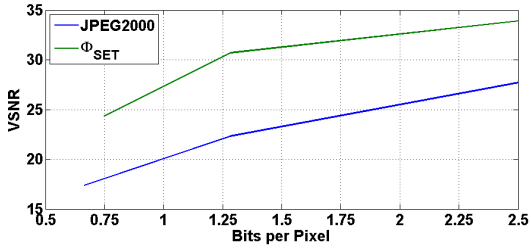
Figure C.6: Color CMU Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: PSNR, SNR, SSIM and UQI.



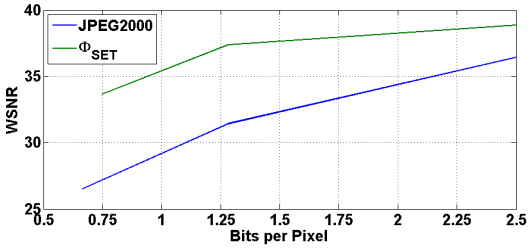
(a) VIF



(b) VIFP



(c) VSNR

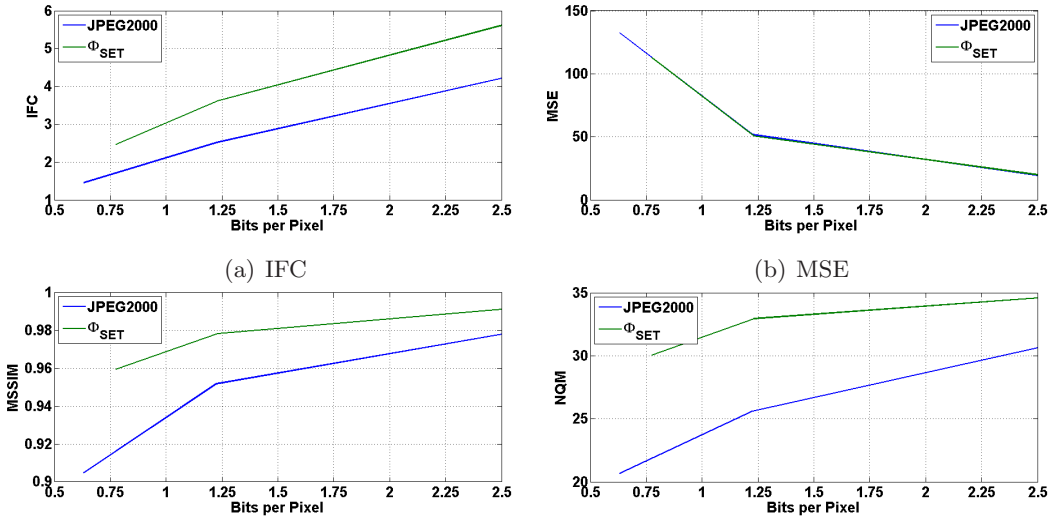


(d) WSNR

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

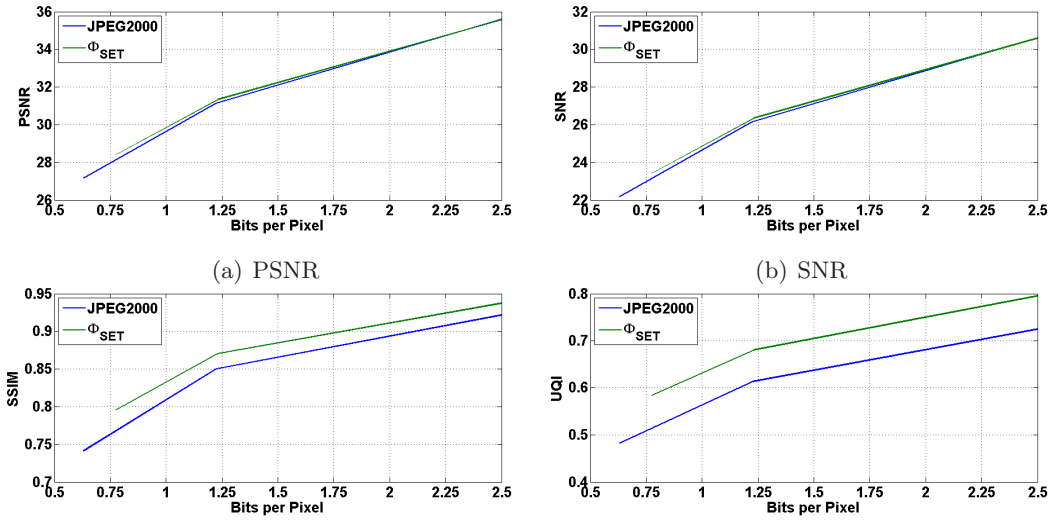
Compression of Color Images vs Image Quality Assessment. Green functions represent results obtained by Φ_{SET} coder, while blue functions by JPEG2000 coder.



