# A Human Visual System Based Image Coder

**A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy**


## Chin Soon Tan

**Master of Information Technology**

**School of Electrical and Computer Engineering**

**RMIT University**


**March 2009**

# Abstract

Over the years, society has changed considerably due to technological changes, and digital images have become part and parcel of our everyday lives. Irrespective of applications (i.e., digital camera) and services (information sharing, e.g., Youtube, archive / storage), there is the need for high image quality with high compression ratios. Hence, considerable efforts have been invested in the area of image compression. The traditional image compression systems take into account of statistical redundancies inherent in the image data. However, the development and adaptation of vision models, which take into account the properties of the human visual system (HVS), into picture coders have since shown promising results.

The objective of the thesis is to propose the implementation of a vision model in two different manners in the JPEG2000 coding system: (a) a Perceptual Colour Distortion Measure (PCDM) for colour images in the encoding stage, and (b) a Perceptual Post Filtering (PPF) algorithm for colour images in the decoding stage. Both implementations are embedded into the JPEG2000 coder. The vision model here exploits the contrast sensitivity, the inter-orientation masking and intra-band masking visual properties of the HVS. Extensive calibration work has been undertaken to fine-tune the 42 model parameters of the PCDM and Just-Noticeable-Difference thresholds of the PPF for colour images. Evaluation with subjective assessments of PCDM based coder has shown perceived quality improvement over the JPEG2000 benchmark with the MSE (mean square error) and CVIS criteria. For the PPF adapted JPEG2000 decoder, performance evaluation has also shown promising results against the JPEG2000 benchmarks. Based on subjective evaluation, when both PCDM and PPF are used in the JPEG2000 coding system, the overall perceived image quality is superior to the stand-alone JPEG2000 with the PCDM.

# A Human Visual System Based Image Coder

## *Declaration*

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; and, any editorial work, paid or unpaid, carried out by a third party is acknowledged.

Chin Soon Tan
23 March 2009

# Acknowledgements

I thank my God, the Father of our Lord Jesus Christ, who has helped me throughout my darkest hour. He helped me through my most depressing state. He gave me wisdom and encouragement so much so that I can see the day of the completion of my thesis writing. To Him be the Glory, Amen.

I will also like to thank the various people: Prof. Hong Ren Wu for his input, Dr. Damian Tan for his guidance and patience throughout this period; my fellow research mates, James Mei and David Wu, for their encouragement and support.

A special thank you also to my two children, Claudia and Moses, who had been most understanding when I had to spend hours slogging over my thesis; the brothers and sisters in Christ who have prayed and encouraged me. To my dearest wife, Kok Nee, I owe her for her support, patience, and understanding.

# List of Publications by Author

1. C. S. Tan, D. M. Tan, and H. R. Wu, "Perceptual Coding of Digital Colour Images Based on a Vision Model," in *Proceedings of IEEE International Symposium on Circuits and Systems*, Vancouver, Canada, 23-26 May 2004, pp. V-441-V-444.

2. C. S. Tan and H. R. Wu, "Vision Model Based Perceptual Post Filtering of JPEG2000 Coded Colour Images," in *Proceedings of SPIE Conference: Visual Communications and Image Processing 2005*, Jul 2005.

3. C. S. Tan and H. R. Wu, "Common and Separate Parameterizations of Vision Model Based Perceptual Post Filtering for Digital Colour Images," in *Proceedings of the TENCON2005 IEEE Region 10 Conference*, Melbourne, Victoria Australia, Nov 2005.

4. C. White, R. Martin, D. Wu, C. S. Tan, D. M. Tan, H. R. Wu, and J. Cai, "Subjective Image Quality Assessment at Threshold Level," in *Proceedings of the TENCON 2005 IEEE Region 10 Conference*, Melbourne, Victoria, Australia, Nov 2005.

5. D. M. Tan, C. S. Tan, and H. R. Wu, "Perceptual Color Image Coding With JPEG2000," *IEEE Transactions on Image Processing,* vol. 19, pp. 374 - 383, Feb 2010.

# Table of Contents

# List of Tables

# List of Figures

# List of Common Abbreviations

| | |
|---|---|
| 1-D | One Dimension |
| 2-D | Two Dimension |
| bpp | Bit Per Pixel |
| CI | Confidence Interval |
| CGC | Contrast Gain Control |
| CMP | Common Model Parameterisation |
| CSF | Contrast Sensitivity Function |
| CVIS or VDM | Visual Distortion Metric |
| DCT | Discrete Cosine Transform |
| DCTune | see page 65 |
| d.f. | Degree of Freedom |
| DFT | Discrete Fourier Transform |
| DPCM | Differential Pulse Code Modulation |
| DICOM | Digital Imaging and Communications in Medicine |
| DWT | Discrete Wavelet Transform |
| EBCOT | Embedded Block Coding with Optimised Truncation |
| EZW | Embedded Zero-tree Wavelet |
| GQMF | Generalised Quadrature Mirror Filter |
| HDTV | High Definition TV |
| HVS | Human Visual System |
| JNCD | Just Noticeable Colour Difference |
| JND | Just Noticeable Difference |
| JPEG | Joint Photographic Experts Group |
| JPEG2000 | Still Image Compression Standard developed by the Joint Photographic Experts Group |
| JPEG-LS | Image Compression Standard for Lossless and Near Lossless Compression of Continuous-tone, Gray Scale and Colour Still Images |
| JBIG2 | Image Compression Standard developed by the Joint Bi-level Image Expert Group |
| KLT | Karhunen-Loeve Transform |
| LAPIC | Locally Adaptive Perceptual Image Coding |
| LGN | Lateral Geniculate Nucleus |
| LSB | Least Significant Bit |
| LSBP | Least Significant Bitplane |
| LSF | Linespread Function |
| LWT | Lifting Wavelet Transform |
| M Cells | Magnocellular Cells |
| MAE | Mean Absolute Error |
| MND | Minimally Noticeable Distortion |
| MSB | Most Significant Bit |
| MSBP | Most Significant Bitplane |
| MSE | Mean Square Error |
| MTF | Modulation Transfer Function |
| P Cells | Parvocellular Cells |

| | |
|---|---|
| PCDM | Perceptual Colour Distortion Measure |
| PCRD | Post Compression Rate Distortion |
| PDF | Probability Distribution Function |
| PDM | Perceptual Distortion Metric |
| PIDM | Perceptual Image Distortion Metric |
| PPF | Perceptual Post Filtering |
| PSF | Point Spread Function |
| Q MF | Quadrature Mirror Filter |
| R-D | Rate Distortion |
| RMS | Root Mean Square |
| SMP | Separate Model Parameterisation |
| SNR | Signal-to-Noise Ratio |
| SPIHT | Set Partitioning in Hierarchical Tree |
| TvC | Target Contrast Threshold Versus Masker Contrast |
| V1 | Primary Visual Cortex |
| VDP | Visible Difference Predictor |
| VQ | Vector Quantisation |
| WITCH | Wavelet-based Image/Texture Coding Hybrid |
| WVDP | Wavelet Visible Difference Predictor |

# Chapter 1 Introduction

Data Compression is concerned with the removal of redundancies [1]. Data compression has become prevalent since the advent of the digital age with dependency on digital data. With the prevalence of digital media in our everyday lives and the use of images to convey information, images are now an integral part of our modern lifestyle. One can relate how an image of war-torn country speaks louder than a thousand words describing the scene. Moreover, with the increase in popularity of websites like Facebook[1], where one shares information and digital images freely over the internet, and the Google Earth[2] where one can find satellite images for maps and directions, the need for image compression becomes clear.

With the surge of the internet and intranet use, there exists a possibility that network traffic volume may exceed its capacity, thereby affecting transmission speed. Some have argued against the need for image compression as there is now greater availability of high-bandwidth broadband cable networks. However, as the issues surrounding the cost of providing and maintaining broadband access to the wider community (e.g., who is to bear the cost, cost of subsidies to Telcos) have been so greatly contested at both the local and higher governments[3], the need for image compression still persists. This is evident with the total switch of analog to digital High Definition TV (HDTV) in the near future in some countries, thus, the need for picture compression looms greater. Limitation in electronic data storage space also dictates the need for data compression to prevent an overflow of data storage [1].

Even at the individual consumer level, the need for electronic data storage space will always exist. With the increased use of digital images, e.g., digital photography used in cameras and mobile phones, there will always be a problem of "not enough disk space" or "not enough memory space". Hence the research of image compression has

[1] Facebook is social networking website launched on February 4, 2004. http://www.facebook.com/facebook.
[2] Google Earth is a virtual globe program. It maps the earth by the superimposition of images obtained from satellite images and aerial photography. http://earth.google.com.
[3] State or Federal Governments.

much bearing in the application for the world of consumer electronics such as digital image cameras. Furthermore, image compression has also gained inroads into other areas in medical imaging such as JPEG-LS [2, 3] and DICOM [4] for medical field especially in the areas of medical imaging [5-7], FBI Compression for finger printing [8, 9] for defence, security, and law enforcement.

## *1.1 Research Areas in Image Compression*

Image compression involves the removal of data redundancies in an image. This is also referred to by Shannon as statistical redundancy with "noise" [10]. In the premise of this thesis, two approaches of image compression are most poignant: lossy and lossless compression. Both compression philosophies seek to remove redundancies within images. However, in lossy compression, image quality is compromised to allow for a higher compression ratio. The loss of information accompanying the lossy compression is the result of quantisation. Conversely, a lossless compression seeks to achieve an optimal compression ratio without compromising image quality. The JPEG baseline [11] (established to standardise image compression techniques) uses the block based DCT approach and concentrates on removing the statistical redundancies which are computed from the mean squared error (MSE) [12]. More recently, in the JPEG2000 standard [12, 13], the embedded block coding with optimized truncation (EBCOT) [14] has been adopted. The EBCOT uses the rate-distortion function to achieve optimal quality for a given bit rate [12, 14]. Consequently, EBCOT's main features are scalability in quality and resolution.

However, there has been a growing research in the area of an image coder based on the human visual system (HVS). Apart from the statistical redundancies, there are some redundancies which are imperceptible to the human eye. These redundancies are known as psychovisual redundancies. Removal of these redundancies gives rise to perceptually lossy [15] or perceptual lossless compression [16]. Being modeled after the human eye, this vision model [15, 16] takes into consideration the physiological and psychological studies in relation to the human visual systems and the interactions of these visual signals with our human brain [17, 18]. The neural responses that form the visual images are arranged in a manner which is both frequency and orientation

selective [19-21]. One particular neural phenomenon that has direct bearing on our visual perception is masking, which intrinsically decreases the strength of some neural signals. This masking effect has been modelled by some researchers [22-26]. For example, the contrast gain control model booted by Watson and Solomon attempts to incorporate the quantifiable properties of the HVS, namely contrast, frequency, orientation and masking sensitivities [27].

Having established the HVS model, there is the next step of applying the HVS model to a coding structure. Several approaches have been identified, such as pre-filtering to reduce visual redundancies, post-filtering to reduce distortions or designing quantisation matrices specific to aspects of the HVS. In some cases, vision model is incorporated into the distortion function. For optimisation of the vision model, parameterisation is required, i.e., the parameters of the model are calibrated to attain optimal visual quality.

## *1.2 Objective and Organisation of Thesis*

The objective of this thesis is to design a perceptual colour image coder based on the Human Visual System (HVS). The proposed coder employs the JPEG2000 [12] structure. As the coder is based on the HVS, there is a need to underline the physiology and psychophysical studies relating to the human eye. Chapter 2 gives a detailed account of the physical eye and its interactions with the human brain to form neural images. Psychophysical experiments related to mammalian visual system are outlined in the chapter [19-21, 28-30]. This chapter provides insights into the human eye and lays the premises relating to the HVS model.

Chapter 3 begins with a general description of the image compression systems, namely lossy and lossless compressions and the underlying theory of image compression, i.e., Shannon's theory of noiseless source coding and rate distortion theory [10]. The rate distortion theory is concerned with the relation between bit rate and image quality. An overview of the various elements in an image compression system is also discussed, with particular emphasis on the various transform and quantisation methods [12]. In particular, the block-based transform and bitplane

quantisation forms part of the framework of the Perceptual Colour Distortion Measure (PCDM) discussed in chapter 4.

A comparison of the various image bitplane coders are presented, beginning with the Embedded Zero-tree Wavelet (EZW) [31], the Set Partitioning in Hierarchical Tree (SPIHT) [32] and the Embedded Block Coding with Optimised Truncation (EBCOT) [14]. The EBCOT is regarded as superior to EZW and SPIHT in terms of its Signal-to-Noise Ratio (SNR) and resolution scalability [14]. Consequently, the JPEG2000 which is based on the EBCOT structure is now hailed as the current state-of-art coder.

Since human observers are the ultimate judges of image quality, perceptual image coders based on the HVS have gained attention. Ultimately, the goal of these perceptual models is to improve perceived image quality. A literature review of perceptual image coders is provided in chapter 3 to give an overview of the current development of perceptual image coders. The model proposed by Tan et al. [15], which forms the basis of the development of the PCDM model for colour image and the Perceptual Post-Filtering (PPF) algorithm, is also presented.

Chapter 4 presents the Perceptual Colour Distortion Measure (PCDM) coder for colour image and the parameterisation of its HVS model. It is extensively calibrated to improve visual quality at medium to low bit rates. The subjective assessment results and the test images involving about thirty participants are also presented to ascertain the performance of the PCDM based coder.

In chapter 5, a perceptual post-filtering (PPF) algorithm based on the HVS model is developed to attempt to recover the loss of visual information. The preliminary subjective assessment tests show promising results for the algorithm.

Finally, chapter 6 concludes with an overview of the contribution of this thesis and directions for future research.

## 1.3 Contributions

The contributions of this thesis are as follows:

a. An adaptation of the monochromatic based PIDM (Perceptual Image Distortion Metric) into colour based PCDM model in the *YCbCr* colour space. The resulting model, PCDM, is adapted to JPEG2000 coder.
b. The calibration of the 42 PCDM parameters. Two sets of sub-optimal values were obtained.
c. Subjective assessment of proposed PCDM based coder vs JPEG2000-MSE and JPEG2000-CVIS was carried out with 30 subjects for performance evaluation. Results showed that the PCDM produces image with better perceived quality than the benchmarks.
d. Adaption of the PPF algorithm to the JPEG2000 decoder to recover the loss of visual information due to compression operation.
e. Threshold points of PPF were obtained through subjective experiment. The thresholds are set at the Just-Noticeable-Difference (JND) level.
f. Performance evaluations of the PPF based decoder and the PCDM with the PPF codec were conducted through subjective tests against JPEG2000-MSE and JPEG2000-CVIS. Perceptual improvement in picture quality is obtained for both proposed implementations against the JPEG2000 benchmarks.
g. Subjective evaluation of the PPF algorithm with separate model parameterisation (SMP) against the PPF algorithm with common model paramterisation (CPM). The SMP implementation did not show better perceived picture quality than the CMP.

# Chapter 2 Studies of Human Visual System

## *2.1 Overview of the Human Visual System - Physiological view*

Even at this moment, when one is reading this page, the light that is reflected from this page is focused by the lens of the eyes to form retinal images [18]. Light reaching the retina must pass through all other layers of the retina tissues before reaching the light sensitive photoreceptors. The fovea, a small dip in the retina about 1 mm away from the posterior pole of the eye and near the centre of the retina, has the highest concentration of photoreceptors that are exposed to light. Once illuminated, these photosensitive cells response by converting the light energy into electro-chemical signals. These signals are further processed by multiple retinal connections before being transported through the visual pathway via the optic nerve, the axons of the ganglion cells. The retina ganglion cells then send their signals to the lateral geniculate nucleus (LGN), a part of the thalamus in the midbrain, where further synaptic connections are formed from the LGN to neurons that project to the primary visual cortex (V1 region) in the occipital lobe of the cerebral cortex. The visual signals are then processed by the brain to produce visual perception of object structures, location, motion, colours, etc. Hence the human visual system (HVS) (see Figure 2.1) can be seen in 3 parts: the eyes (the window to the outside world), the visual pathway (the linkway where an image is conveyed and processed) and the final destination – the visual cortex of the brain (where images are perceived by the individual).

Being one of the most sophisticated and intricate system of the body, it is impossible due to limitations in technology and ethical issues to fully unravel the mystery of the functional processes of the HVS. Instead much of the theories postulated concerning the HVS are based on empirical studies on primates, felines and other animals, psychological studies of the HVS or even educated guesses [17, 18, 33-36].

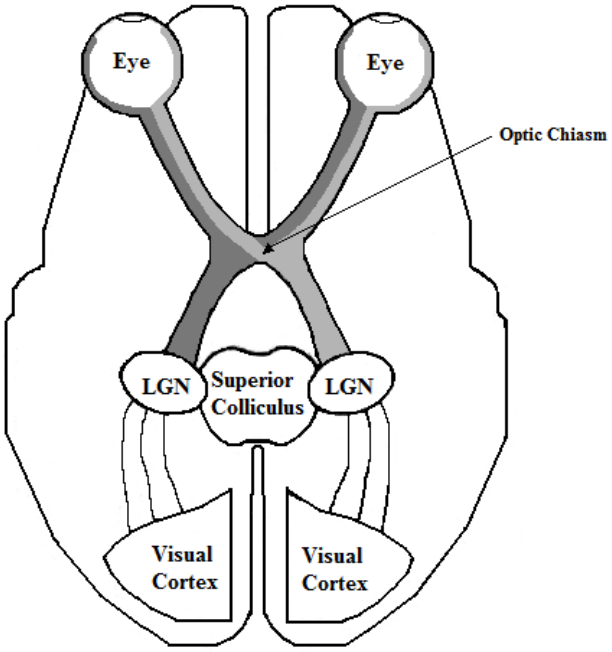



Figure 2.1 Visual pathways: retina to cortex.
(Adapted from Forrester et al. [37])

### 2.1.1 The Human Eye

***a. The physical structure of the Human Eye***

Light enters the eye through the cornea, a thin transparent film which acts as a protective barrier for the inner eye from the external world. It also acts as a refractive surface of the eye whereby external light source is refracted toward and away from the lens. Eventually an image representing the external world is formed at the retina as an inverted retina image on the fovea. The cornea provides two-thirds of the eyes' refractive power [38].

In Figure 2.2, the area between the cornea and the lens is the anterior chamber which is filled with a liquid substance called aqueous humour. The aqueous humour provides nutrients to the cornea, iris and lens. In addition, it keeps the eyeball rigid by maintaining interior pressure at around 10 to 20 mm Hg [38].

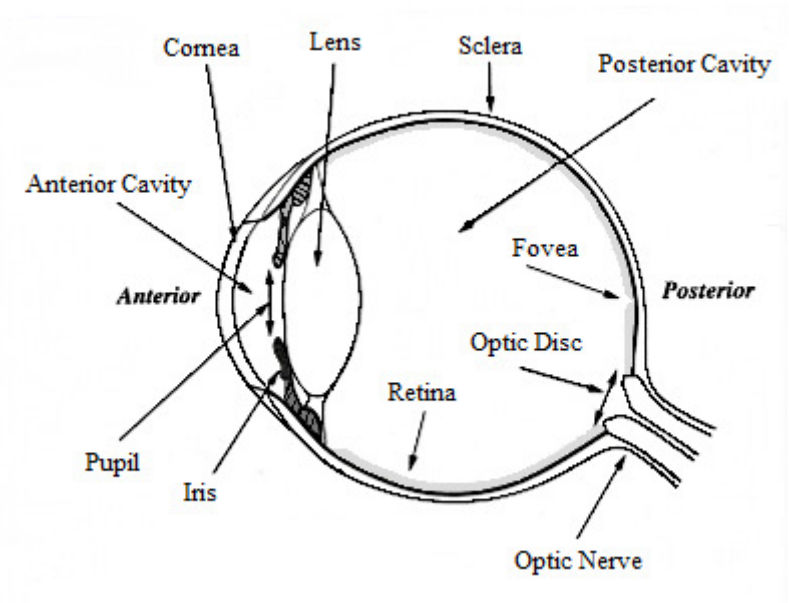



Figure 2.2 A generalized cross section of a human eye.
(Adapted from Malacara [39])

The iris forms an aperture in front of the lens. At its centre is a circular opening called the pupil. Though, the iris can dilate or constrict the pupil to as little as 1 mm diameter, it normally functions in the range of 3 to 7 mm as the adjustment depends on the prevailing light level and influences of the autonomic nervous responses [38]. The dilation and constriction of the pupil size control the exposure area of the lens to external light. This mechanism can change this area by as much as a factor of 5. A smaller pupil size has the effect of restricting the amount of light onto the lens to the peripheral region of the retina, and hence reduces spherical aberration and peripheral blurring [40, 41]. Spherical aberration occurs due to different focal length variations between the fovea and peripheral parts of the retina while chromatic aberration occurs due to different focal lengths for light of different wavelengths [38]. However, reducing the pupil size reduces the amount of light reaching the retina and causes more diffraction, and hence blurring as well. The pupil is automatically adjusted according to light intensity to minimise the blurring effect. The iris which regulates the pupil size thus helps to control the overall sharpness of the images formed at the retina. T lens, suspended at the circular ciliary muscle, is made up of ribbon-like fibres arranged in concentric laminae. Unlike the cornea which has a constant refractive power, the refractive power of the lens varies. It changes through a process called accommodation. Accommodation is controlled by the ciliary muscle, causing

the anterior surface of the lens to either bulge forward or backward, thereby increasing or decreasing respectively the optical power of the lens. The purpose of accommodation is to focus the image onto the retina. The lens focus objects at a distance from about 6.5 metres down to about 10 centimeters. Containing yellow pigments, the lens can also absorb light at ultraviolet region near the wavelength of 365 nm. Hence ultraviolet radiation is usually invisible to the human visual perception [38].

The interior area between the lens and retina is occupied by the vitreous body (vitreous humour). The liquid filled vitreous humour maintains the structural integrity of the eye by ensuring sufficient pressure is maintained to prevent the collapse of the cavity wall. The content of the liquid and its concentration is similar to that of the aqueous humour, and hence both have the same refractive power. The cavity wall contains its neural structures and composes of three layers, the sclera, choroid and the retina. For this thesis, the point of interest is the retina which will be discussed in greater details in the next section.