(a) IFC (b) MSE

(c) MSSIM (d) NQM

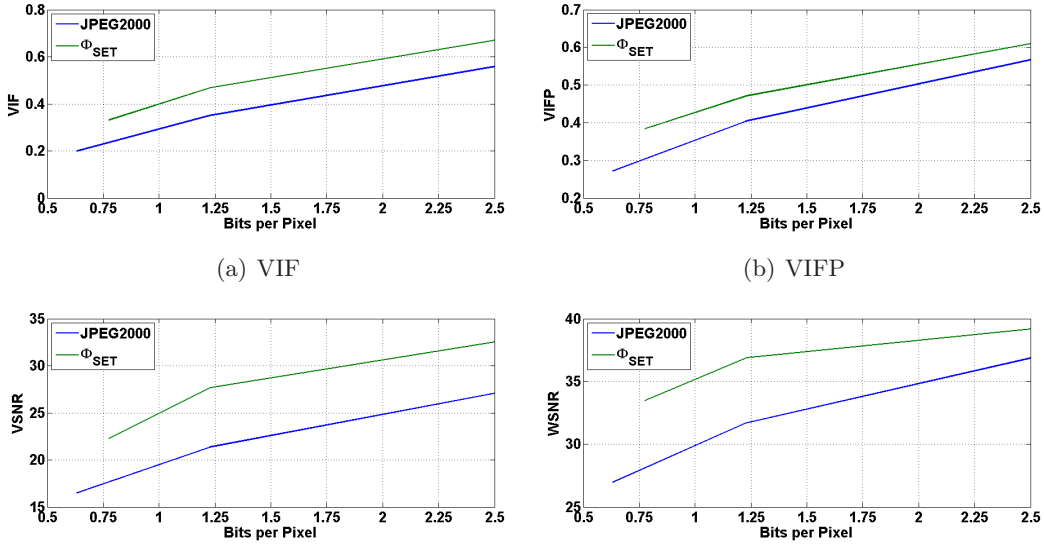
Figure C.8: Color IVC Image Database: JPEG2000 vs Φ_{SET}. Metrics employed: IFC, MSE, MSSIM and NQM.



(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. 24

[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, 57, 82, 93

[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. 29

[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. 71

[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

applied to map overlapping in geographic information systems. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 5001 –5004, 23-28 2007. 80

[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. xiv, 44, 45

[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. 1, 29, 32, 43, 55, 56, 57, 71

[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. 71, 77

[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. 24, 52, 67

[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. 71

[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. 24, 52, 67

[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. 80

[16] David Hilbert. Über die stetige Abbildung einer Linie auf ein Flächenstück. *Mathematische Annalen*, **38**(3):459–460, Sept. 1891. xiv, 33

[17] M. Hollander and D.A. Wolfe. *Non-parametric Statistical Methods*. Wiley, 2nd edition, 1999. 24

[18] Q. Huynh-Thu and M. Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, **44**(13):800–801, 2008. 24, 52, 67

[19] ISO/IEC 12640-1. Graphic technology - prepress digital data exchange - CMYK standard colour image data (CMYK/SCID), 1997. 51

[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. xiv, 44, 45

[21] Cornell University Visual Communications Laboratory. MeTriX MuX visual quality assessment package , available at *http* : *//foulard.ece.cornell.edu/gaubatz/metrix_mux/*, 2010. 24

[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, 92

[23] Patrick le Callet and Florent Autrusseau. Subjective quality assessment IRCCyN/IVC database, 2005. http://www.irccyn.ec-nantes.fr/ivcdb/. 9, 23, 89

[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. 31

[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. 24, 52, 59, 67

[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. 88

[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. 71

[31] David Nister and Charilaos Christopoulos. Lossless region of interest coding. *Signal Processing*, **78**(1):1 – 17, 1999. 71

[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. 27, 32, 35

[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. 27, 35

[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. 27

[37] PEIPA. Pilot european image processing archive, available at http://peipa.essex.ac.uk/. 81

[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, 22, 90

[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, 22, 46, 90

[40] Cannon Research, École Polytechnique Fédérale de Lausanne, and Ericsson. JJ2000 implementation in Java, available at http://jj2000.epfl.ch/, 2001. 46, 57, 70

[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. 27, 31, 88

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

[43] 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. 27, 88

[44] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, **15**(2):430 –444, feb. 2006. 9, 24, 52, 67

[45] 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, 91

[46] 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. 24, 52, 67

[47] Athanassios Skodras, Charilaos Christopoulos, and Touradj Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine*, **18**(5):36–58, September 2001. 29, 71, 88

[48] Wim Sweldens. The lifting scheme: A custom-design construction of biorthogonal wavelets. *Applied and Computational Harmonic Analysis*, **3**(2):186 – 200, 1996. 29

[49] David Taubman. Kakadu software, available at http://www.kakadusoftware.com/, July 2010. 46, 70, 77

[50] David S. Taubman and Michel W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. ISBN: 0-7923-7519-X. Kluwer Academic Publishers, 2002. 2, 8, 26, 29, 44, 67, 71

[51] Bryan E. Usevitch. A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000. *IEEE Signal Processing Magazine*, **18**(5):22–35, 2001. 27

[52] 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. 14

[53] 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, 24, 52, 67

[54] Zhang Wang and Alan Bovik. A universal image quality index. *IEEE Signal Processing Letters*, **9**:81–84, 2002. 24, 52, 67

[55] 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. 74

[56] 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

[57] 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, 24, 52, 67

[58] 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, 29

[59] 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. 73

[60] 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

[61] Bettye Wilson. *Ethics and Basic Law for Medical Imaging Professionals*. F.A. Davis Co., 1997. 80

[62] 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. 82

Publications

Journals

- Jaime Moreno and Xavier Otazu, Image Coder Based on Hilbert Scanning of Embedded quadTrees, *Digital Signal Processing*, Submitted.
- Jaime Moreno and Xavier Otazu, Full-Reference Quality Assessment using a Perceptual Chromatic Induction Model. Application to JPEG and JPEG2000 images, *Journal of the Optical Society of America A*, 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.

PUBLICATIONS

- 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