***b. Retina***

The retina is part of the central nervous system. It consists of five main groups of neural cells arranged into three cellular layers and two synaptic layers. The innermost layer contains light sensitive photoreceptors called rods and cones, named according to their physical appearances. (Refer to Fig. 2.4) [18]. Each retina has about 100-120 million rods and 7-8 million cones [37, 42]. The rods are sensitive to light at low level of illumination and are responsible for scotopic vision (e.g. “night” vision). On the other hand, being less sensitive than rods, the cones are responsible for colour vision (photopic vision) at high level of illumination. According to Forrester et al. [37], both the rods and the cones are sensitive to light with wavelengths from about 400nm to 700nm, with the rods having peak sensitivity at about 498nm. The cones have bandpass spectral response characteristics. There are three types of cones with three different photopigments to absorb different wavelengths of light to different degrees. The three types of cones, being sensitive to lights of short, medium and long wavelengths, are respectively labelled as S (or “blue”) cones, M (or “green”) cones and L (or “red”) cones. The sensitivities of these cones cover the entire visible

spectrum of the human eye, with peak sensitivities at 420nm for "blue" cones, 534nm for "green" cones, and 564 nm for "red" cones. It has been found that the S cones have different spectral sensitivity than the L and M cones that share similar spectral sensitivities.

The strength of the cone's response is proportional to the amount of light energy absorbed by its pigment [18]. The perceptual quality of colour relates roughly to the wavelength's physical properties, i.e., colour as perceived in our nervous system is the result of the differing profile of responses of each type of cone [18]. Red colour is an example of increased activity in the long wavelength cones coupled with minimum activity in the small and medium wavelength cones (see Fig 2.3).

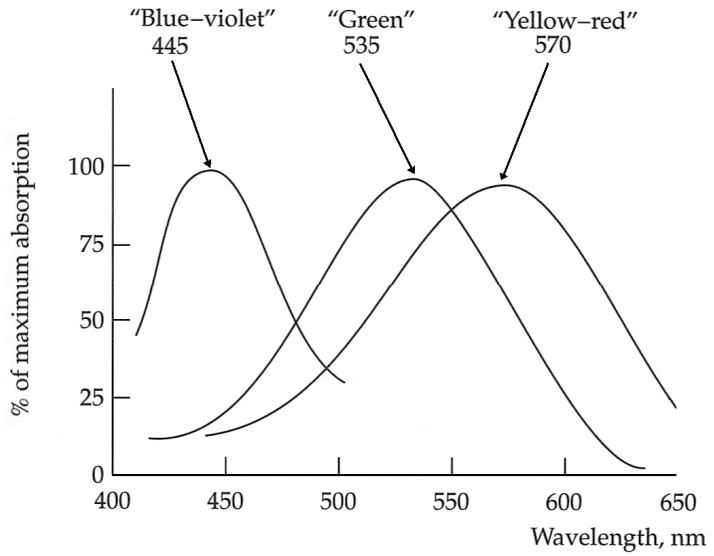



Figure 2.3 Absorption spectra of the three types of cones.
(Adapted from Farah [18])

Apart from the nasal retina where the optic disc (the blind spot where no rods and cones are present) resides, the density and distribution of rods and cones are not uniform throughout the surface of the retina. At the fovea, the cones density is the highest but without any presence of rods. With increasing eccentricity from the fovea, the cones density decreases in an exponential manner until it reaches a constant low level at about 20 degrees from the fovea, while the rods concentration increases until it reaches a maximum level at about 20 degrees from the fovea. Thereafter, the rods concentration decreases to a minimum at about 75 degrees from the fovea [37]. It is clear that the eyes are focused in a manner so that the retina image of any object is

always formed at the fovea where the concentration of cones is highest, and hence sharpest vision and colour discrimination is possible. Away from the fovea, the rest of the retina is responsible for peripheral vision. However, at a very low level of illumination, the image formation at the fovea region does not ensure high visual acuity because of the absence of the rods and insensitivities of the cones at low levels of illumination.

A closer observation of the structures of the photoreceptors and the optic nerve reveals that some form of signal processing does occur before visual information is transmitted to the visual pathway. Each photoreceptor, rod or cone, is composed of an outer segment, a narrow neck, an inner segment, a cell body, and a synaptic base (see Figure 2.4). The outer segment contains photopigments. For the cones, there are 3 pigments that have maximum absorptions for blue, green and red. Photo-chemical reaction to light illumination takes place at the outer segment to produce generator potential. The retina are organised into two synaptic layers, i.e., the outer and inner plexiform layers, which provide both direct and lateral interconnections from the photoreceptor to ganglion cell. The outer plexiform layer consists of horizontal and bipolar cells. One bipolar cell forms a synapse to multiple rods. In contrast, only one cone makes multiple synapses to a bipolar cell. The horizontal cells in the outer plexiform layer provide lateral interconnections between photoreceptors. The second layer consists of amacrine and ganglion cells. The bipolar cells in the outer layer are synapsed to the ganglion cells in the inner layers, while the amacrine cells provide lateral interconnections between the bipolar cells. The synapse of multiple rods to a single bipolar cell increases the sensitivity of photonic energy since any response of any connected rod would activate the bipolar cell. However, less visual acuity is evident as it is less likely to precisely identify between the responses of more than one connected receptors. Hence the rods are more sensitive to low level illumination but less sensitive to discriminate sharper details, while the converse is true for the cones. In the inner plexiform layer, the axions of the ganglion cells extend to form the fibers of the optic nerve.

The differing photosensitive chemicals as well as differing patterns of connectivity to other cells in subsequent layers give rise to the differing functions of rods and cones. Farah [18] postulated a trade-off between sensitivity to light and spatial resolution.

Amazingly, the HVS multiplexes an image into two channels: one that favours sensitivity and one that favours resolution. Hence, the rods with higher sensitivity and convergence onto bipolar collector and ganglion cells give us a low resolution image when there is little light. Conversely, the cones, due to their lower convergence, provide us a high resolution image in the presence of good lighting [18]. Moreover, since colour relies on the cones, which trades off resolution for sensitivity, there is the phenomenon of achromative vision blindness that may occur when lighting is poor.

### 2.1.2 The Visual Pathways

As mentioned earlier, the visual pathway is the linkway that conveys information from the eye to the visual cortex. The bundle of axons connecting the retina to the visual pathway, also known as the optic nerve, splits into numerous pathways [18], of which only two are crucial to visual perception. The first is the geniculostriaye pathway, consisting of the LGN and the primary visual cortex. The other is collicular pathway, which affects spatial orienting and eye movement. In this thesis, only the geniculostriaye pathway will be discussed as it is the most dominant pathway of the HVS [18].

***a. Retinal Ganglion Cells – Center surround Receptive Fields***

The concept of center surround receptive fields was used by Kuffler [43] to describe the interactions of neuron within the visual systems of mammals. Before an image leaves the eye, absolute levels of illumination are laundered off, leaving a retinotopic map of differences: points in the visual field where an illuminated region abuts a dark region. At the individual retinal ganglion cell level, this is represented as the center-surround organisation of its receptive fields (See Figure 2.5) [18].

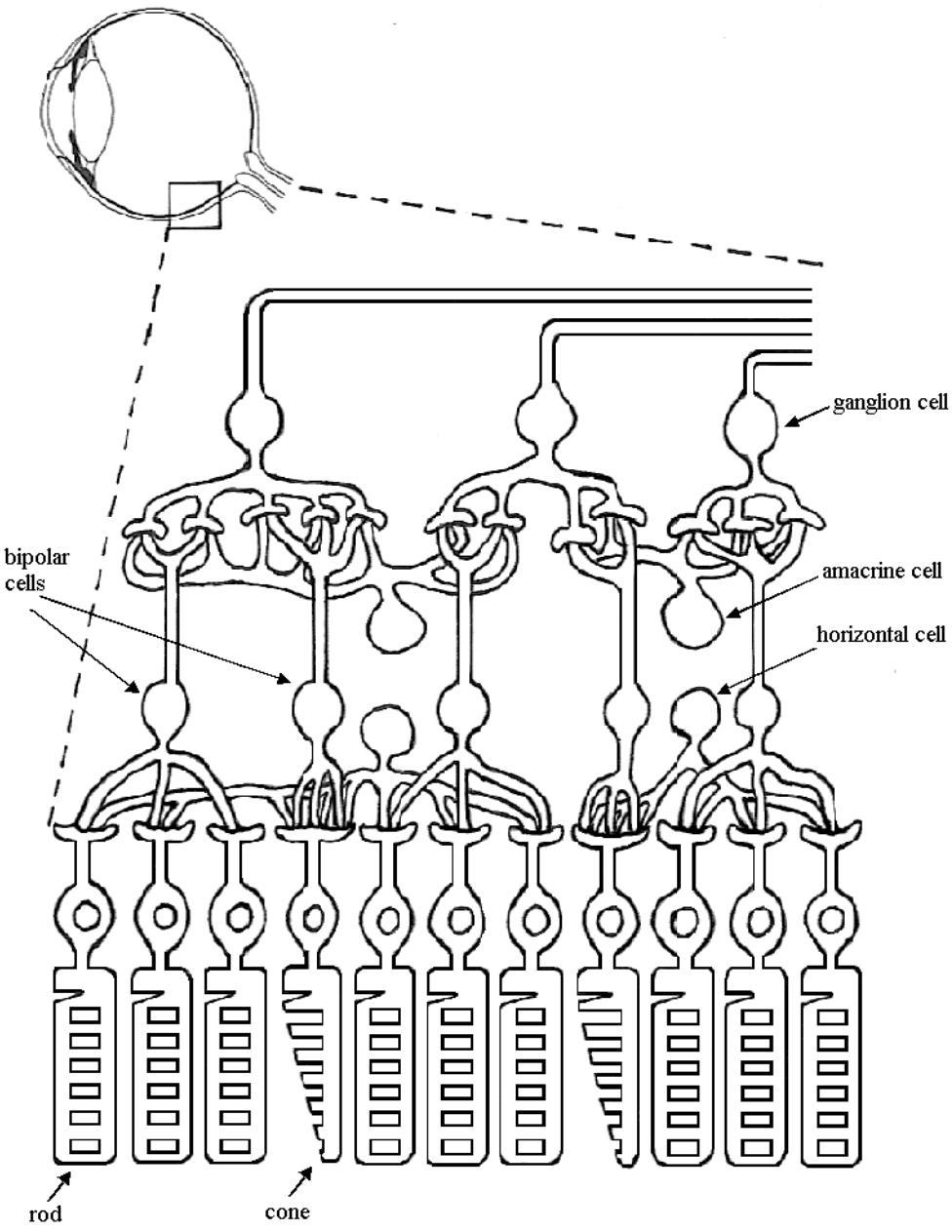



Figure 2.4 Cross-section through the retina
(Adapted from Farah [18])

The human retinal ganglion cells comprise of three distinct classes that are known as X, Y, W cells [44, 45]. These cells are of different sizes. Both X and Y cells project to the dorsal lateral geniculate nucleus and the pretectum. The W ganglion cells project to the superior colliculus and the pretectum. It is also known that the X cells have slower conduction velocities than the Y type cells, with the W cells having the lowest of the three. It is believed that both X and Y cells contribute to high vision discrimination. X cells are more likely to be responsible for resolving higher spatial

frequencies, while the Y cells are more responsive to moving stimuli. The X ganglion cells are concerned with central vision [46].

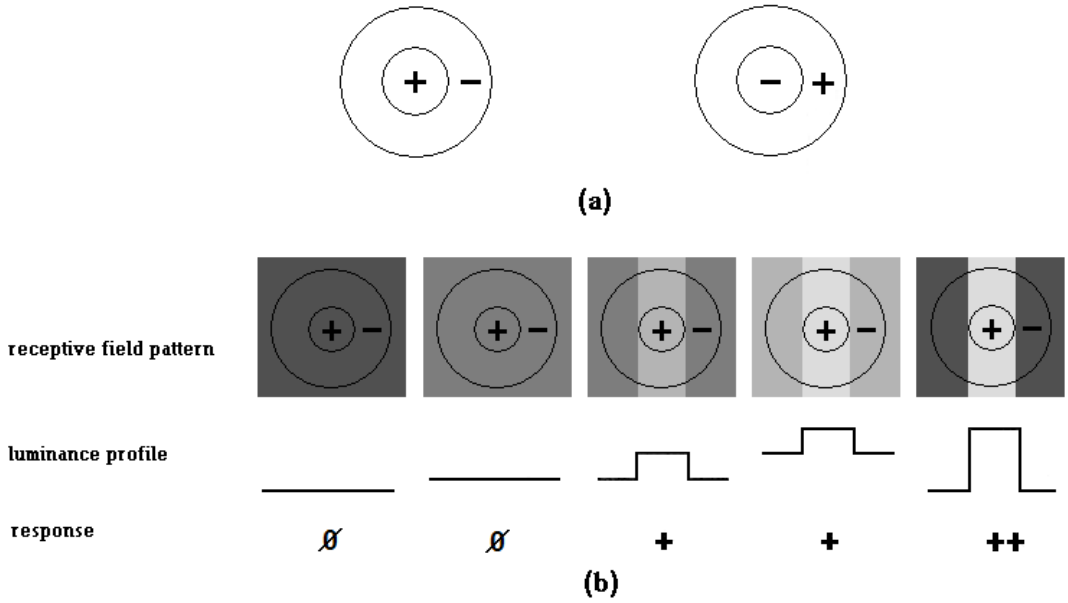



Figure 2.5 (a) Depiction of on-centre/off-surround (left) and off centre/on-surround (right) receptive field structures; (b) Contrast processing of receptive fields (Adapted from Farah [18] )

As stated previously, the photoreceptors in the retina transform light energy into electrical impulses from the ganglion cells. These electrical impulses can be determined by using microelectrodes [30, 34] which measures the response as active potentials or spikes over a time period, when the receptors are subjected to a stimulus. The results showed that the spontaneous firing rate or average rate of occurring spikes increases when a neuron is subjected to a spot of light. However, when the spot of light shifts to the surrounding region, the spontaneous firing rates diminish [34, 38]. Referring to Figure 2.5, the "on-center" cells are stimulated by light in a small area throughout the visual field (on- center) while inhibited by light in the surrounding areas (off- surround). Conversely, the "off-center" cells works in the opposite way [18, 47]. Hence, in the eventual visual perception of objects, it is not the level of absolute brightness, but the differences in brightness between central and the surrounding regions of receptive fields that matter. In Figure 2.5(b), the greater difference in brightness on the right hand side of the on-centre/off-surround receptive field results in higher response (++) than the 'no' response ($\Phi$) of the left hand side on-centre/off-surround receptive field pattern which has the same absolute brightness

in both the on-centre/off-surround regions. In the same way, the perception of colour images is also based on the groundwork of the output of the on-off receptors cells of the various cone types [18].

***b. The Lateral Geniculate Nucleus (LGN)***

The Lateral Geniculate Nucleus (LGN) consists of six layers - four parvocellular (P cells) layers visible at the top and two layers of magnocellular cells (M cells) visible from the bottom [17, 18]. Compared to the P cells, the M cells are larger and have broader axons, resulting in a faster nerve conduction velocity and more transient response. However, in terms of colour perception, the P cells exhibit colour sensitivity while the M cells do not. Moreover, the M cells receive input from a greater number of photoreceptors, giving rise to greater light sensitivity or in other words, better temporal resolution. On the other hand, the P cells, receive input from a smaller number of receptors, producing better spatial resolution. [18]. The temporal resolution of the M cells creates the perception of motion and redirects spatial attention to any unexpected stimulus (e.g., tracking), while the spatial resolution, colour sensitivity and pattern detection of the P cells caters for object recognition where pattern, colour and texture are dominant characteristics [18, 34]. Experiments carried out on primates have also shown the above characteristics of the M and P cells. In the experiments, sections of the monkeys' LGN layers were lesioned with ibotenic acid to create impairment in the M or P cellular layers. The primates are then subjected to psychophysical test to map their impaired and preserved visual perceptual abilities [48]. Recent Studies has also indicated the presence of another separate layer, the Koniocellular layer [49], which exhibited similar behaviour to the P cells. The Koniocellular layer bypasses the primary visual cortex, V1, and instead connects directly to the V2 layer [50]. The functionality of this layer is as yet unknown.

The neurons in the LGN layers exhibit the same center-surround organization as the retinal ganglion cells. Though some researchers think that the cells in the LGN layers have more powerful inhibition towards the surrounding regions [34, 51], there should not be any major distortion of the neural image as it moves from the retina to LGN. Currently, researchers do not fully understand the full function of the LGN though

many concur that it is positioned to amplify visual input to the cortex [51]. This then leads us to our next section where the primary visual cortex is discussed [34, 38] – (the final destination of the visual signal from the retina and LGN) .

### 2.1.3 The Primary Visual Cortex

The optic fibers from the two retinas merge at the optic chiasm where the fibers are separated into two groups that connect to each side of the brain. Here the retinal ganglion cells send images from the left optic field to the right side and of the brain and vice versa. A large part of the visual signal from the retina and LGN is sent to a single area in the occipital lobe of the cortex. This area is called V1 or the primary visual cortex [34]. Other cortical areas have also been identified by researchers over the years, of which V1 through to V5 are most prominent. V4 is generally associated with colour while V5 with motion [18, 34, 37, 52, 53] (See Figure 2.6).

The discussion here shall center on V1 and V4. V1 consists of six layers based on the differing densities of neurons, axons, synapses and interconnectivities with the rest of the brain. According to Livingstone and Hubel [54], layer 4B received signals from the M cells, specializing in the motion and depth perception. Layer 4C continues the parvocellular processing, specializing in colour and shape perception. These two streams then project to different parts of V2 and even possibly project to other higher level of association cortices. However, recent studies have shown that the hypothesized segregation at each level of processing is not always true [18].

V4 is commonly associated with the perception of colour. Perception of colour starts with the absorption of different wavelength light corresponding to the three cone types. The P cells in the retinal ganglion cells with the center-surround field responds to the differing profile of responses towards colour. Colour contrast is further processed and becomes more pronounced in the LGN.

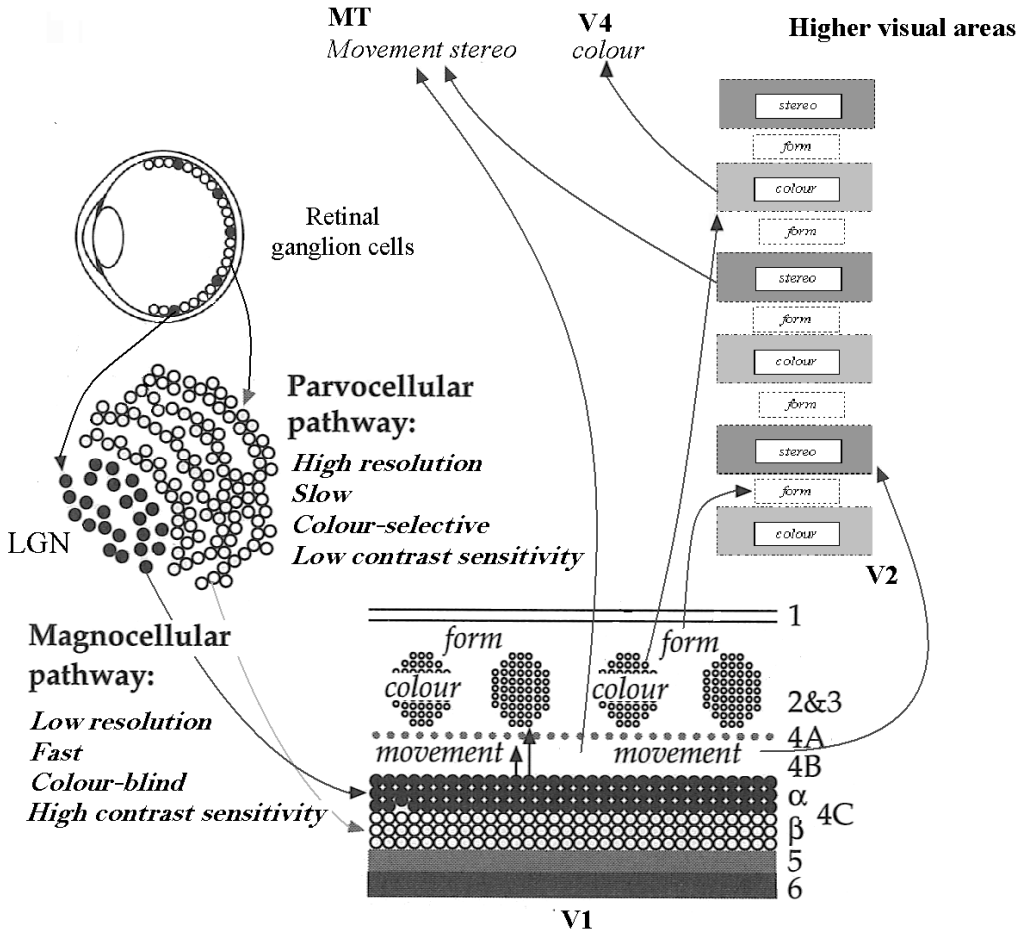



Figure 2.6 Anatomically and physiologically defined subdivisions of the visual system (Adapted from Livingstone and Hubel [54] )

In the primary visual cortex, layers 2 and 3 carry colour information and project it into V2 which in turn is translated to V4. Although many researchers have accepted the hypothesis of V4 being a main player in colour perception or even the colour centre, nothing can be said about the exact nature of V4's role [18]. Thus this gives rise to a hypothesis of the specialization of higher cortical processes in the HVS [18]. Similar to the retina ganglion cells, the cells of the primary cortex exhibits some characteristics - the orientation and frequency selectivity nature of the cells in the primary cortex (discussed in the next section).

### 2.1.4 Characteristics of Neural Responses - Orientation and Frequency Selectivity

***a. Simple, center-surround and complex cells in the primary visual cortex***

When visual signals travel from the LGN through the visual pathway to the primary cortex, there is a major change in the image representation [18]. Hubel and Weisel discovered in 1958 that the receptive fields of the visual cortex cells are different from that of the retina and LGN when they conducted experiments on a cat's eye [55]. Basically the cells in the visual cortex are classified into 3 categories [21]: simple cells, center-surround cells and complex cells.

Within a visual field, simple cells respond to edges at certain specific locations and orientations (see Figure 2.7). The excitatory and inhibitory regions are elongated and thus spots of light or edges at the wrong orientation have little effect on their response levels. As regards to center-surround cells, they response similarly to the retinal ganglion on-off cells (discussed earlier), i.e., specific regions of the visual field either excite or inhibit them [18]. Complex cells, as the name suggests, have responses more complex than the previous two types. Representing more abstract visual information, they are more selective to particular lengths of contour and thus are sometimes called "hypercomplex" or "end-stopped" cells [18]. In fact, Hubel and Weisel [21] suggested that there could be a feed-forward sequential and hierarchical visual processing between the three types of cells (see Figure 2.8). The responses of the cells are specific to the form of stimulus (e.g., from constant luminance to an oriented edge or bar) and the viewing conditions (from a point to a range of location in reference to a fixation). Thus a simple pattern of excitation would channel signals from one level to another, and the simple and center-surround cells would converge on a complex cell, giving rise to object recognition at a higher level of visual processing. From experimental data, Hubel and Weisel found that the stimuli that incite strongest responses from simple and complex cells were oriented edges and bars [21].

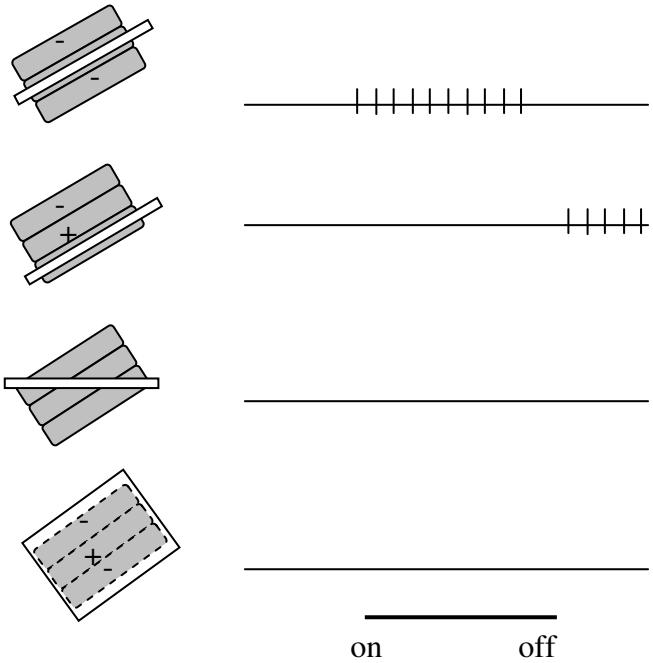



Figure 2.7 Bar stimuli of different orientations (left) and the responses they evoke from a simple cell in primary visual cortex (right). (Adapted from Hubel [56])

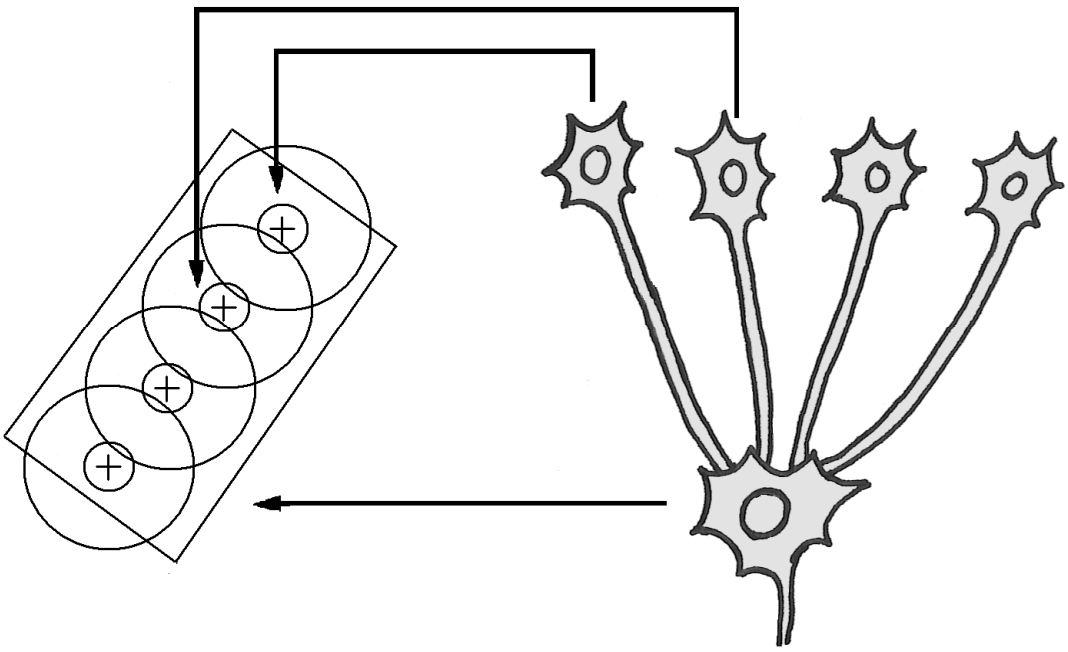


Figure 2.8 Illustration of the idea that simple cells result from the feedforward convergence of a set of centre-surround cells. (Adapted from Hubel and Wiesel [21])

***b. Orientation selectivity***

A visual signal (electrode penetration) which is perpendicular to the cortical layer will attune to cells with the same orientation preference. At each level, there is a column with a particular orientation preference [18, 21]. The orientation preferences of each successive column vary in a smooth and systematic way and are by no means random. Hence, Hubel and Weisel [21] used the term "columns" to portray the organisation of orientation selectivity in the human visual system.

On the psychophysical front, Valois, Yund and Hepler [19] derived quantitative data on the orientation and directional responses of cells in the striate cortex (primary visual cortex of monkeys). Their studies reveal that the orientation bandwidth of cells at half amplitude ranges from 6 to 36 degrees, with a median of 40 degrees. Most cells also show excitations to some particular orientations and inhibitions to other orientations, with maximum inhibitions present side by side of excitatory orientations. Some cells are also found to be isotropic.

***C. Frequency selectivity***

Many psychophysical studies have shown that the "visual system operates in a quasi-linear fashion over a realistic range of contrasts, producing multiple, fairly narrow tuned, spatial frequency channels. (Presumably, cells are selectively sensitive to different restricted portions of the spatial frequency spectrum)." [20]. Thus it can be said that the HVS (up to the region of the primary visual cortex) performs a spatial frequency filtering of the visual information.

## *2.2 Overview of Human Visual system – Psychophysical View*

Visual adaptations include changes over time in the areas of visibility, colour appearance, visual acuity and sensitivity. These changes can be be measured using psychophysical experiments [37]. Therefore, the study of the HVS is not complete without observing the psychophysical aspect. The psychophysical studies and experiments undertaken in the areas of visual acuity, contrast sensitivity and visual masking will be discussed in the following sections.

### 2.2.1 Visual Acuity

When an image is captured by the eye, three factors (i.e., optical filtering, receptor sampling and the receptive organization at the retinal level) determine the clarity of the captured image. Thus visual acuity is the measurement of this clarity [37].

Campbell and Gubish [57] measure the optical quality of the eye by recording the faint light emerging from the eye that was reflected on the fundus. The basic idea behind this is to capture the retinal image. However, due to the problem of the double passage of light (light entering and leaving the eye) and the optical imperfections inherent to the eye, the clarity of an external object is slightly diminished. For example, an infinitesimally, self-luminous object will be degraded to a smooth illuminance distributiontermed as the linespread function (LSF) [57]. Using Fourier transform, the line images were translated to modulation transfer functions (MTF). Results show that the MTF gives rise to a better optical quality estimate. Other studies have also confirmed that for a given pupil size, the retinal image of a thin line is twice as broad as the line's diffracted image [57-60]. Moreover, a further study by Campbell and Gubisch [57] not only shows that the retinal image is a blurred version of the original input image due to imperfections of the human's optic, but it also shows that the linespread function is related to the pupil size, i.e., a larger pupil will give rise to more blurring of the image.

However, as most images do not consist of weighted sums of line, Wandell [34] suggested the use of a set of points as better descriptors for two-dimensional (2-D-image. Thus the use of the point spread function (PSF) [61] is a more general representation for real life images (see Figure 2.9) [34].

The derivation of the MTF either from the LSF or the PSF is an optical transfer function which defines the scale factors applied to each spatial frequency. The MTF is the magnitude of the Fourier Transform of the PSF. Due to difficulty of determining the MTF from PSF, a common approach is to determine the MTF by taking the Fast Fourier Transform (FFT) of the LSF at various angles. In Manos and Sakrison [62],

the MTF of the PSF has been used to measure perception distortion of images. Based on the modulation curves of the HVS, derived through experiments, the MTF could serve as a good estimate of optical sensitivities relative to frequency. According to Mannos and Sakrison [62], the MTF which is an empirical model often used in experiment to fit CSF data is shown as,

$$MTF \approx 2.6(0.0192 + 0.114 f_r)e^{-(0.114 f_r)^{1.1}} \tag{2.1}$$

where $f_r = \sqrt{f_x^2 + f_y^2}$ . $f_x$ and $f_y$ are the horizontal and vertical spatial frequencies, measured in cycles/degrees.

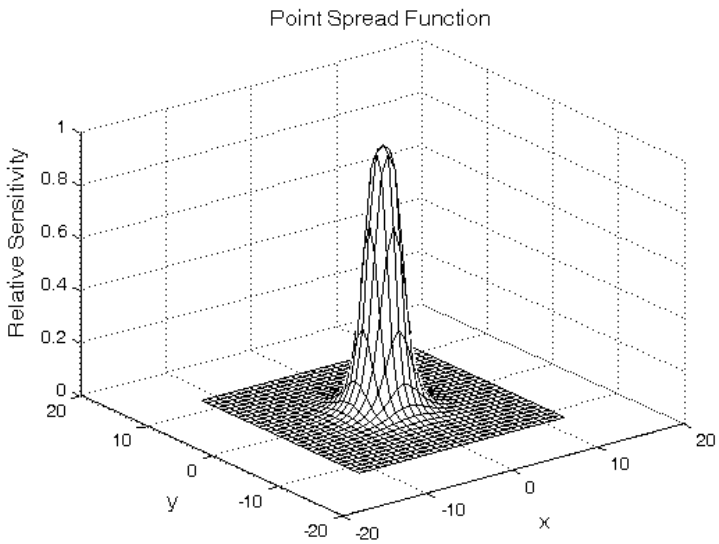



Figure 2.9 Point spread function (Adapted from Wandell [34])

From the characteristics of the MTF (See Figure 2.10), the human optics have a band pass characteristic with a peak sensitivity estimated to be about 8 cycles per degree of visual angle. This sensitivity attenuates rapidly at both the lower and higher frequency band with a cut off frequency at around 50 cycles per degree. This is consistent with the contrast sensitivity function [63, 64]. The low frequency cut-off is due to lateral suppression in the retina ganglion cells. The high frequency cut-off is

due to the MTF of the optics and the integration process of the retina photoreceptive cells (i.e., the cones).

### 2.2.2 Contrast Sensitivity Function

The HVS is able to perceive very minute differences in luminance. Contrast threshold is thus defined as the contrast needed to elicit a visual response in the wake of differences in intensity/luminance. By inversing the contrast threshold, the contrast sensitivity function is obtained [34]. Contrast can be measured at the luminance level and has several forms of expression. Two commonly used definitions are the Weber-Fechner contrast [65] and the Michelson's contrast functions [66].

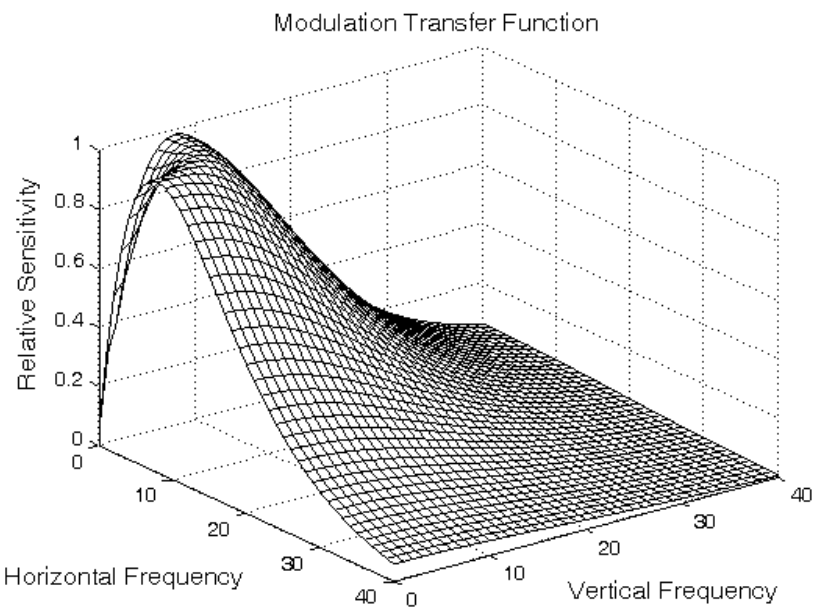



Figure 2.10: Modulation Transfer Function of the Human Eye.
(Based on MTF function of Mannos and Sakrison [62])

Weber's contrast function is derived from a psycho-visual experiment. An observer looks at a stimulus like the one shown in Figure 2.11. The stimulus consists of a constant uniform background with luminance L and a varying patch in the foreground with luminance L + ΔL. As the foreground luminance increase in brightness, the Just

Noticeable Difference (JND) - ΔL/L which is the minimum luminance needed to see the patch, is measured. Thus the Weber's constant function is defined as

$$C_{weber} = \frac{\Delta L}{L} = k \tag{2.2}$$

where $L$ is the background luminance, $k$ is the Weber-Fechner fraction, and the JND is 1-3% for a constant region of $L$ values between 0.1 – 1000 cd/m$^2$

Michelson's contrast is usually used to measure contrast of sinusoidal grating:

$$C_{Michelson} = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \tag{2.3}$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum luminance, respectively.

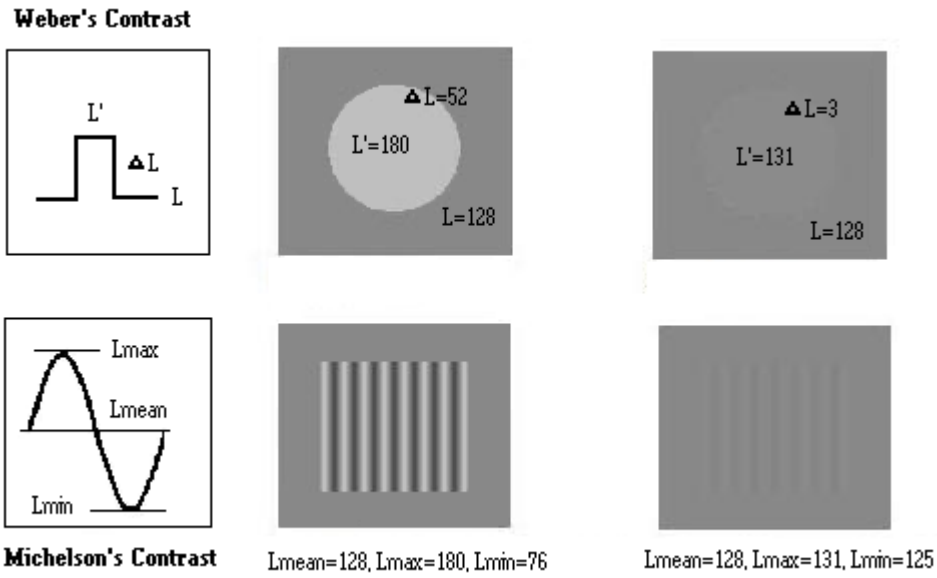



Figure 2.11 Contrast measures of simple patterns

However, both Weber-Fechner and Michelson's contrast functions are designed for simple patterns. As the images in our real world have more complex patterns, these functions have limited effectiveness. In fact, Winkler [67] highlighted that both Weber's and Michelson's functions are affected by changes in luminance extremities

and fluctuations. Note that, as reported by Peli [68], although both definitions of contrast are similar, they are not equivalent and the dynamic range for both are not the same.

Peli provided a definition for contrast for complex images – the band-limited contrast ($C^{blc}$) [68], which defined contrast at any frequency band. The band-limited contrast, $C_i^{blc}$, at any spatial frequency, *i*, is as follows,

$$C_i^{blc}(x, y) = \frac{a_i(x, y)}{l_i(x, y)} \tag{2.4}$$

where $l_i(x, y) > 0$. In the space domain, $a_i(x, y)$ is the bandpass-filtered image, and $l_i(x, y)$ is the low pass filtered version of the image containing all energy at bands below the current scale. In Peli's work [68], a pyramidal structure of 1-octave wide bandpass filter centred at different scales that are 1-octave apart is used. A definition of the bandlimited contrast with the pyramidal structure is included in Appendix H. Interested readers may refer to Peli's work [68] for an extensive coverage.

Contrast sensitivity is a function of spatial frequency, temporal frequency and mean luminance [34]. Van Nes and Bouman described the CSF in two parts: "the optical modulation transfer function responsible for the image formation on the retina, and a retina-perception-center contrast sensitivity function." [64]. The contrast threshold increases according to mean luminance [64]. Since the CSF is the inverse of the contrast threshold, when the mean luminance increases, the contrast sensitivity of high spatial frequency signals decreases (Fig 2.12).

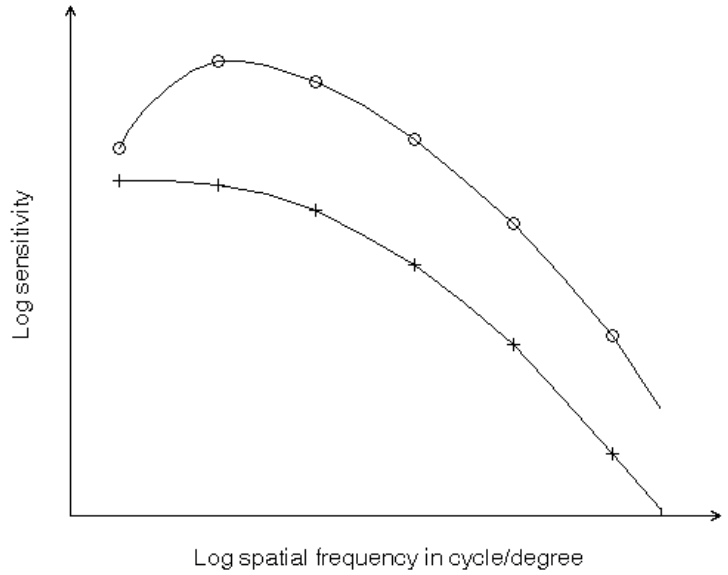



Figure 2.12 Contrast Sensitivity of sine-wave gratings. Cross for lower mean luminance. Circle for higher mean luminance. (Adapted from Wandell [34])

### 2.2.3 Visual Masking

In the presence of other visual stimuli, the strength of a visual stimulus can be either enhanced or diminished. The enhancement or deterioration of the visual stimulus is due to the responses of receptive fields in the visual cortex being triggered either positively (excitation) or negatively (inhibition). The enhancement and deterioration of visual stimulus in this manner is commonly known as facilitation and masking, respectively. In the experiment conducted by Legge and Foley [22] with sinusoidal gratings, the frequency and orientation of the target signal and masker are closely related as to affect the level of facilitation and masking. In Figure 2.13, the target contrast threshold versus masker contrast (TvC) profile, no masking occurs at low masking contrast level (masking contrast below c1). Facilitation occurs between c1 and c2, and masking occurs beyond c2. It has been found in [22] that for high contrast maskers and signals at medium and high spatial frequencies, signal threshold elevation increases when the frequency and orientation of the target signal and masker are similar, and being maximal when both signal and masker have the same frequency. The effect of masking diminishes as the masking frequencies deviate away from the target signal frequency.

The masking model proposed by Legge and Foley includes both low contrast detection and high contrast discrimination in a nonlinear transducer as follows,

$$F(r) = \frac{a_1 |r|^p}{|r|^q + a_2^q} \tag{2.5}$$

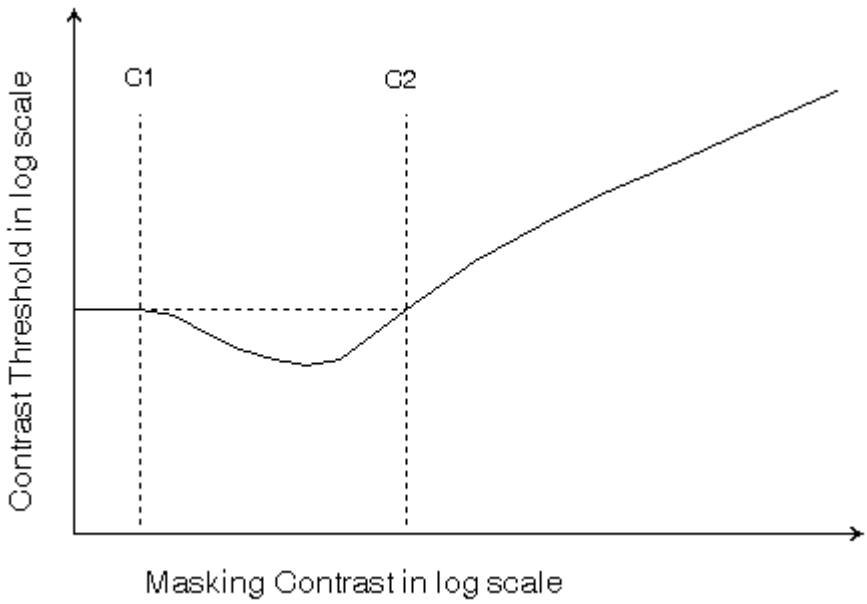



Figure 2.13 Target contrast thresold vs masker contrast (TvC) curve.
No masking is observed to the left of C1. Facilitation occurs between C1 and C2.
Masking occurs to the right of C2. (Adapted from Legge and Foley [22] )

where $r$ is the input signal (signal + masker or signal without masker) to the transducer. It is derived from the output of a presiding linear filter. $a_1$ and $a_2$ are constants. $p$ and $q$ are the exponents for the excitatory and inhibitory terms, respectively, with $p > q$. The exponents $p$ and $q$ are set to 2.4 and 2, respectively,

at low input to account for low contrast, $r > a_2$, $F(r) \approx a_1 |r|^{0.4}$. At high input, which accounts for high contrast, $r < a_2$, $F(r) \approx \frac{a_1 |r|^{2.4}}{a_2^2}$.

The output, $F(r)$, from the transducer is added with Gaussian noise, e, to account for observers giving the same response in identical force-choice trials. The output of the detector is $E(r) = F(r) + e$. The force-choice trials are conducted whereby an observer is presented with one interval containing target signal plus masker, and with another interval containing masker alone.

Essential, the decision rule is based on $E(r_s + r_m) - E(r_m)$, where $r_s$ and $r_m$ are input signals representing target signal and masker, respectively.

***a. Foley's Model***

Based on the work of Legge and Foley [22], Foley [25] conducted experiments to investigate two prediction (1) a change in spatial waveform of the masker causes a left or right shift of the TvC function by a multiplicative constant, and (2) a shift of the TvC function to either left or right by an additive constant in the presence of an additional constant masker. However, tests with Gabor patterns for both the target and the masker did not support the above predictions. Instead, Foley developed two new models incorporating a divisive inhibition that described better fits to observed data than that of Legge and Foley's model [22]. The new models were based on the finding that cells in the visual cortex have both the excitatory and a broadband divisive input. In one of the proposed models, the excitation function, $\boldsymbol{E}$, is the half-wave rectified sum of the individual excitation function, of which the individual excitation function is defined as the product of component contrast, $\boldsymbol{C_i}$, and the sensitivity due to the normalized luminance profile, $\boldsymbol{S_{Ei}}$ of component, $\boldsymbol{i}$, that is,

$$E = \sum_i C_i s_{Ei} \tag{2.6}$$

The contrast component, $\boldsymbol{C_i}$, is defined as

$$C_i = \frac{L_i(x,y)_{\max} - L_o}{L_o} \tag{2.7}$$

where $L_i(x,y)_{\max}$ and $L_o$ are the maximum and average luminance, respectively, for component, $i$ .

The broadband divisive inhibition function, $I$ , is defined as the sum of the product of individual inhibition. The individual inhibition function is defined as the product of the component contrast $C_i$ and sensitivity $S_{Ii}$ for pattern $i$ .

$$I = \sum_i C_i s_{Ii} \tag{2.8}$$

The response function is given by,

$$R = \frac{E^p}{I^q + Z} \tag{2.9}$$

where $p$ and $q$ are constant exponents, with $q = 2$ , and $Z$ is a positive constant parameter to prevent any likelihood of division by zero. In general, $E \neq I$ , $S_{Ei}$ and $S_{Ii}$ , due to excitation and inhibition, respectively, are different, in general.

An elaboration of the above model gives rise to another model that includes components from the same orientation as well as that pooled from different orientation, $j$ , as part of the sum for the division term in the response function. Hence the inhibition becomes,

$$I_j = \max\left(\sum_i C_{ij} s_{Iij} \ , \ 0\right) \tag{2.10}$$

where $i$ is the index for components of the same orientation and $j$ is an index for orientation. The response is defined as

$$R = \frac{E^p}{\sum_j I_j^q + Z} \tag{2.11}$$

The inhibitory input terms are summed together for components with the same orientation, $i$ , as in equation (2.10). For pattern components across different orientations, the input is raised to a power, $q$, before it is summed across different orientations, $j$ . The elaborated model with response function in equation (2.11) resulted in better fit to experimental data than that of equation (2.9).

***b. Teo and Heeger's Model***

Teo and Heeger [23, 69] developed a perceptual distortion measure based on the HVS that fits empirical psychophysical data of spatial masking experiments [70] . The model is closely based on the work of Heeger [71] , in which the neuronal response is the result of an accelerating nonlinear response of a cortical neuron's excitation and suppressed divisively by pooled responses of other cortical neurons.

The model consists of a front-end linear transform, squaring of the transform coefficient, a divisive contrast normalization (similar to that of Legge and Foley [22]) across orientations, and finally a detection stage. The model initially uses the Hexagonal QMF filters [72] for frequency decomposition, creating subbands of 0, 60 and 120 degrees orientations for each resolution level. However, the bandwidths for the 60º and 120º orientations were too wide to provide good fit to data. The frequency transform is subsequently replaced by steerable pyramid transform. The steerable pyramid transform is used to decompose the image into several spatial frequency levels, each of which is further divided into six orientations at 0, 30, 60, 90, 120, and 150 degrees. The neuronal response function takes the form as follows,

$$R_\theta = k_i \frac{(X_\theta)^2}{I_\theta + (\sigma_i)^2} \tag{2.12}$$

where $i \in \{1,2,...,N\}$ denotes the contrast discrimination band with $N = 4$. $X_\theta$ is the transform coefficient at orientation, $\boldsymbol{\theta}$, and $\sigma_i$ is the saturation constant. $k_i$ is the scaling constant. $I_\theta = \sum_\phi (X_\phi)^2$ is the inhibition function, with $\Phi = \{0, 30^0, 60^0, 90^0, 120^0, 150^0\}$ as the orientations. Since each normalized sensor can only discriminate contrast differences for a narrow contrast range, the contrast discrimination level is set to $N$=4 so as to cover the full range of contrasts. With the inclusion of numerator term, $X_\theta$, as part of the $\sum_\phi (X_\phi)^2$, and $\boldsymbol{\sigma_i} > 0$, the range for the response function, $\boldsymbol{R_\theta}$, is $[0, \boldsymbol{k_i})$.

The final detection, $\boldsymbol{D}$, adopts the $\boldsymbol{l}_2$ norm,

$$\boldsymbol{D} = \left\| \boldsymbol{R^\alpha} - \boldsymbol{R^\beta} \right\| \tag{2.13}$$

where $\boldsymbol{R^\alpha}$ and $\boldsymbol{R^\beta}$ are the vectors of normalized responses due to the distorted image ($\alpha$) and the reference image ($\beta$), respectively.

***c. Watson-Solomon's Model***

While Foley's model [25] mainly considers spatial masking localised with individual oriented bands, that is, masking contribution due to components within the same spatial frequencies, but without components from the same spatial but different orientation subbands, Teo and Heeger's model [23, 69] only considers masking contribution from across different oriented subbands, but does not include masking contribution from different spatial frequencies. Considerations of both spatial frequencies as well as across different orientations as pooled candidates in the divisive inhibitory function are necessary to achieve better fit to psycho-physical data. All these considerations are subsequently included in Watson-solomon's model [27] through the contrast gain control (CGC) process. In Watson-solomon's model, the inhibitory function includes multiple channel inputs from spatial, frequency and orientation domains. The input signals of two-dimensional image are filtered according to the contrast sensitivity of the HVS followed by either the cortex

transform or the Gabor Array into frequency domain, creating multiple frequency and orientation subbands.

The neuronal excitation, $\boldsymbol{E}_{(\bar{u},\bar{x},\phi)}$, similar to that of Teo-heeger's model [23, 69]. It is defined as,

$$E_{(\bar{u},\bar{x},\phi)} = t^{p}_{(\bar{u},\bar{x},\phi)} \tag{2.14}$$

where $\boldsymbol{t}_{(\bar{u},\bar{x},\phi)}$ is the transformed coefficient of the input image, obtained by either cortex transform (see Appendix I) or Gabor filtering. $\bar{\boldsymbol{u}} = (\boldsymbol{L}, \boldsymbol{\Theta})$ refers to the subband of frequency $\boldsymbol{L}$ and orientation $\boldsymbol{\Theta}$, $\bar{x}$ the spatial location, $\boldsymbol{\phi}$ the phase, and $p$ the excitation exponent. The phase, $\boldsymbol{\phi}$, refers to the four hypothetical phases (0. 90, 180, 270 degrees) of the individual receptive fields [27].

The inhibitory function, $\boldsymbol{I}$, pools transformed coefficients from within individual frequency subband, across different orientation bands and between different frequency bands. It is computed as a convolution with a pooling kernel $H_{(\bar{u},\bar{x},\phi)}$ as follows,

$$I_{(\bar{u},\bar{x},\phi)} = t^{q}_{(\bar{u},\bar{x},\phi)} * H_{(\bar{u},\bar{x},\phi)} \tag{2.15}$$

where $\boldsymbol{H}_{(\bar{u},\bar{x},\phi)}$ is the pooling kernel, and $q = 2$ is the inhibitory exponent.

The overall response, $r_{(\bar{u},\bar{x},\phi)}$, after pooling is defined as,

$$r_{(\bar{u},\bar{x},\phi)} = \frac{E_{(\bar{u},\bar{x},\phi)}}{b^{q} + I_{(\bar{u},\bar{x},\phi)}} \tag{2.16}$$

where $\boldsymbol{b} > 0$ prevents the response from saturating. In general, $\boldsymbol{p} > \boldsymbol{q}$.

## 2.3 Chapter Summary

This Chapter presents an overview of the human visual system. The physiology of the human eye is discussed in detail. Of particular interest is the process of how an image is transformed from a light image to a neural image by the human visual system (HVS). The three aspects involved in this transformation are discussed in detail namely, the retina (where an image is first captured), the visual pathway (where the image is conveyed and processed through LGN) and the primary cortex (where the image is perceived by the human brain). Some neural cells responsible for image formation in the HVS are frequency and/or orientation selective [19-21]. One particular neural behavior that has direct bearing on visual perception is masking. Some of these properties are important visual characteristics which are taken into account during the development of the perceptual models presented in the later chapters.

The study of the physiological mechanisms of the human eye establishes the basis of visual adaptation. Examples of visual adaptations include changes over time in the areas of visibility, colour appearance, visual acuity and sensitivity. Some of these changes can be observed and quantified with psychophysical experiments [37, 42]. Therefore, the study of the human visual system is incomplete without observing the psychophysical aspect.

The psychophysical studies and experiments undertaken in the areas of visual acuity, contrast sensitivity and visual masking have been discussed in this chapter. The Contrast Gain Control Model by Watson and Solomon [27] is an example of a vision model which attempts to incorporate certain quantifiable properties of the HVS such as contrast sensitivity, frequency and orientation selectivity of neurons, and masking phenomenon. Other models following this approach are also discussed [22, 25, 69, 73, 74]. These models formed the basis of the Perceptual Colour Distortion Measure (PCDM) and Perceptual Post-Filtering (PPF) algorithm developed in chapters 4 and 5.

# Chapter 3 Review of Contemporary Image Coders

## *3.1 Overview of image compression systems*

Digital images or pictures are prevalent in modern day life. However, they require significant storage and transmission bandwidth. For example, a 512×512 resolution colour image with 24-bit per pixels occupies 786,432 bytes. Thus, at a resolution of 1024×1024, the size of the image becomes four times as large. With the increased need for digital storage and the use of images in most applications, image compression then becomes important [12, 75, 76].

There are two approaches to image compression: lossy and lossless. Lossy compression allows for some loss of information during encoding. On the other hand, the lossless compression maintains integrity of information during the encoding process, i.e., the reconstructed image from a lossless compression is identically equal to the original uncompressed image. For lossless compression, statistical redundancies in a given data set are removed.

Given that there are limitations in transmission bandwidths and storage capacity, a higher level of compression ratio is desirable and perhaps necessary in some applications. Inevitably, there is a need to accept a certain amount of distortion (information loss) in order to achieve higher compression as evident in the Rate Distortion (R-D) Function [12], i.e., compression ratio is related to the level of distortion. As the encoding process in the lossy compression is selective, meaning not every single piece of information is encoded, lossy compression can achieve higher compression ratio as opposed to the lossless compression. The general approach for lossy compression is to encode information according to importance, i.e., most important information over less important.

In recent years, another school of thought for image compression (i.e., perceptual coding) [77] has emerged which strives to maintain better perceived image quality (vis a vis that of the lossy compression) whilst achieving a higher compression ratio

(compared to that of the lossless compression). Essentially, "*perceptually*" lossless compression is achieved by removing information that is "*perceptually*" irrelevant to the HVS. Perceptually lossless compression attempts to remove statistical and psychovisual redundancies

The focus of the discussion in this chapter is the review of the various image coders in the literature (sections 3.4 and 3.5). An overview of the information theory which forms the basis for image coding is also provided.

## *3.2 Information Theory*

### 3.2.1 Theory of entropy

Image compression is achieved through the removal of statistical redundancies in the data set. Shannon theory of entropy [10] describes the relationship between data, information and redundancy. All data contains certain amount of information which is measured in bit per pixel (bpp). If data used to describe the information exceeds the entropy, redundancy exists. Given a data set with *n* different symbols of probability of occurrence, $p = \{p_1, p_2, ..., p_n\}$, where $\sum_1^n p_i = 1$, there is a minimum amount of bits required to represent each symbol. This is referred to as self information [10], and is defined as,

$$I_i = -\log_2 p_i \tag{3.1}$$

Hence, symbols with higher probability can be represented with shorter length code words and vice versa. The summation of all self-information in a data set is equal to the entropy, $H$. $H$ and is defined as,

$$H = -\sum_1^n p_i \log_2 p_i \tag{3.2}$$

The entropy for a given input source is the minimum average number of bits required to represent each data sample. When all symbols in a data set have equal probability (i.e., the worst case scenario), $H = -\log_2 \frac{1}{n}$ corresponds to the maximum $H$. The redundancy ($R_d$) in data is defined as,

$$R_d = -\log_2 \frac{1}{n} - \left( -\sum_{1}^{n} p_i \log_2 p_i \right) = \log_2 n + \sum_{1}^{i} p_i \log_2 p_i \qquad (3.3)$$

If no redundancies exist, e.g., random noise, then $R_d$ would have been zero, resulting in $\log_2 n + \sum_{1}^{i} p_i \log_2 p_i = 0$.

Since that for a certain interval of finite length of codes, fixed length coding cannot ensure that all source outcomes are represented efficiently, variable length codes are used [12]. Examples of variable length codes are Prefix Codes [78, 79], Unary Code, Golomb Code [80] , Shannon-Fano Code, Huffman Code [81] and Adaptive Huffman Code [82], Arithmetic Code [83-85]. For most practical implementation of lossless compression, Huffman Coding, Adaptive Huffman Coding, and Arithmetic Coding are widely used . Similarly, examples of fixed-length codes are Run Length Encoding [80], Tunstall Code [86].

While the theoretical coding efficiency is at the entropy, in practice, coding at entropy has never been achieved due to practical limitations of modelling accuracy and coding overhead. However, the entropy bound can be nearly achieved with the use of arithmetic coding to the extent that source statistics can be accurately modeled.

### 3.2.2 Rate distortion theory (R-D)

"The primary goal of lossless compression is to minimize the number of bits required to represent the original samples without any loss of information" [12]. However, there are three reasons why information loss is acceptable: (1) Loss of information is allowed as long as it is not perceptible by the HVS, (2) lossless compression is unable

to provide high compression ratio for many practical applications. Consequently, the existence of compression standards, such as JPEG baseline [11] and JPEG2000-lossy [12] came to being, and (3) in the first place, any digital input to the compression algorithm is itself not a perfect representation of the original image.

Given that small errors or distortion are permitted, lossy compression thus strives to provide a balance between distortion levels versus compression ratio [12].

Consider the case of the mutual information, $I(U;V)$, between two random variables $U$ and $V$, which is defined as:

$$I(U;V) = H(U) - H(U \mid V) \tag{3.4}$$

where the entropy, $H(U) = -\sum_u P_U(u) log_2 P_U(u)$, and the conditional entropy, $H(U \mid V) = -\sum_v P_V(v) \sum_u P_{U|V}(u,v) log_2 P_{U|V}(u,v)$. $P_V(v)$ and $P_U(u)$ are the probabilities of occurrence for $V$ and $U$, respectively. $P_{U|V}(u,v)$ is the joint probability. The mutual information $I(U,V)$ in equation (3.4) becomes,

$$I(U,V) = \sum_u \sum_v P_{V \mid U}(v,u) \cdot P_U(u) log_2 \frac{P_{V \mid U}(v,u)}{P_V(v)} \tag{3.5}$$

In source coding with lossy compression, the loss of information is most notably due to quantisation. Consider a source sample, $X = \{x_1, x_2, ..., x_N\}$, subjected to quantisation process such that $\hat{X} = Q^{-1}(Q(X))$, where $Q(.)$ and $Q^{-1}(.)$ are the quantisation and dequantisation operations, respectively. The distortion measure based on square error between $x_i$ and $\hat{x}_i$ is given as $d(x_i, \hat{x}_i) = (x_i - \hat{x}_i)^2$. The mean square error between $X$ and $\hat{X}$ is computed as:

$$MSE(X, \hat{X}) = \frac{1}{N} \sum_{i=1}^{N} d(x_i, \hat{x}_i) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \tag{3.6}$$

Applying equations (3.4) and (3.5) with square error distortion, $d(x_i, \hat{x}_i)$, to a memoryless source, the rate distortion (R-D) function is obtained by solving the minimization problem as follows:

$$R(D) = \inf_{P_{\hat{X}|X} \subset \left\{ P_{\hat{X}|X} : \sum_x \sum_{\hat{x}} P_{\hat{X}|X}(\hat{x}, x) \cdot P_X(x) \cdot d(x, \hat{x}_i) \le D \right\}} I(X; \hat{X}) \tag{3.7}$$

The discrete case in equation (3.7) can be extended to the general case for continuous function. Typically, the R-D function is a continuous and monotonically decreasing convex function in the interval [0, $D_{max}$] as shown in Figure 3.1. $D_{max}$ is the value of $D$ after which $R(D)=0$. $R(D=0)$ is the rate at which distortion is zero, and in this case for lossless compression. The inverse of R-D function is the distortion rate (D-R) function which sets the theoretical limit on distortion, subject to the constraint of a given coding rate.

For a memoryless source, X, with squared error as distortion measure, Shannon lower bound states that:

$$R(D) \ge h(X) - h(D) \tag{3.8}$$

Where $h(D)$ is the differential entropy of a Gaussian random variable with variance, $D$. Consequently, for the memoryless source where $P_X(x)$ is Gaussian with variance, $\sigma^2$, subject to the constraint, $E\left[(X - \hat{X})^2\right] \le D$, the R-D function is as follows:

$$R(D) = \frac{1}{2} \log_2 \frac{\sigma_x^2}{D} \quad 0 \le D \le \sigma_x^2 \tag{3.9}$$

The function in equation (3.9) has a similar shape as in Figure (3.1). The rate distortion theory essentially shows us that any compression system can only perform within the shaded area in Figure (3.1). For a given distortion $D$, it is the design of a lossy compression system to attempt to operate as close to the R-D curve (i.e.,

reaching the lower bound). Note that in transform based image coding [87, 88], distortions are usually generated as a result of quantisation noise. (This will be discussed further in section 3.3.2).

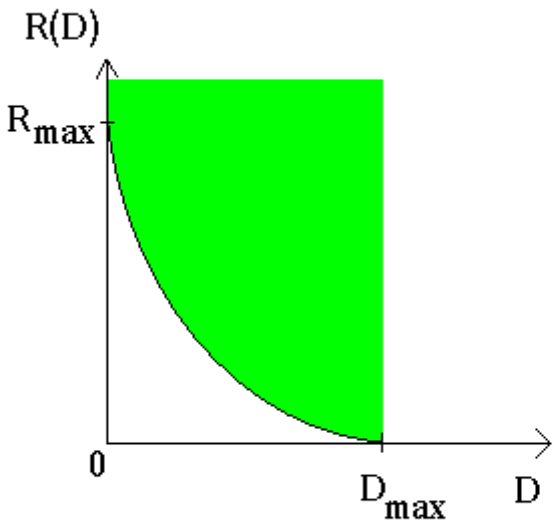



Figure 3.1 A typical rate distortion (R-D) function curve

## 3.3 Elements of an Image Compression System

Figure 3.2 shows the elements in an image compression system. The following sections focus on each of the main elements during the process of image compression.

Pixels of natural images are usually correlated with their neighbouring pixels [12]. The first step in a transformed based image compression system is to project these correlated pixels into a representation so that the sample data are decorrelated [87] with a large quantity of the image energy compacted at a few coefficients (i.e.. DCT transform). The transformed samples are then subjected to a process of quantisation which essentially decreases the precision of the sample data, and thereby reshaping the probability distribution function (PDF) and hence the entropy [89]. Quantised coefficients are then entropy encoded to form the compressed bit-stream.

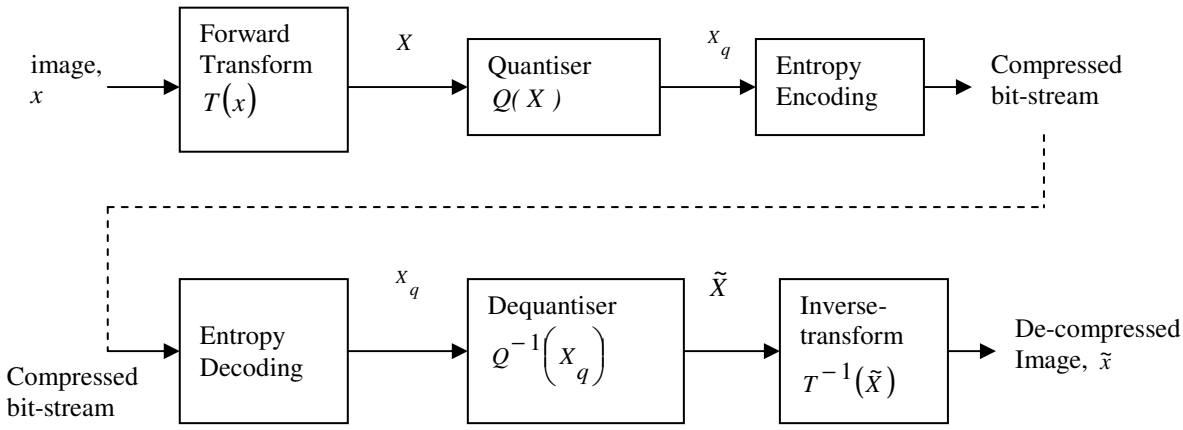



Figure 3.2 Structure of an image compression system

During the de-compression process, the compressed bit-stream is entropy decoded, followed by dequantisation, and then the inverse transform to reconstruct the input image. While quantisation contributes to compression gain, it is also the main contributor to distortion due to quantisation error.

### 3.3.1 Transform

A linear transform ($T(.)$) on an input signal, $x$, and its invertible transform ($T^{-1}(.)$) on the transform coefficients, $\boldsymbol{X}$, can be expressed as,

$$\boldsymbol{X} = T(\boldsymbol{x}) \tag{3.10}$$

$$\boldsymbol{x} = T^{-1}(\boldsymbol{X}) \tag{3.11}$$

In transform based image coding, where a recovery process is required to reconstruct compressed images, it is desirable to have an invertible transform kernel [90], i.e., perfect reconstruction. Both the orthogonal and bi-orthogonal transforms [89-92] are classes of all invertible transform. The perfect reconstruction Quadrature Mirror Filter (QMF) [93] which has been used in both audio and image coding [94] in the literature is also invertible.

From matrix perspective, orthogonal transforms must fulfill the following conditions:

$$\boldsymbol{A} \cdot \boldsymbol{A}^T = \alpha \boldsymbol{I} \tag{3.12}$$

where $\boldsymbol{A}$ is a $M \times M$ square matrix, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{\alpha}$ is a diagonal matrix. Both $\boldsymbol{A}$ and $\boldsymbol{\alpha}$ are of the form,

$$\boldsymbol{A} = \begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1M} \\ a_{21} & a_{22} & \dots & a_{2M} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{M1} & a_{M2} & \dots & a_{MM} \end{Bmatrix}, \tag{3.13}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & 0 & \dots & 0 \\ 0 & \alpha_{22} & \dots & 0 \\ . & . & & . \\ . & . & . & . \\ . & . & & . \\ 0 & 0 & \dots & \alpha_{MM} \end{bmatrix} \tag{3.14}$$

Orthogonality of a transform can also be viewed from vector perspective as inner product of two vectors satisfying the condition below,

$$\left\langle a_i, a_j^T \right\rangle = \sum_{k=1}^{N} a_{ik} \cdot a_{jk}^T = \alpha \delta_{ij} = \begin{cases} 0, & when\ i \neq j \\ \alpha, & when\ i = j \end{cases} \quad for\ \forall i, j, \tag{3.15}$$

where $a_i$ is the row vector of $\boldsymbol{A_m}$ with $i = \{1,2,...,M\}$, $a_j^T$ is the transpose of $a_j$, and $\alpha \in R$. $\boldsymbol{A_m}$ is the square matrix equivalent to equation (3.13).

Matrix $\boldsymbol{A}$ in equation (3.13) is orthonormal if $\alpha_{ii} = 1$. Consequently, the analysis vector ($\boldsymbol{A}$) and the synthesis vector ($\boldsymbol{S} = \boldsymbol{A}^{-1} = \boldsymbol{A}^T$) of orthonormal transforms are equivalent in a sense that the analysis filters are time-reversed, complex conjugate

versions of the synthesis filters, and they are mutually orthogonal with a unit length [12]. That is,

$$\boldsymbol{A} \cdot \boldsymbol{A}^T = \boldsymbol{I} \tag{3.16}$$

or

$$\left\langle a_i, a_j^T \right\rangle = \sum_{k=1}^{N} a_{ik} \cdot a_{jk}^T = \delta_{ij} = \begin{cases} 0, & when\ i \neq j \\ 1, & when\ i = j \end{cases} \quad for\ \forall i, j \tag{3.17}$$

All orthogonal transforms are linear. An important characteristic of an orthogonal transform is the energy preserving property also known as Parseval's relationship [12, 90, 91, 95]. In short, this means,

$$\|\boldsymbol{Ax}\| = (\boldsymbol{Ax})^T \boldsymbol{Ax} = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{Ax} = \boldsymbol{x}^T \boldsymbol{Ix} = \boldsymbol{x}^T \boldsymbol{x} = \|\boldsymbol{x}\| \tag{3.18}$$

where $\boldsymbol{x}$ is the input signal vector in the time domain, the $\boldsymbol{Ax}$ is the transform coefficient vector and $\boldsymbol{A}$ is the orthogonal matrix. Hence, if the MSE in the transform domain is minimised, the MSE of the reconstructed image is also minimised. Examples of well-known orthogonal transforms in the field of image coding include the Discrete Fourier Transform (DFT) [90], Discrete Cosine Transform (DCT) [96], Hadamard Transform, Haar Transform [97], Slant Transform and the Karhunen-Loeve Transform (KLT) [98].

A Biorthogonal transform [90] is invertible, like an orthogonal transform. Specifically, for a non-orthogonal matrix $\boldsymbol{B}$ (i.e., $\boldsymbol{B}^{-1} \neq \boldsymbol{B}^T$), if there exists a dual basis non-orthogonal matrix $\tilde{\boldsymbol{B}}$ (i.e., $\tilde{\boldsymbol{B}}^{-1} \neq \tilde{\boldsymbol{B}}^T$, and $\boldsymbol{B} \neq \tilde{\boldsymbol{B}}$), that satisfies the condition,

$$\boldsymbol{B}\tilde{\boldsymbol{B}}^T = \alpha \boldsymbol{I}\ , \tag{3.19}$$

it is said that matrix $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$ are biorthogonal, where $\alpha \in \Re$. From the vector perspective, vector $\boldsymbol{B}$ and its dual basis $\tilde{\boldsymbol{B}}$, are biorthogonal if,

$$\left\langle b_i, \tilde{b}_j \right\rangle = \alpha\delta_{ij} = \begin{cases} 0, & i \neq j \\ \alpha, & i = j \end{cases}, \tag{3.20}$$

If $\alpha = 1$, matrices $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$ are said to be biorthonormal, and the analysis and synthesis filters are dual basis of each other. Biorthogonal filters do not preserve vector length. Also, Parseval's relation no longer holds for biorthogonal system, therefore, it is important to design a biorthogonal system so that the norms are close [91].

Biorthogonal transform is advantageous over the orthogonal transform with respect to regularity and phase linearity. Regularity is a filter characteristic which measures the degree of filter smoothness under iterations. This means minimum fluctuation, resulting in better reconstructed image. A filter's length affects its regularity and the longer the filter length, the more regular the filter will be. However longer filters increase the computation load of transform [90].

Though regularity is desirable, Rioul [99] argued that excessively regular filters are not needed in image compression since they do not offer significant improvement in the quality of reconstructed images. Since the biorthogonal filters allow for phase linearity, they eliminate phase distortion especially along the sharp edges of images. Though phase misalignment can occur during an orthogonal transform, this problem can also be avoided by using symmetrical filters [100-102].

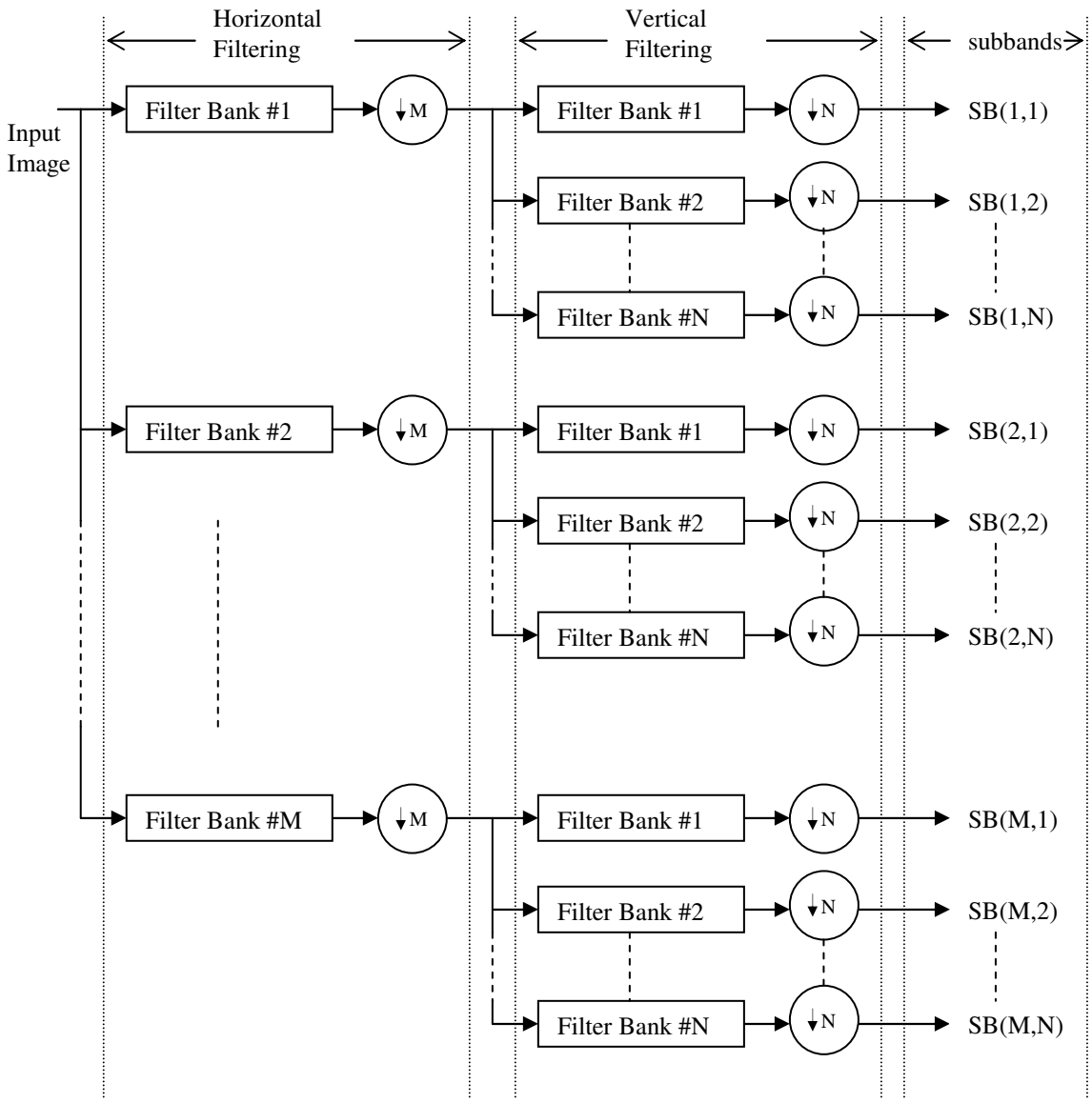



Figure 3.3 Structure of subband coding. The result is a collection of M × N numbers of subbands. ↓N means down sampling by a factor of N.

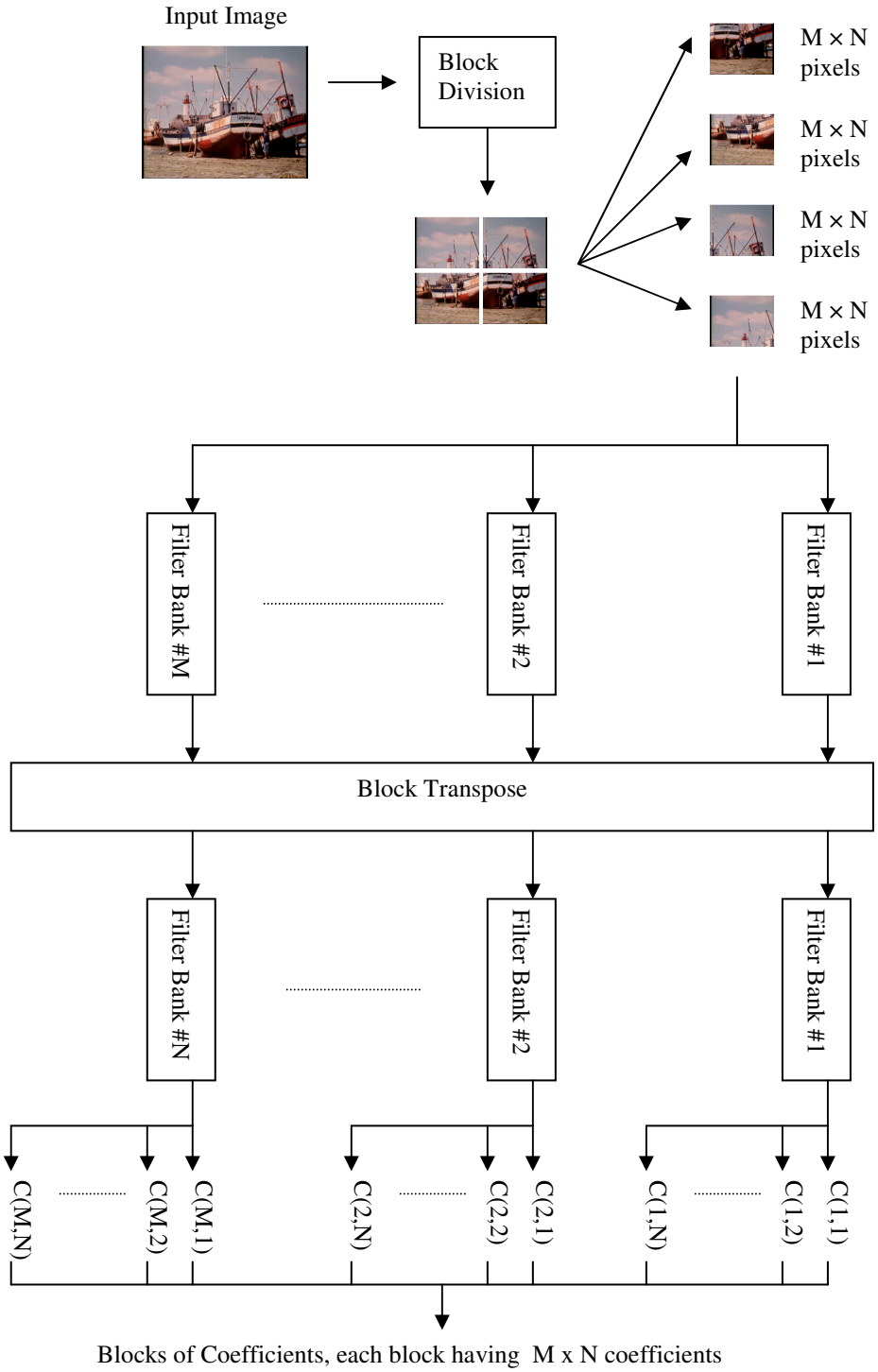



Figure 3.4 Block based decomposition. An input image is sub-divided into blocks of M×N pixels before decomposition takes place. The output is a set of blocks of M×N coefficients.

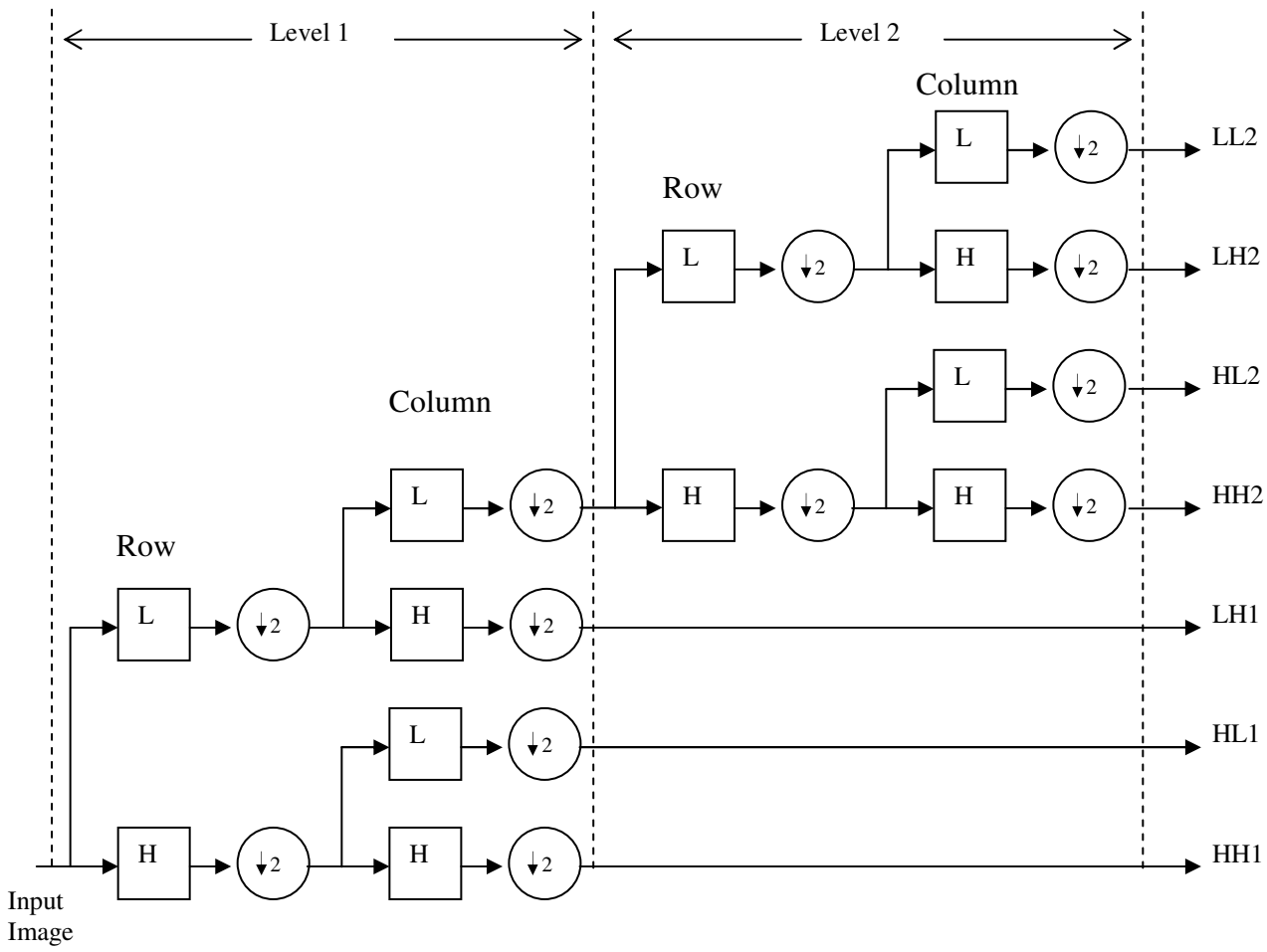



Figure 3.5 Frequency decomposition in multiresolution representation. ↓2 means down sampling by a factor of 2.

The various spectral decomposition structures can be categorised into: subband, block-based, and hierarchical structures. The subband structure organises the spectral coefficients into groups of frequency bands, such that coefficients of the same frequency band are grouped together (See Figure 3.3). For a block-based structure (see Figure 3.4), an image is first divided into blocks of $M \times N$ size, each of which is independently decomposed into spectral coefficients, forming $M \times N$ number of subband coefficients. The hierarchical structure follows the wavelet-based multi-resolution analysis (see Figure 3.5) according to Mallat decomposition [103] .

**a. Block-based Transform**

The Discrete Cosine transform (DCT) Transform [96], the Karhunen-Loeve Transform (KLT) [98] and the Haar transform [87, 104, 105] are the various common block-based transforms. In theory, the KLT is noted for its excellent pixel decorrelation. Though KLT is the optimal transform in terms of energy compaction and decorrelation, it is nevertheless not used in practical applications due to its complex computation. As the KLT Kernel has to be computed for an individual image and transmitted along with the compression stream, calculation of the KLT kernel is slow and cumbersome [1] since there are no fast algorithms. Furthermore, the application of KLT becomes impossible in some situations where the statistics of the source data may not be known in advance, since the optimum transform kernel must be constructed from the statistics of the source data.

In terms of decorrelation and energy compaction, the DCT transform [106] is second only to KLT [12]. With good decorrelation and the availability for fast algorithms, the DCT [106] has been used extensively in picture compression applications such as JPEG [11] and MPEG [107].

The 2-D DCT [1] is defined as,

$$X[k,l] = \sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}} C_k C_l \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} x[i,j]\cos\left[\frac{(2i+1)\pi k}{2M}\right]\cos\left[\frac{(2j+1)\pi l}{2N}\right] \tag{3.21}$$

$$C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0 \\ 1, & f > 0 \end{cases} \tag{3.22}$$

where $0 \le k \le N-1$ and $0 \le l \le M-1$. $x[i,j]$ belongs to the pixel element of an $M \times N$ pixel block, and $[i,j]$ denotes the position of the pixel element in the block. Usually, an image is divided into $k$ blocks of 8×8 pixels [11].

The inverse DCT [1] is defined as,

$$x[i,j]=\sqrt{\frac{2}{M}}\sqrt{\frac{2}{N}}\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}C_k C_l \cdot X[k,l]\cos\left[\frac{(2i+1)\pi k}{2M}\right]\cos\left[\frac{(2j+1)\pi l}{2N}\right] \qquad (3.23)$$

**b. Subband Transform**

The main disadvantage of the block-based transform is that the images are processed in independent blocks. These blocks are seen as disjointed blocks, and thus assumed to be uncorrelated with neighbouring blocks. However, this assumption does not seem plausible as neighbouring pixels may show high correlation. Generally, the correlation decreases when the block size increases [12].

Subband transform uses input from multiple vectors through a sequence of transform matrices known as the filter bank [12]. It filters the source data with a set of $m$ bank analysis filters. For each filter output, only the $m^{th}$ sample is retained through decimation (or down-sampling) [108]. These decimated output values of the $m^{th}$ filter form the $m^{th}$ subband. In the reconstruction stage, coefficients in subbands are up-sampled, then inverse transformed to reconstruct the data [12].

**c. Separable Image Transform**

Multi-dimensional signal processing uses both separable and non-separable filters [109]. In a two dimensional separable transform, the analysis vector is formed by taking the tensor product of one dimensional analysis vectors. Similarly, the synthesis vector is the tensor product of one dimensional synthesis vectors. In separable filtering, input signals can be processed separately in a cascaded manner. Conversely, input signals of non-separable filtering are applied directly in all dimensions. Specifically, consider the case of a separable filter for a two dimensional image arranged in a row and column form, a 1-D subband transform is first performed on the input image along each row to produce an intermediate 2-D array. Then the 1-D transform is applied to each column of this intermediate 2-D array to produce the final

samples. This structure is illustrated in Figure 3.3. Since the implementation and computation of separable filters are less complicated than that of the non-separable filters, separable filters are most widely used in most image coders.

**d. Multiresolution Transform**

A commonly used subband transform is the hierarchical subband transform based on the multiresolution representation of Mallat [103]. Unlike uniform subband transform, this tree-structured transform subjects each successive low frequency band to further decomposition to form a hierarchical resolution chain. Figure 3.5 illustrates a dyadic hierarchical decomposition. A feature of this transform is that a compressed image can be partially reconstructed with higher successive resolutions of the source image. The "resolution scalability" feature in these dyadic decompositions thus makes this transform desirable for image compression applications.

### 3.3.2 Quantisation

Scalar quantisation is most commonly used in lossy compression systems. A scalar quantisation function maps each element, $x_i \in \Re$, on the real line to a particular value within the same subset of data [12]. For a given real number line that is divided into a set of M disjoint intervals, $I = \{I_0, I_{1,} I_2, ..., I_M\}$, with $I_q = \lfloor x_q, x_{q+1})$ and $0 \le q \le M$, the scalar quantisation process maps all real number input values, $x_i \in \Re$, with $x_q \le x_i < x_{q+1}$ and $q$ being the integer-valued quantisation index, into a particular value in $\hat{x}_q$, where $x_q \le \hat{x}_q < x_{q+1}$. Hence scalar quantisation is a many-to-one mapping. Specifically, the quantisation maps all the values in the $M$ disjoint intervals, $I = \{I_0, I_{1,} I_2, ..., I_M\}$, with $I_q = \lfloor x_q, x_{q+1})$, into a subset of $M$ single-real-valued numbers, $\hat{x}_r = \{\hat{x}_0, \hat{x}_1, \hat{x}_2, ..., \hat{x}_M\}$. In practice, the quantisation index, $q$, is being transmitted after a scalar quantiser is applied. At the receiving end, an inverse quantiser is applied to $q$ to produce the output, $\hat{x}_q$. If we denote $Q$ and $Q^{-1}$ as the operators for the uniform linear quantiser and inverse quantiser, respectively, then,

$$q = Q(x_i) = sign(x_i)\left\lfloor \frac{|x_i|}{s} \right\rfloor \tag{3.24}$$

and

$$\hat{x}_q = Q^{-1}(q) \tag{3.25}$$

where $s$ is the quantisation step size. A scalar quantisers can be classified as either uniform or non-uniform quantiser, and mid-rise or mid-thread quantiser [110] as shown in Figure 3.6. Due to the many-to-one mapping of quantisation process, both the input value, $x_i$ and its output value, $\hat{x}_q$ are not equal in general. The error between the input and the output values, $e_i = x_i - \hat{x}_q$, is known as the quantisation error.

Quantisation contributes significantly to the actual compression of data by decreasing the precision of the input data, leading to a reshaping PDF which alters the entropy [89]. The distortion due to quantisation is commonly computed by either the mean-squared-error (MSE) or mean-absolute-error (MAE) metrics. For a set of $N$ sample input data ( $X$ ) and its quantised output values ( $\hat{X}$ ), the MSE and MAE are defined as,

$$MSE = \frac{1}{N}\sum_{u=0}^{N-1}\left(X(u) - \hat{X}(u)\right)^2 \tag{3.26}$$

and

$$MAE = \frac{1}{N}\sum_{u=0}^{N-1}\left|X(u) - \hat{X}(u)\right| \tag{3.27}$$

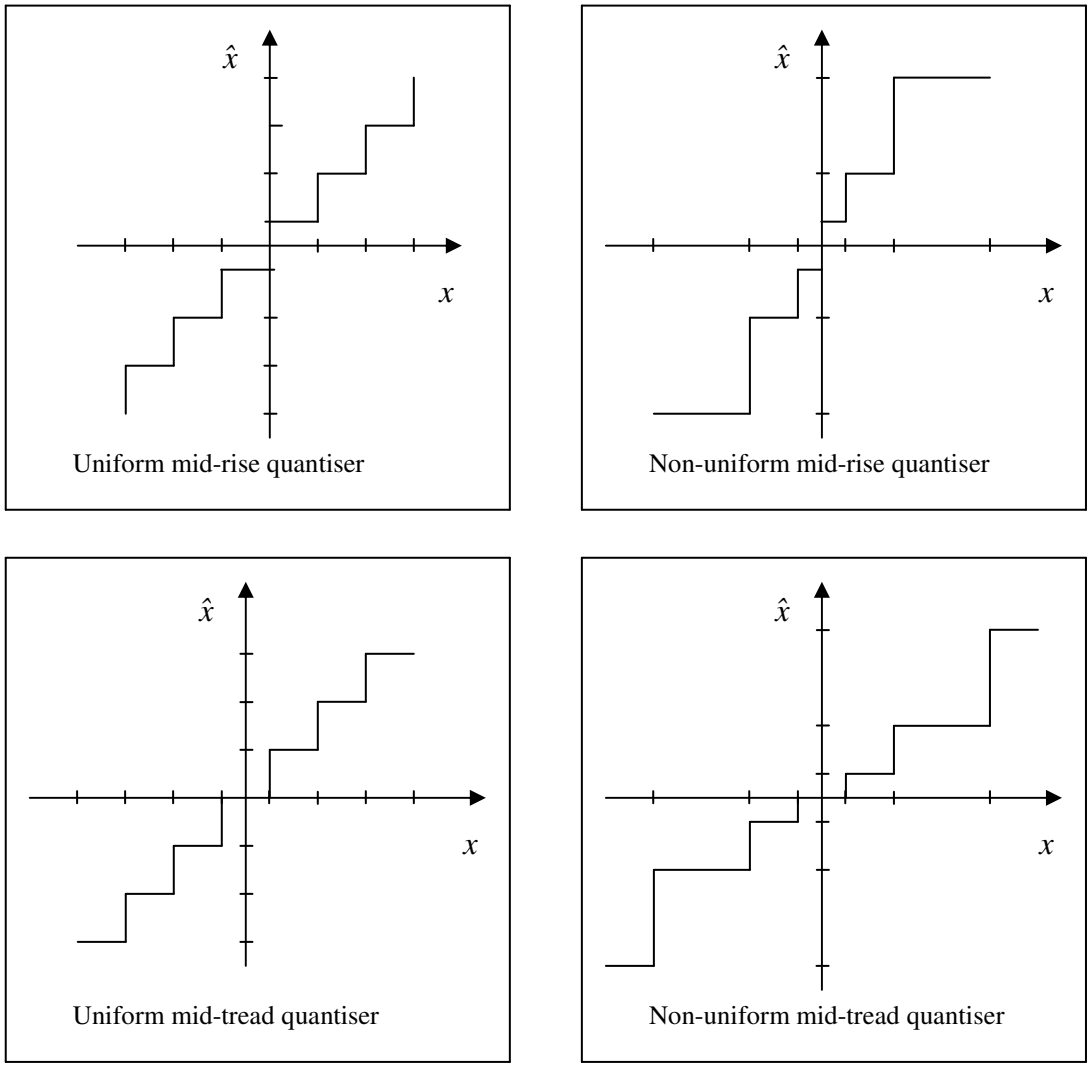



Figure 3.6 Different classification of quantisers

The Lloyd-Max quantiser [111, 112] provides an optimal scalar quantiser when the probability density function of the source input is known. Under the optimal conditions, the Lloyd-Max quantiser minimises the MSE [12, 111]. The other commonly used quantisation scheme for image compression is vector quantisation (VQ) [113], whereby, it selects a codeword, $c_i = \{\hat{x}_1^i, \hat{x}_2^i, ..., \hat{x}_m^i\}$, from a codebook, $C = \{c_1, c_2, ..., c_n\}$, such that, the selected codeword gives the best approximation to the vector of input data, $\vec{x} = \{x_1, x_2, ..., x_m\}$. The key to VQ lies in the vector codebook. Therefore, optimizing this codebook will lead to error minimisation – a process that can be accomplished by the Linde-Buzo-Gray algorithm [113] . A

detailed coverage of VQ can be found in [114]. It is noted that the scalar quantisation is a special case of vector quantisation with vector length equals 1.

### 3.3.3 Bitplane Coding and Bitplane Quantisation

Bitplane coding [115] is an approach for encoding bit layers of data, starting from most significant bit layer to the least significant bit layer, in a progressive manner. Each coefficient is represented in a series of binary digits, starting from the most significant bit (MSB) to the least significant bit (LSB). When all the data set are represented in their binary representation, they collectively form layers of bitplanes, starting from most significant bitplane (MSBP) to the least significant bitplane (LSBP) as shown in Figure 3.7.

For instance, with a block of coefficients, $X$, and hence $X_M$ being the magnitude portion of the coefficients of $X$ arranged in a row and a column format as follows,

$$X_M = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,N} \\ x_{2,1} & x_{2,2} & \dots & x_{2,N} \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix} \tag{3.28}$$

where $x_{u,v}$ is the coefficient in location $(u,v)$ in $X_M$. If all elements in $X_M$ can be sufficiently represented by $k$-binary bits, there will be $k$ binary bitplanes for $X_M$ starting from the MSBP ($p=(k-1)^{th}$) to the LSBP ($p$=0), and hence $X_M$ can be arranged in bitplane layers as,

$$X_M = \{X_{k-1}, X_{k-2}, ..., X_2, X_1, X_0\} \tag{3.29}$$

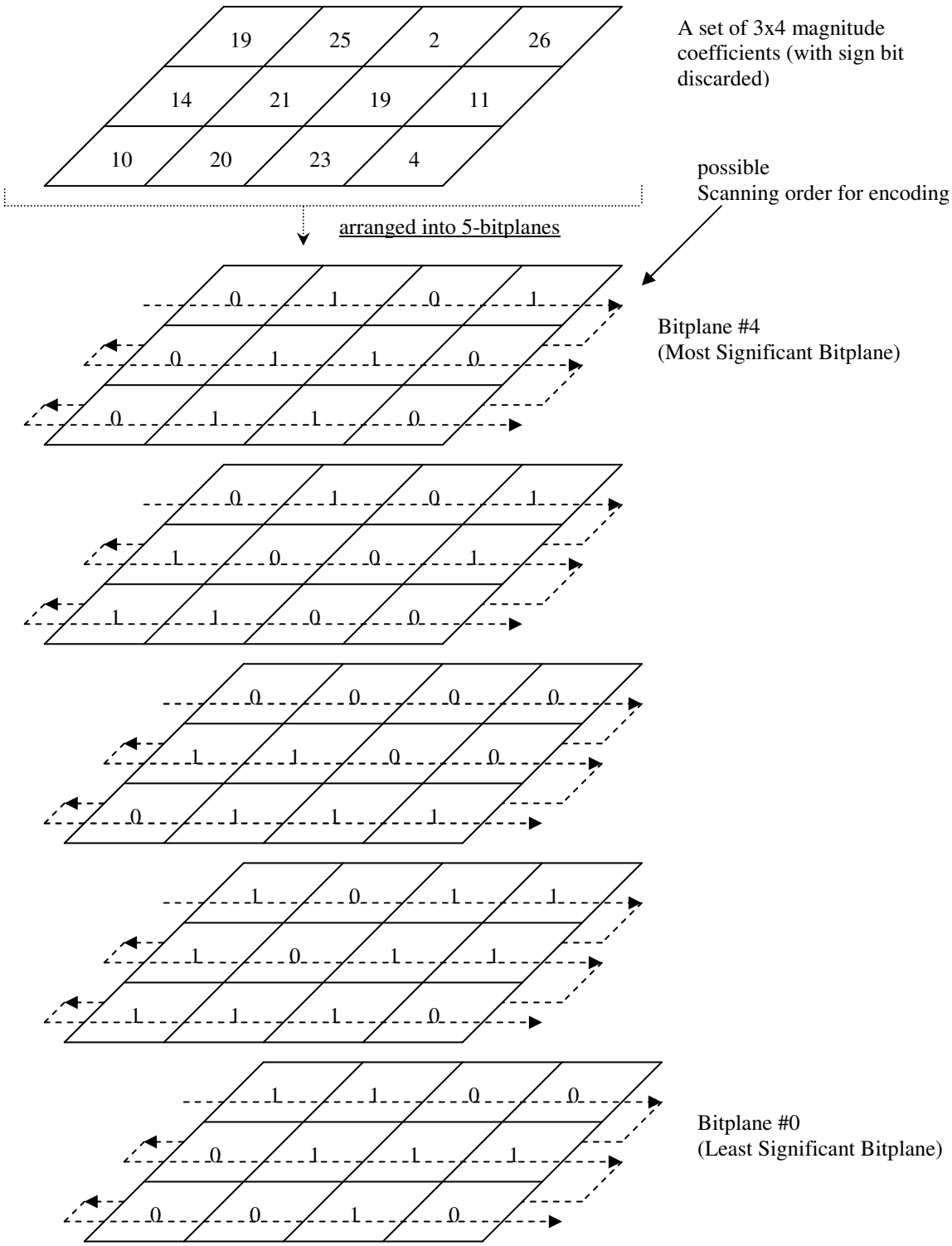



Figure 3.7 An example of bitplane quantiser and its encoding order on the magnitude bitplane. Encoding begins at the MSB Plane, then progressively reaches the LSB Plane. Within each bitplane, scanning order begins from top left corner at the first row until it reaches bottom right at the last row in a zig-zag scanning sequence.

Figure 3.7 shows an example of a group of 3×4 data in their bitplane representation and a possible order in which they may be encoded. Essentially, the bitplane representation re-organises the source symbols into bitplane symbols that are encoded within each bit plane successively with traditional entropy encoding techniques [81, 104, 105], resulting in either information-preserving (i.e., lossless) operation or information-destroying (i.e., lossy) operation. In the case of lossy operation, where successive bitplane coding with bitplane levels lower than $l$ are truncated, the distortion is equivalent to having a scalar quantiser in operation, and the quantised value, $\hat{X}_q(m,n)$, of coefficient, $X(m,n)$, produced by the bitplane quantisation amounts to

$$\hat{X}_q(m,n) = sign(X(m,n))\left\lfloor \frac{|X(m,n)|}{2^l} \right\rfloor \cdot 2^l . \tag{3.30}$$

where $\lfloor\ \rfloor$ means rounding down to the nearest integer value. In equation (3.30), the bitplanes used are magnitude bitplanes of sign-magnitude representation. With partial bitplane truncation, it is obvious that some coefficients may be zero while others with lower significant bits being set to binary '0's because of bit rate constraint as in the case of EZW[31], SPIHT [32], and EBCOT[14]. Hence, progressive encoding with bitplane quantisation has the effect of successive approximation [116]. Consequently, encoding from higher bitplane levels first before lower bitplane levels result in lower MSE. Examples of hierarchical bitplane coders include EZW [31], SPIHT [32], EBCOT [14], JPEG2000 [12].

## 3.4 Hierarchical Bitplane coders

This section shall focus on the discussion of hierarchical bitplane coders (i.e., EZW, SPIHT, and EBCOT) that share some common principles coding strategies in the following way:

(a) wavelet transform the image data,

(b) encoding transform data in progressive bitplane quantisation scheme, and thus provides bit streams that have rate scalability for EZW and SPIHT coders, and rate, resolution and quality scalability for EBCOT. The core coding design principles of EBCOT have been adopted by the state-of-the-art JPEG2000 image compression standards [12].

### 3.4.1 Embedded Zero-tree Wavelet (EZW)

Shapiro [31] has noted that zeros symbols in subbands can be predicted from low resolution level to high resolution level across scales, and hence he proposed the EZW algorithm with a hierarchical bitplane coding scheme for still images. It is found that wavelet coefficients at the higher resolution subbands of the same orientation belonging to the same spatial location have high probability of being insignificant if the corresponding wavelet coefficient at the lower resolution subband is insignificant with respect to a given threshold, $T$ [31]. From this empirical evidence, a zero-tree data structure is used to define the relationship between coefficients across scales. Dependencies between wavelet coefficients across subbands at different resolution levels are depicted in Figure 3.8. In Figure 3.8, every coefficient in the LL3 (i.e., isotropic DC band at the lowest resolution level) is directly related to coefficient in the three orientation bands (LH3, HL3, HH3) at the same spatial location. Each coefficient in the orientation subbands of HL3, LH3, and HH3 is related to four coefficients in the HL2, LH2, and HH2 subbands, respectively. The dependencies of coefficients across resolution levels are classified as,

(1) Parent – Any coefficient at a lower resolution subband of the tree with same spatial and/or orientation position than the current coefficient. In Figure 3.8, a coefficient in LL3 is a parent of coefficients in HL3, LH3 and HH3 at the same spatial location. A coefficient in LH3 is a parent of four coefficients in LH2, and a coefficient in LH2 is a parent in LH1. However, all the coefficients in LH1, HL1 and HH1 cannot be parents as they are the leaves of the tree.

(2) Child – a coefficient is a child if it has a parent coefficient in the next lower resolution subband at the same spatial and/or orientation position. The children

in HL2 have a parent coefficient in HL3. In the case of coefficients in LL3, they have no parents.

(3) Descendants – For a given parent, the set of all coefficients at all higher resolution subbands of same spatial and/or orientation locations are defined as descendents. A coefficient in HH3 in Figure 3.8 has 20 descendants (i.e., 4 in HH2 and 16 in HH1).

(4) Ancestors – For a given child, the set of coefficients at all the lower subbands of the same spatial and/or orientation locations are called ancestors. A coefficient at LH2 has two ancestors (i.e., 1 each at LH3 and LL3).

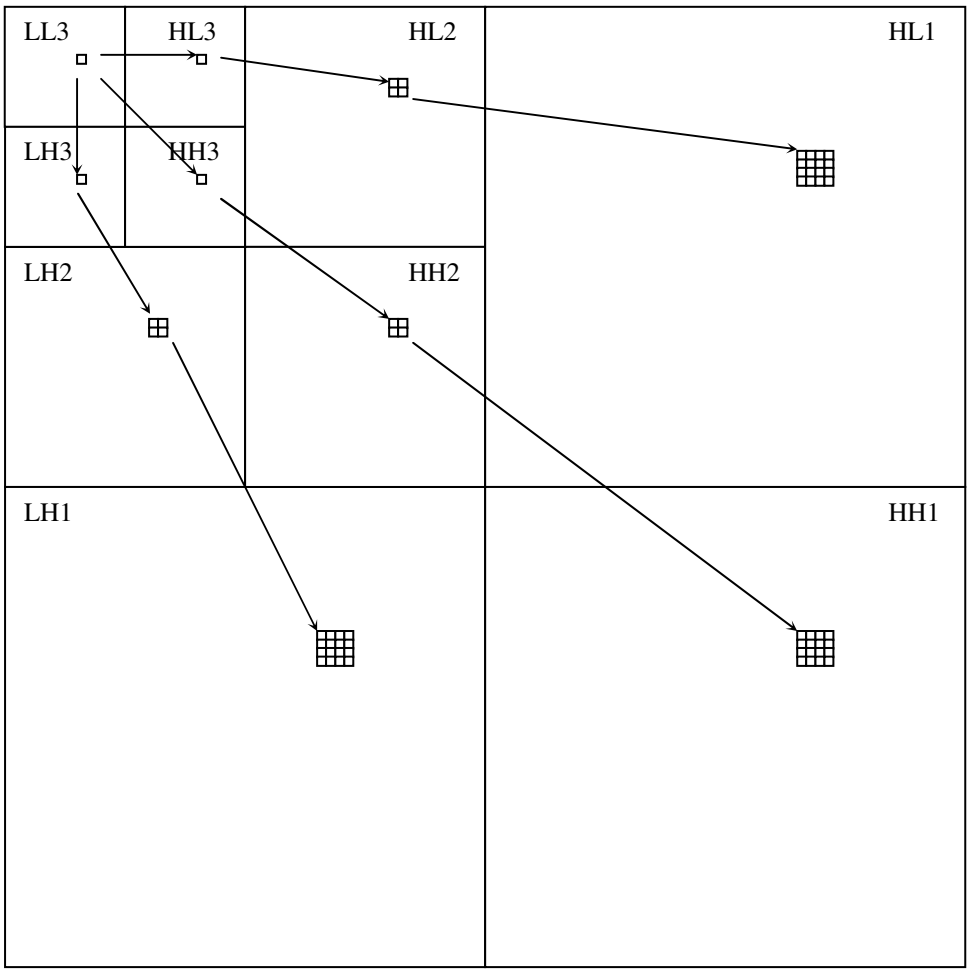



Figure 3.8 Parent-child relationship in EZW.

The bitplane encoding process starts at the lowest resolution band, denoted by $LL_n$, and the orientation bands in the order of $LL_n$, $HL_n$, $LH_n$, and $HH_n$ at resolution level $n$. It then moves on to the next higher resolution ($n$-$1$) at $HL_{n-1}$, $LH_{n-1}$ and $HH_{n-1}$. During the scanning process, no coefficient is scanned before its parent and all coefficients within a subband must be scanned in a raster format before scanning moves to the next subband. The bitplane encoding involves a two-pass process, namely a dominant pass followed by a subordinate pass, commencing from the MSBP and ending at the LSBP. At the highest bitplane, $P_{max}$, the dominant pass updates the significant map by determining if a coefficient is significant or not with respect to a threshold level, $T_{p_{\max}}$. An insignificant coefficient is one whose magnitude is below a threshold level, $T_{p_{\max}}$. Once the status of a coefficient is determined, it will be updated on the significant map with one of the four coding symbols defined for dominant pass.

For any other bitplane, $p$, coefficients that have not been found to be significant during the previous bitplane will be scanned during the dominant pass to determine if they are significant or not with respect to threshold level, $T_p = (T_{p+1})/2$, where bitplane $p+1$ is higher than bitplane $p$.

The four coding symbols defined for the dominant pass are (1) zerotree root (ZTR), (2) isolated zero (IZ), (3) positive significant (POS), and (4) negative significant (NEG). The ZTR is used when a coefficient and all its descendants are insignificant, but itself is not the descendant of a previously found zero-tree root. If an insignificant coefficient has significant descendant(s), it is coded with IZ symbol. The POS symbol is used for coding a significant coefficient that has a positive value, and the NEG symbol is used for a negatively signed significant coefficient. In addition, a Magnitude Refinement (MR) symbol, which is used in the subordinate pass, is used for coding the bitplanes of coefficients that are found to be significant in the dominant pass. Figure 3.9 shows the flow chart for encoding a coefficient of the significant map.

The zerotree coding effectively reduces the cost of encoding the significant map with the use of self-similarity that exists between coefficients across scales as the

appearances of insignificant coefficients across scales are not strictly independent events. When a zerotree root is coded, all the descendants following the "zerotree root" symbol of the insignificant coefficient need not be coded. Effectively, only the significant map and the significant coefficient of the current bitplane along with their children are coded. The two-pass approach in the bitplane coding also allows the different PDF to be used in the dominant pass and subordinate pass separately. This provides a better statistical model and thus entropy coding is expected to be more efficient.

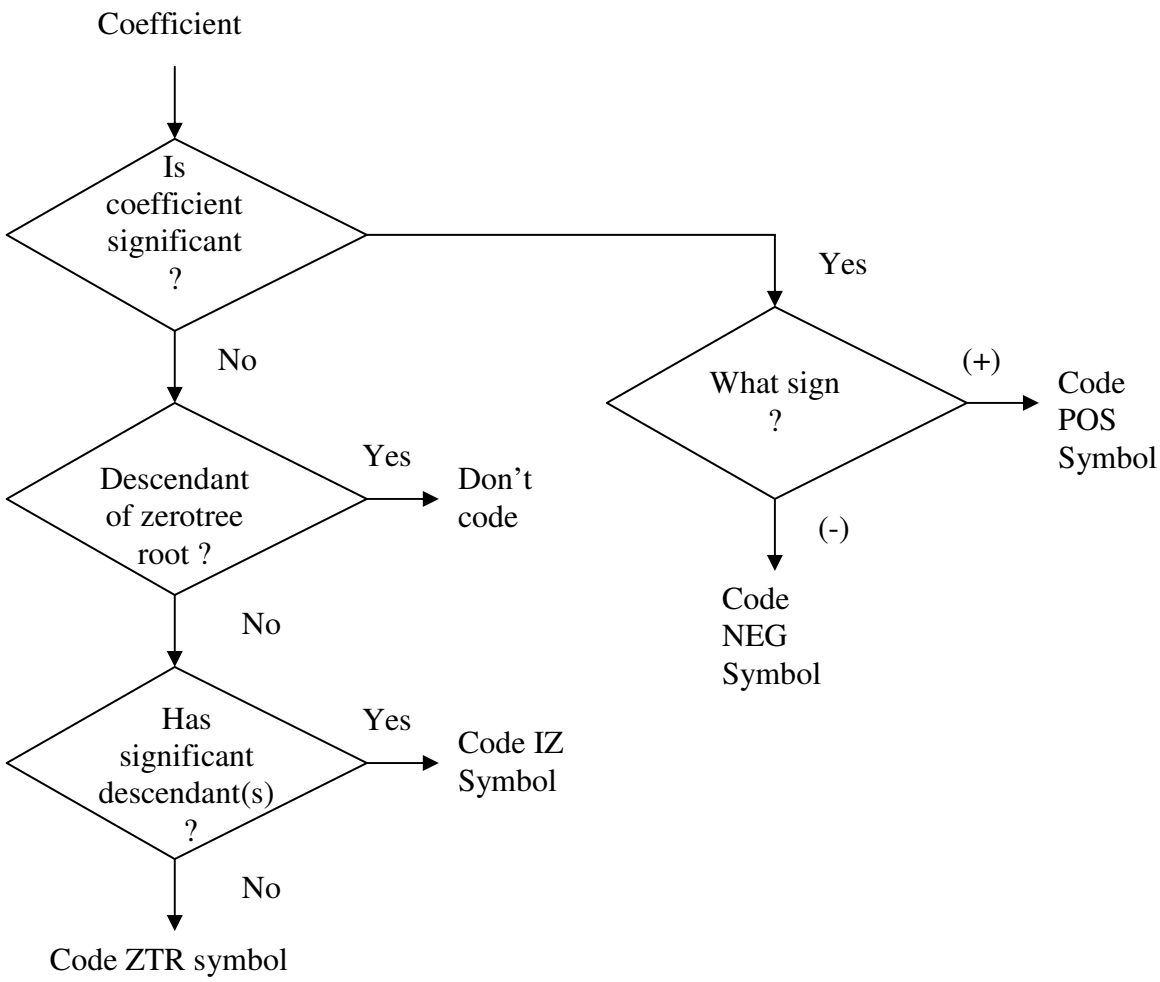



Figure 3.9 Flow chart for encoding a coefficient of the significant map.

Undoubtedly, EZW represents a significant contribution and novelty in the design of hierarchical bitplane coders. Subsequent improvement of this algorithm can be found in [117]. Its popularity has motivated the development of SPIHT [32] and the EBCOT [14] coders. Monro et al. [118] has also extended the EZW approach to block-based transform coding, where zero-tree coding for DCT coefficients is proposed.

### 3.4.2 Set Partitioning In Hierarchical Tree (SPIHT)

The SPIHT coder [32] offers an extension to the EZW coder [31]. In the EZW coding, there is the partial ordering of the transform coefficients with respect to a set of threshold values. In SPIHT, however, a set partitioning sorting procedure is used, and a significant test is performed on the partitioned set, $\Gamma$, of coefficients. The magnitude of the maximum coefficient in the given partitioned set, $\Gamma$, is tested against a threshold, $T_n$, and the set is considered significant if $\max_{(i,j)\in\Gamma}\{|c_{i,j}|\} \geq T_n$. If the test is insignificant, all the other coefficients in the partitioned set are also considered as insignificant. With the exception of the relationship in the $LL_D$ (the lowest isotropic DC band), the parent-child relationships in the SPIHT are similar to that of the EZW. Referring to Figure 3.10 on the SPIHT, one quarter of the coefficients (with even horizontal and vertical coordinates) in the $LL_D$ have no children, while the rest of the coefficients each have four children. For the other three regions, the HH, HL and LH, the parent-child relationships for SPIHT are similar to that of the EZW.

There are three ordered lists in SPIHT:

1) List of significant coefficients (LSC)
2) List of insignificant coefficients (LIC)
3) List of insignificant sets of coefficients (LIS)

The set of coordinates of immediate children, descendents and non-immediate-offspring descendents are represented as $I(i,j)$, $D(i,j)$, and $D_{ni}(i,j) = D(i,j) - I(i,j)$, respectively.

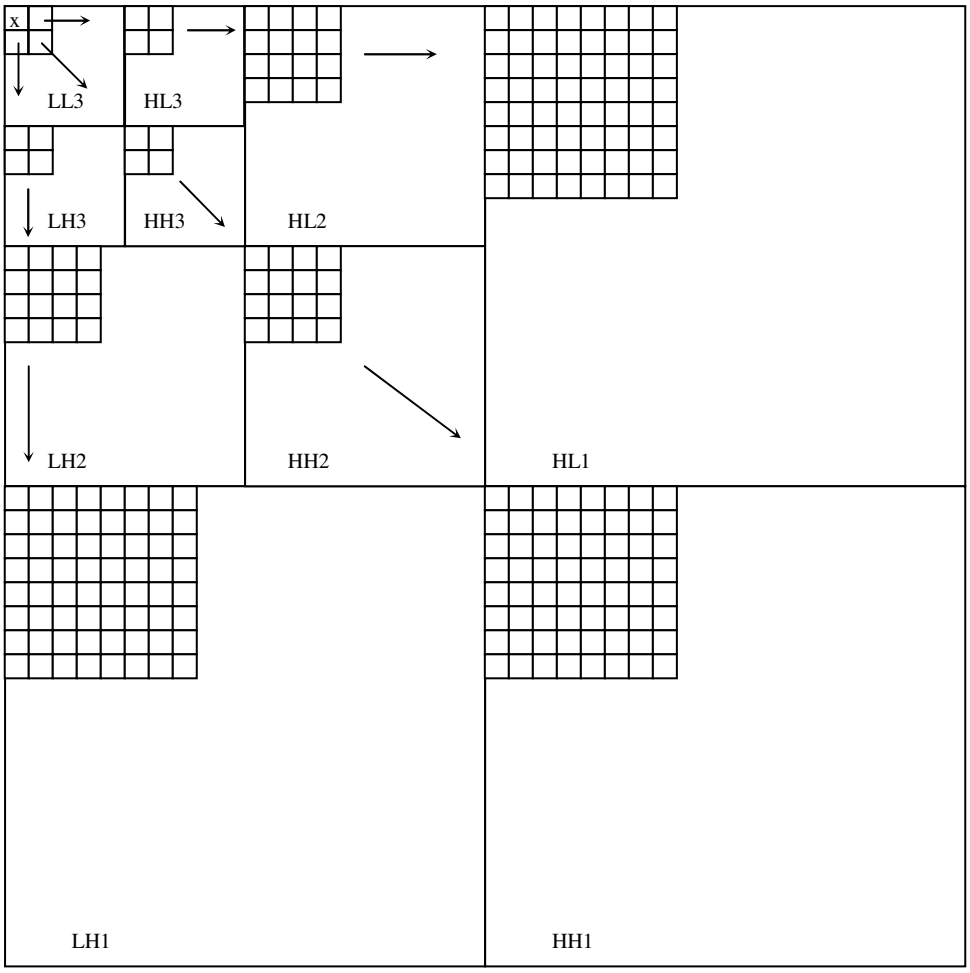
Beginning with the highest bitplane, each bitplane is treated with the significance and the refinement pass. At the initialization stage, the LSC is reset as an empty set, the coordinates $(i, j)$ of all coefficients in the $LL_D$ region enters the LIC, and those coefficients with children are added to the LIS as roots of Type A. Next, all coefficients in LIC are examined and coded starting from the MSB plane. For significant coefficients, their signs are output, and the significant coefficients are moved to LSC. All the set of coefficients in the LIS are also examined and coded in sequential order, one set at a time. If a set of coefficients in the LIS is significant and belongs to type A, two possible outcomes arise:

(a) if the set of its immediate children $I(i, j)$ is significant, the coordinates of children coefficients are moved to LSC and the signs of their coefficients are output. Otherwise,

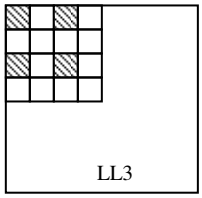
(b) the coordinates of the immediate children coefficients are moved to LIC.

If the set only has immediate children but no other descendents, the set would be removed from LIS. If the set has non-immediate offspring (i.e., $D_{ni}(i, j) \neq 0$), the coordinate (*i, j*) is moved to LIS as type B. If a set of coefficients in the LIS is insignificant, a 0 bit is coded. If a set of coefficients (*i, j*) in the LIS belonging to type B and the set of its non-immediate-offspring descendents ($D_{ni}(i, j)$) are significant, the coordinates of its immediate children are added to the end of LIS as type A. The entry of the set of coefficients (*i, j*) is removed from LIS.

At the refinement pass, all LSC coefficients are coded, except those that have just been added to LSC. The coding proceeds for the next lower bitplane by visiting entries in the LIP, LIS and LSC.




(a)




(b)

Figure 3.10 (a) Parent-child relationship in SPIHT. (b) Shaded region indicates coefficients in the $LL_3$ (the lowest DC Level) that have no children.

Similar to the EZW, encoding can halt at any time when the desired coding rate is achieved. Empirical studies too have shown that the SPIHT has achieved better coding results than the EZW [1, 12, 32]. While there are 3 coding passes in SPIHT as opposed to 2 coding passes in EZW, the extra coding pass in SPIHT can provide fine embedding of information which can potentially be exploited for HVS-based rate control scheme.

### 3.4.3 Embedded Block Coding with Optimized Truncation (EBCOT)

The EBCOT [14] algorithm employs DWT with either the Mallat dyadic [103] or packet wavelet decomposition structure [90]. The DWT samples are then bitplane quantised and encoded with context arithmetic coding. Similar to EZW and SPIHT, the EBCOT is a scalable coder. While EZW and SPIHT generate bitstreams that are rate scalable, the EBCOT produces bitstreams that are quality and resolution scalable. The output bit stream consists of embedded subsets (codeblock layers) which are independently compressed.

Resolution scalability translates to the ability to reconstruct an image at different resolution levels. Quality scalability means that images can be reconstructed with different quality levels, relative to some quality measure. When the bitstream is both resolution and quality scalable, it means that the compressed bit stream can be decoded to different resolution or quality levels [12, 14, 103].

EBCOT utilizes a two-tier coding strategy. During tier one coding, each subband is divided into independent code-blocks of 32x32 or 64x64 samples each. Each codeblock is encoded bitplane layer by bitplane layer. Each bitplane layer is further segregated to fractional bitplane layers to form addition truncation points on the R-D curve. Associated with each fractional bitplane layer is the rate (in bits) required to encode the layer and the distortion reduction resulting from the encoding of the layer. The rate increase and the distortion reduction for all truncation points are then used in the Post Compression Rate Distortion (PCRD) optimisation in the tier two coding to optimise the final bitstream. The Partitioning of codeblock has the advantage of

minimising the use of memory [14]. Compressing individual blocks as opposed to the whole image is more resource efficient.

Every sample in the codeblock is coded by four different types of coding primitives: Zero Coding (ZC), Run-Length Coding (RLC), Sign Coding (SC) and Magnitude Refinement (MR). While it may be reasonable to assume the correlation between the current codeblock and its neighbours as insignificant (in order to ensure that each block's bit-stream remains independent), this does not hold for the neighbours of each subblock. In the presence of an insignificant sample, the ZC is used. However, if a horizontal run of insignificant samples is encountered, the RLC is used instead of ZC. SC is employed to determine the sign of the sample and is used only once for each sample. Conversely, significant samples are subjected to the MR primitive coding operation [14].

Starting from the MSB, bitplane coding is carried out through four coding passes, each generating its own truncation point. As shown in Figure 3.11, more truncation points do provide finer approximations to the R-D curve.

The four coding passes are described as follows:

1) Forward Significance Propagation Pass ($P_1^P$): This pass proceeds through the sub block samples in a scan-line fashion, omitting all samples which are insignificant. Here, the ZC or RLC is employed to identify the significance of the sample, and if found to be significant, the SC coding operation is executed.
2) Reverse Significance Propagation Pass ($P_2^P$): Similar to the coding pass in ($P_1^P$), this scanning is done in the reverse order. Samples which are coded in the previous pass are omitted, while samples with at least one significant neighbour (of the 8 immediate neighbours) are added.
3) Magnitude refinement Pass ($P_3^P$): All samples which were previously found to be significant are coded with the MR coding operation.
4) Normalisation Pass ($P_4^P$): The least significant bit of the remaining samples which were not visited by the preceding three passes is coded using the RLC primitive, and if a sample is significant, its sign will also be coded immediately with a SC primitive.

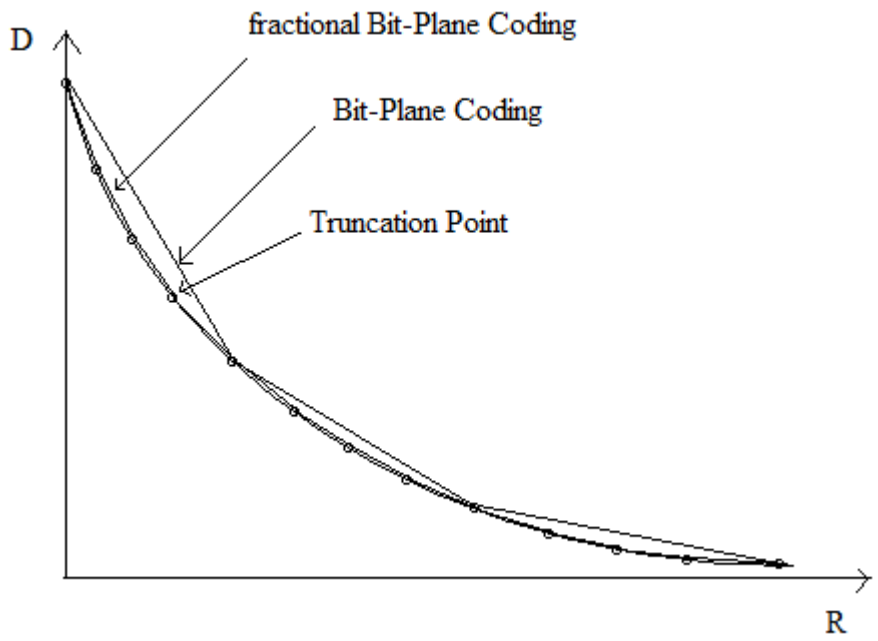



Figure 3.11 Rate Distortion Curve with Bitplane.

## *3.5 Perceptual Coders and Psychophysical Quality Metrics*

Traditional transform coders achieved excellent compression ratio by exploiting the statistical redundancies exists in the image data. However, reduction of statistical redundancies does not necessary equate to the reduction of psychovisual redundancies. Since the human observers are the ultimate judges of picture quality, picture coders should ideally remove psycho-visual redundancies, and thus retain visually relevant information in image data. Hence, it would be beneficial to incorporate aspects of the HVS into the coding process to improve picture quality of coded images. Perceptual coders can be widely classified into rate driven or quality driven.

### 3.5.1 Watson's DCTune

Watson's DCTune [119] is based on the standard DCT coder with vision modeling for quantisation matrix. In the earlier publication by Paterson et al. [120, 121], the threshold for DCT basis functions is measured. It is found that there exists a smallest coefficient that shows psychophysical visible distortion for a certain DCT basis function at index (*u, v*). This value is known as the threshold, $t_{uv}$. The highest possible quantisation error at this threshold point is,

$$e_{uvk} = \frac{q_{uv}}{2} \tag{3.31}$$

where $e_{uvk}$ is the maximum quantisation error for $k^{th}$ DCT block at index (*u,v*). If the element in the quantisation matrix is set at $\frac{q_{uv}}{2} = t_{uv}$, it will ensure that errors are visually imperceptible. Hence,

$$q_{uv} = 2t_{uv} \tag{3.32}$$

The quantisation matrix (QM), $q_{u,v}$, is thus dependent on the visually perceptible maximum possible quantisation errors at various DCT basis functions, but independent of the image. Watson called it the "image-independent perceptual" (IIP). However, Watson in DCTune [119] proposes an image dependent perceptual (IDP) approach for formulating a QM tailored to specific images. The IDP approach gives rise to a given perceptual error, based on the DCT coefficients by considering both the effects of contrast and luminance masking. The model for the masked threshold, $m_{uvk}$, is as follows,

$$m_{uvk} = \max\left(t_{uvk}, \left|c_{uvk}\right|^{w_{uv}} t_{uvk}^{\;1-w_{uv}}\right) \tag{3.33}$$

where $w_{uv}$ is the exponent having a value between 0 and 1, $t_{uvk}$ and $c_{uvk}$ are the luminance masking threshold and the DCT coefficient, respectively. Note that the image is first divided into blocks of size 8x8, and $k$ denotes the index of a block (size of 8x8) of image, $u$ and $v$ are indices of the DCT frequency (or basis function). The DCT coefficient, $c_{uvk}$, can be computed by equation (3.21) (i.e., the $X[k,l]$ in equation (3.21)). The luminance masking threshold, $t_{uvk}$, can be found by the formula supplied by Ahumada and Peterson [122]. The perceptual distortion due to quantisation error when considering the effect of masking is thus expressed as,

$$d_{uvk} = \frac{e_{uvk}}{m_{uvk}} \tag{3.34}$$

Minkowski metric is used to pool the Just-Noticeable-Differences (JND), $d_{uvk}$, for a particular frequency at ($u$, $v$) over all DCT block, k, as follows,

$$D_{uv} = \left( \sum_{k} \left| d_{uvk} \right|^{\beta} \right)^{\frac{1}{\beta}} \tag{3.35}$$

Where $D_{uv}$ is the perceptual error at ($u$, $v$). Pooling all the elements of ($u$, $v$) of the perceptual error leads to the overall distortion as,

$$D = \left( \sum_{u} \sum_{v} D_{uv}{}^{\lambda} \right)^{\frac{1}{\lambda}} \tag{3.36}$$

If the exponent, $\lambda \to \infty$, $D$ is $\max(D_{uv})$, and the minimum bitrate for a given $D = \psi$ is achieved when $D_{uv} = \psi$, where $\psi$ is the perceptual error.

The optimisation of the quantisation matrix (QM) can be determined by assuming $\lambda \to \infty$, and the QM becomes separate optimisation of individual elements of the matrix. Each entry of the perceptual error, $D_{uv}$, is an independently monotonically increasing function of the respective elements in the QM.

When coding Lena at 0.25 bpp separately by IDP and IIP approaches, Watson [119] has reported that the IDP approach produced better perceived quality improvement over the IIP approach.

### 3.5.2 Subband Image Coder by Safranek and Johnston

This coder [123] presents coding of wide selection of images with rates of less than 1 bit per pixel (bpp). It employs differential pulse code modulation (DPCM), entropy coding, perceptual-threshold calculation, and quiescent block rejection.

Each image is transformed using the GQMF filter bank [94, 124] into four bandpass sub-images. The RMS noise sensitivity threshold (also called based noise sensitivity) for each subband was determined through a series of informal sensitivity testing. By adjusting the luminance level and base sensitivity, both frequency content and image brightness for a flat-field image, which the human eye is sensitive to, are accounted for. The perceptual threshold calculation is expressed in dB as follows:

$$pt(b,u,v) = B(b) - 0.15\,log(T(u,v)) - W \cdot C(u,v) \tag{3.37}$$

where $b$ is the subband, $u$ and $v$ correspond to the pixel location. $B(b)$ is the base noise sensitivity for subband, $b$. $W$ and $C(u,v)$ are the brightness weighting factor and the brightness correction, respectively, The brightness factor takes into consideration luminance variations. Notice also the function has a texture energy variable, $T(u,v)$, for textural masking adjustment as Safranek and Johnston [123] generally believe that textured regions are over coded. The texture energy function is:

$$T(u,v) = \sum_{b=1}^{15} W_{mtf}(b) \cdot E(b,u,v) + W_{mtf}(0) \cdot var((u,v),(u+1,y),(u,v+1),(u+1,v+1)) \tag{3.38}$$

The weights, $W_{mtf}$, are assigned based on the modulation transfer function (MTF) [125]. The $var(,,,)$ is the variance taken over a 2x2 area with the target pixel in the

upper left corner at ($u$, $v$), $E(b,u,v)$ is the local energy in the subband, $b$, except subband zero. Essentially, the texture masking function is the weighted sum of the texture energy at each image location.

### 3.5.3 Perceptually Tuned Subband Image Coding by Chou and Li

Chou and Li [126] propose a method to estimate the JND and minimally noticeable distortion (MND) profiles of monochromatic images. The JND/MND profiles are used to remove perceptual redundancy in their subband coding algorithm. The JND profile is computed as follows,

$$JND_{fb}(x,y) = max\{f_a(x,y), f_d(x,y)\} \tag{3.39}$$

$$f_a = ag(x,y)(0.0001 \cdot bg(x,y) + 0.115) + \lambda - 0.001 \cdot bg(x,y) \tag{3.40}$$

$$f_d(x,y) = \begin{cases} T_0 \cdot \left(1 - \sqrt{\dfrac{bg(x,y)}{127}}\right) + 3, & if \quad bg(x,y) \le 127 \\ \gamma \cdot (bg(x,y) - 127) + 3, & if \quad bg(x,y) > 127 \end{cases} \tag{3.41}$$

where $ag(x,y)$ and $bg(x,y)$ are the weighted average luminance differences and mean background luminance around pixel $(x,y)$, respectively. The parameters, $\lambda$, $T_0$, and $\gamma$, were derived from subjective experiments and curve fitting. The values of these parameters increase with increasing viewing distance. While $JND_{fg}$ profile encodes images to an imperceptible difference level, the MND profile encodes images to a target bitrate while minimising visual distortion. The MND profile is computed as follows,

$$MND_{g,fb}(x,y) = JND_{fb}(x,y) \cdot g \tag{3.42}$$

where $g$ is the distortion index ranging between 1.0 and 4.0. After the JND or MND have been computed from the image data, it is decomposed into respective subbands

(i.e., each JND or MND per subband) in the frequency domain with their MTF weights, where each MTF weight is the average MTF value of its subband. The decomposed JNDs or MNDs in the subbands are used in the DPCM encoding to achieve the desired bitrate and visual quality.

### 3.5.4 Locally Adaptive Perceptual-based Image Coding by Hontsch and Karam

Hontsch and Karam's Locally Adaptive Perceptual Image Coding (LAPIC) [127] is an extension of their earlier work [128, 129] that uses adaptive quantisation scheme with DPCM coding within the domain of Generalised Quadrature Mirror Filter (GQMF) Bank [94, 124]. The earlier work is based on the concept of JND [130], incorporating aspects of the HVS of contrast sensitivity, luminance and contrast masking. The quantisation scheme estimates the JND threshold at the encoding stage. A similar process is carried out to estimate its JND threshold during dequantisation at the decoding stage without side information, and hence eliminating the need to transmit adaptive quantisation step sizes. This quality driven coder produces superior quality images than its predecessor [123].

Being an expansion of the previous work [128, 129] that are based on GQMF, the LAPIC is based on discrete cosine transform (DCT) and uses JND threshold for DCT coefficients. Contrast sensitivity and contrast masking are the two visual mechanism employed in the computation of the JND thresholds denoted as $t_{JND}(b,n_1,n_2)$. It is defined as,

$$t_{JND}(b,n_1,n_2) = t_{DCT}(b,n_1,n_2) \cdot a_{CM}(b,n_1,n_2) \tag{3.43}$$

where $t_{DCT}(b,n_1,n_2)$ and $a_{CM}(b,n_1,n_2)$ are background luminance-adjusted contrast sensitivity threshold and contrast masking adjustment, respectively. The index $b$ denotes the DCT subband number, $n_1$ and $n_2$ identify the coefficient location within the subband, $b$.

The contrast sensitivity threshold is derived as,

$$t_{DCT}\left(b(i,j),n_1,n_2\right)=\frac{MT_{i,j}\left(n_1,n_2\right)}{2\alpha_i\alpha_j\left(L_{max}-L_{min}\right)} \tag{3.44}$$

where $T_{i,j}\left(n_1,n_2\right)$ is the background luminance-adjusted contrast sensitivity of the luminance error due to quantisation of DCT coefficient, $c_{i,j}$, in DCT block $\left(n_1,n_2\right)$. $M$ being the gray levels, $L_{mn}$ and $L_{max}$ are the minimum and maximum display luminances, and, $\alpha_z=\begin{cases}1, & z=0\\ 1/\sqrt{N_{DCT}}, & z\neq 0\end{cases}$ with $z=\{i,j\}$. $\alpha_i$ and $\alpha_j$ are the DCT coefficient normalization factors. The block size of DCT, $N_{DCT}$, is 8.

$T_{i,j}\left(n_1,n_2\right)$ is based on empirical model [122] that was obtained in psychophysical experiments of fitting CSF data, and it is computed as,

$$T_{i,j}\left(n_1,n_2\right)=\frac{T_{\min}\left(n_1,n_2\right)}{r+\left(1-r\right)\cdot\cos^2\Theta_{i,j}}10^{K(n_1,n_2)\left(\log_{10}f_{i,j}-\log_{10}f_{\min}(n_1,n_2)\right)^2} \tag{3.45}$$

where $f_{i,j}$ is the spatial frequency corresponding to DCT coefficient in location $(i,j)$, and is given as,

$$f_{i,j}=\frac{1}{2N_{DCT}}\sqrt{\frac{i^2}{w_x^2}+\frac{j^2}{w_y^2}} \tag{3.46}$$

the orientation, $\Theta_{i,j}$, $T_{min}\left(n_1,n_2\right)$, $f_{\min}\left(n_1,n_2\right)$, and $K\left(n_1,n_2\right)$ are, respectively, computed as,

$$\Theta_{i,j}=sin^{-1}\frac{2f_{i,0}f_{0,j}}{f_{i,j}^2} \tag{3.47}$$

$$T_{min}(n_1,n_2)=\begin{cases}\left(\dfrac{L(n_1,n_2)}{L_T}\right)^{\alpha_T}\dfrac{L_T}{S_0}, & L(n_1,n_2)\le L_T\\ \dfrac{L(n_1,n_2)}{S_0}, & L(n_1,n_2)>L_T\end{cases} \tag{3.48}$$

$$f_{min}(n_1,n_2)=\begin{cases}f_0\left(\dfrac{L(n_1,n_2)}{L_f}\right)^{\alpha_f}, & L(n_1,n_2)\le L_f\\ f_0, & L(n_1,n_2)>L_f\end{cases} \tag{3.49}$$

$$K(n_1,n_2)=\begin{cases}K_0\left(\dfrac{L(n_1,n_2)}{L_K}\right)^{\alpha_K}, & L(n_1,n_2)\le L_K\\ K_0, & L(n_1,n_2)>L_K\end{cases} \tag{3.50}$$

The local background luminance, $L(n_1,n_2)$, is computed as,

$$L(n_1,n_2)=L_{min}+\frac{L_{max}-L_{min}}{M}\cdot\left(\frac{\sum_{(0,m_1,m_2)\in F_{(0,n_1,n_2)}} c_{(0,m_1,m_2)}}{N_{DCT}\,n\left(F_{(0,n_1,n_2)}\right)}+m_g\right) \tag{3.51}$$

This is based on a fovea region of about 2 degree angle. $n\left(F_{(0,n_1,n_2)}\right)$ that is taken as follows,

$$n\left(F_{(0,n_1,n_2)}\right)=\left(\frac{2DR\tan\left(\dfrac{\theta}{2}\right)}{N_{DCT}}\right)^2 \tag{3.42}$$

where $D$, $R$ and $\theta$ are the viewing distance, display resolution, and visual angle, respectively. The contrast masking adjustment, $a_{CM}(b,n_1,n_2)$, is computed as follows,

$$a_{CM}(b,n_1,n_2)=\begin{cases} max\left\{1,\left|\dfrac{c_{F_{(b,n_1,n_2)}}}{t_{DCT}(b,n_1,n_2)}\right|^{0.6}\right\}, & b\neq 0 \\ 1, & b=0 \end{cases} \tag{3.53}$$

where $c_{F_{(b,n_1,n_2)}}$ is the average magnitude of the DCT coefficients in $F_{(b,n_1,n_2)}$, and $F_{(b,n_1,n_2)}$ denotes the area centre at location $(n_1,n_2)$ in subband $b$ that covers 2 degrees of visual angle.

For imperceptible quantisation errors, the uniform quantisation step size, $s_{JND}(b,n_1,n_2)$, is computed as,

$$s_{JND}(b,n_1,n_2)=2\hat{t}_{JND}(b,n_1,n_2) \tag{3.54}$$

where $\hat{t}_{JND}(b,n_1,n_2)$ is the estimated threshold at location $(b,n_1,n_2)$. $\hat{t}_{JND}(b,n_1,n_2)$ is computed based on equation (3.43) except with $F_{(b,n_1,n_2)}$ being replaced by a causal fovea region. Compared with Watson's DCTune [119], the Locally Adaptive Perceptual Image Coding has improved image quality, especially, at low bitrate as reported in [127] .

### 3.5.5 EBCOT with Visual Masking by Taubman

In EBCOT [14], the default measure for distortion is the MSE. However, it is well known that MSE is not a good measure for visual distortion. Taubman proposes a spatially varying distortion metric [14] that incorporates masking phenomenon within the distortion function. Accordingly, the visual distortion metric (VDM), also known as the CVIS, has the following expression,

$$D_z^n = w_{b_z}^2 \cdot \sum_k \frac{\left(\hat{x}_z^n[k]-x_z[k]\right)^2}{\sigma_{b_z}^2+\left(V_z[j,k]\right)^2} \tag{3.55}$$

where $x_z[k]$ and $\hat{x}_z^n[k]$ denote the subband sample and quantised representation of the subband sample, respectively, in code-block, $B_z$, at location $k = (k_1, k_2)$, where $k_1$ and $k_2$ are the horizontal and vertical positions, respectively, for subbands HH, LH, and LL. In the case of HL, $k_1$ and $k_2$ represent the vertical and horizontal positions, respectively. $w_{b_z}$ is the $L_2$ norm of basis function of wavelet transform for subband, $b_z$ ,which contains the code-block, $B_z$, $\sigma_{b_z}$ is provided for minimum level of inhibition. $V_z[j,k]$ denotes the visual masking strength, and is computed as,

$$V_z[k] = \frac{\sum_{u \in \eta_z[k]} |x_z[u]|^{\rho}}{\|\eta_z[k]\|} \tag{3.56}$$

where $\eta_z[k]$ denotes the neighbourhood of samples about $x_z[k]$, and $\|\eta_z[k]\|$ denotes the size of the neighbourhood. The neighbourhood is obtained by partitioning the code-block, $B_z$, into *8x8* blocks, and the exponent, $\rho$, is set to 0.5. It is noted that the normalized image samples with a range of 0 to 1 has been used for the non-linear operation above.

### 3.5.6 Point-wised Extended Visual Masking by Zeng, Daly and Lei

Embedded into the JPEG2000 coder [131], the Point-wised Extended Visual Masking coding [132] by Zeng et al. incorporates self-contrast masking and neighbourhood masking effects by introducing a non-linear function that maps the wavelet coefficients into perceptual domain. In contrast to EBCOT's Visual Masking [14] where masking effects were considered after quantisation, here a signal that is subject to masking is elevated by a power function and then followed by a divisible neighourhood masking weighting factor. The masking operator modifies the DWT coefficients, and hence an inverse process is required at the decoder. While the neighbourhood masking weighting factor could also include neighourhood coefficients from interbands, the final model that has been adopted by the JEG2000 standard only considers intra-band masking, where the neighourhood masking

weighting factor includes neighbouring coefficients from the same subband. The final model maps the wavelets coefficients as follows,

$$y_k = \frac{sign(x_k) \cdot |x_k|^{\alpha}}{1 + a\sum_i \frac{|\hat{x}_i|^{\beta}}{|\phi_i|}} \tag{3.57}$$

where $x_k$ is the wavelet coefficient, $\alpha$ is the power factor for self-contrast masking having a value between 0 and 1, $sign(x_k)$ gives the sign of the wavelet coefficient, $x_k$ , $a$ being the normalisation factor. $|\phi_i|$ denotes the size of the causal neighbourhood. $|\hat{x}_i|$ are the quantised coefficients of the causal neighbourhood for coefficient, $x_k$. The exponent, $\beta$, is greater than zero. The typical values for $\alpha$ and $\beta$ are 0.7 and 0.2, respectively. A proper choice of $\alpha$, $\beta$ and $|\phi_i|$ enables the coder to distinguish local sharp edges from a locally complex image region. Figure 3.12 shows the selection of causal neighbourhood coefficients that are quantised coefficients $\hat{x}_i$ prior to $x_k$. From the perspective of coefficient recovery, only causal neighbouring samples are used. This is because the decoder requires causal samples to recover the modified DWT coefficients caused by masking operator at the encoder. These neighbourhood coefficients are chosen so that each coefficient of the quantised coefficients, $\hat{x}_i$, can be recovered prior to recovery of $x_k$ at the decoder.

It is noted that the use of neighbourhood quantised coefficients results in some degree of masking inaccuracy especially when coefficients are coarsely quantised and only the first few most significant bits of the quantised indexed are retained while the remaining lower bits are truncated during bitplane coding.

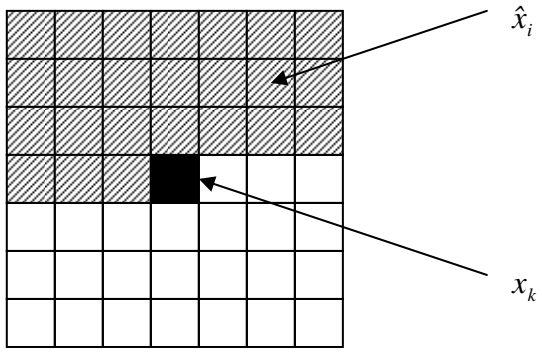



Figure 3.12 Causal neighbourhood coefficients $\hat{x}_i$ (the shaded boxes ▨) for signal $x_k$ ■ in a 7x7 Neighbourhood where $|\phi_i| = 24$. The non-causal coefficients (the unshaded boxes) are not included as the coefficients for computation.

### 3.5.7 Wavelet Visible Difference Predictor by Bradley

In Daly's VDP, an algorithm is developed to determine image fidelity with a vision model by also considering the effect of display parameters and viewing conditions. The output is a probability detection map that provides the location and the degree of visual differences (in the perceptual sense). However, the VDP map does not attempt to discriminate among different suprathreshold visual errors. Three aspects are considered in the VDP: amplitude non-linearity, contrast sensitivity function, and detection mechanism. Basically, two images (an original image and a noisy one) are rescaled by the amplitude non-linearity and CSF functions, before they are filtered by cortex transform. A masked function is applied to the filtered images to determine their masked threshold elevations. The contrast difference and the masked threshold elevation between the two filtered images are used in a psychometric function to compute the probability of detecting the contrast difference. Probability summation is used to pool data over the various cortex channels to create the detection map. A comprehensive coverage of VDP can be found in [133]. The vision model used in the VDP is also included in Appendix G as a reference.

The Wavelet Visible Difference Predictor proposed by Bradley [134] is a modification of the visible difference predictor (VDP), as proposed by Daly [133]. Unlike VDP which is based on the cortex transform, WVDP uses the linear phase 9/7

biorthogonal filter within the hierarchical wavelet transform [135]. Other key modifications are (1) no light adaptation preprocessing is considered in WVDP, (2) adoption of a simplified definition of subband contrast, and (3) the CSF is assumed to have applied directly in the transform domain.

In WVDP, both the original image and noisy image are processed in the three stages before a final probability summation is carried out as outlined in Figure 3.13. During the first stage, discrete wavelet transform is applied to both the original and noisy images. Their output are processed by the threshold elevation (TE) function at the second stage. The TE function determines the amount of quantisation error that can be added without the error being visible after the image is reconstructed. The TE is defined as,

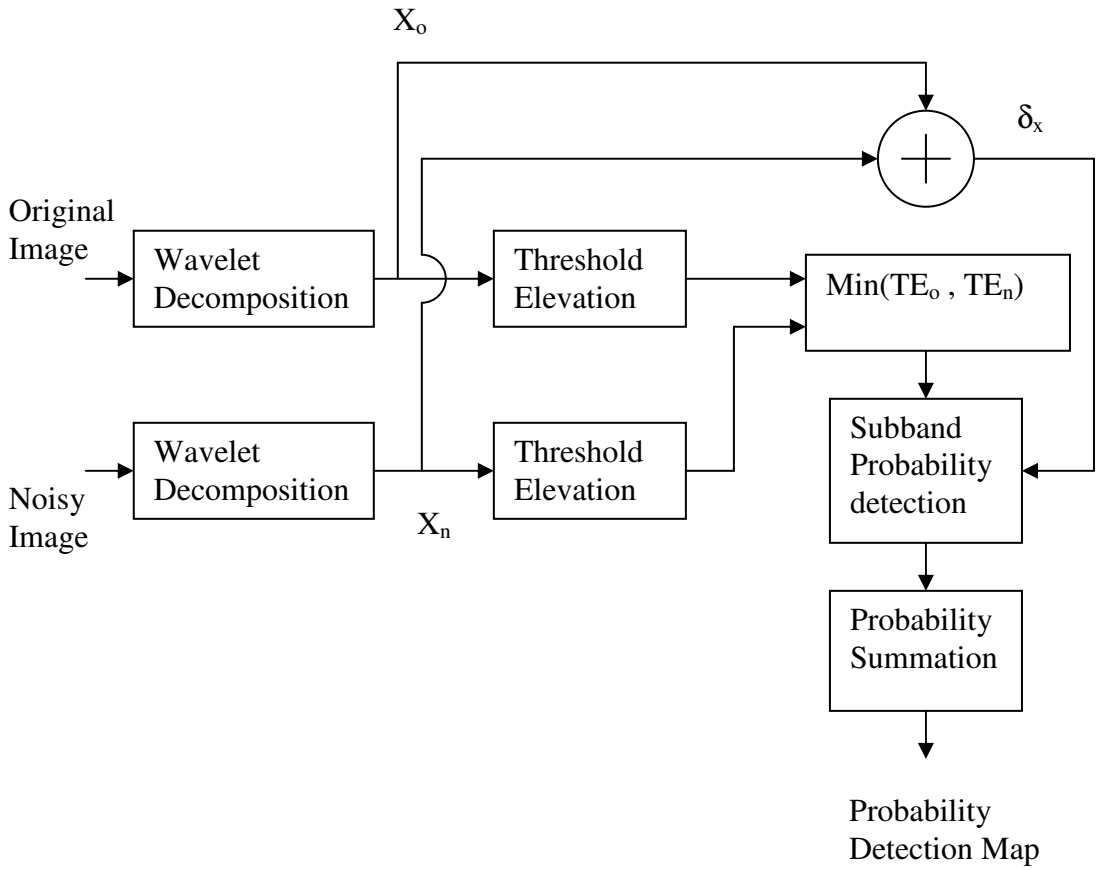



Figure 3.13 The structure of wavelet visible difference predictor

$$TE(\theta, f, m, n) = \max\left(d_c(\theta, f), S_m(f) \cdot |X(m,n)|\right) \tag{3.58}$$

where $\theta$ and $f$ denote the orientation (LL,LH,HL,HH) and the frequency level of decomposition, $X(m,n)$ is the wavelet coefficient at location $m$ and $n$. The $S_m(f)$ is a constant variable that varies according to the frequency, $f$, of the decomposition. $S_m(f)$ can alter the slope of the masking function. For the current model, $S_m(f)$ has been set to one, which corresponds to the derived slope for phase-incoherent masking mentioned in Daly [133]. Together, $S_m(f) \cdot |X(m,n)|$ acts like self masking. $d_c(\theta, f)$ is a coefficient detection threshold defined as,

$$d_c(\theta, f) = \frac{y(\theta, f)}{k_\theta \cdot p_l^{2(l-1)}} \tag{3.59}$$

where $l$ is the decomposition level of the wavelet transform. $k_\theta$ is either $p_l^2, p_h^2$, or $p_l . p_k$ for LL, HH, or LH/HH subband, respectively. The maximum values of $p_l$ and $p_h$ are 0.788485 and 0.852699, respectively. The denominator in equation (3.59) acts like energy gain factors of a wavelet transform and is used to normalized the minimum threshold elevation function, $y(\theta, f)$. The minimum threshold elevation function, $y(\theta, f)$, is obtained from empirical model [122] in psychophysical experiments of noise added directly to wavelet coefficients and viewed from a gamma corrected monitor. $y(\theta, f)$ has the following expression,

$$y(\theta, f) = a \cdot 10^{k\left(log\left(\frac{f}{g_\theta f_0}\right)\right)^{2\,2}} \tag{3.60}$$

where $a$, $k$, $f_o$ are constants having values of 0.495 (minimum), 0.466, and 0.401, respectively. $g_\theta$ has values of 1.501, 1, and 0.534 for LL, LH/HL, and HH subbands, respectively. Equation (3.60) and the values for $a$, $k$, $f_o$ and $g_\theta$ are consistent with those proposed by Liu e. al. in section 3.5.9.

The third stage accounts for mutual masking between the threshold elevations (TE) of both the original ($TE_o$) and noisy ($TE_n$) images by taking the minimum of the two, $T_{em}(\theta, f, m, n) = min(TE_o(\theta, f, m, n), TE_n(\theta, f, m, n))$.

The probability, $P_b(m,n)$, of detecting the visible difference in each subband for each coefficient at location $(m,n)$ is computed as

$$P_b(m,n) = 1 - e^{-\left|\frac{\delta_x(m,n)}{T_{em}(m,n)\cdot\alpha}\right|^{\beta}} \tag{3.61}$$

where $\delta_x(m,n) = X_o(m,n) - X_n(m,n)$, $\beta$ and $\alpha$ are constants having values of 2 and 4, respectively. $X_o$ and $X_n$ are the transform coefficients of the original and noisy images, respectively.

The final output of WVDP is a probability detection map of each pixel at location $(m,n)$. It is computed by combining the probability of detection in each of the subbands as follows,

$$P_{WVDP}(m,n) = 1 - \prod_b (1 - P_b(m,n)) \tag{3.62}$$

Due to aliasing and reduced spatial resolution associated with critical sampling, the critically sample version of WVDP is less accurate when predicting the masking function than the overcomplete version of WVDP. Moreover, the use of 9/7 wavelet transform in WVDP may not be as suitable as the cortex transform, (used in the VDP), for modeling the HVS.

Although, the WVDP is not as reliable and accurate as the cortex transform based VDP, the WVDP can potentially be used to provide a quantitative measure of visual quality in wavelet based coders that do not use the cortex transform. As suggested by Bradley [134], the WVDP can be used to provide a framework for setting a perceptual error below certain visual threshold across the image, so that a wavelet based

compression scheme could operate within this constraint to achieve perceptually lossless compression.

### 3.5.8 JND in DCT Subband Domain by Lin

A JND model incorporating CSF, luminance adaptation, intra-band and inter-band frequency masking effects based on the HVS was proposed by Lin [136] to compute a distortion measure in the DCT domain. The JND, $s(n,k,l)$, is defined as,

$$s(n,k,l) = t_{s-csf}(n,k,l)\prod_{\wp}\alpha_{\wp}(n,k,l) \tag{3.63}$$

where $s(n,k,l)$ is the JND for a DCT subband, $t_{s-csf}(n,k,l)$ is the base threshold due to CSF, and $\alpha_{\wp}(n,k,l)$ is the elevation parameter for all the $\wp \in \{lum, intra, inter\}$ due to luminance adaptation, intra-band frequency masking and inter-band frequency masking. $n$ denotes the position of a $NxN$ DCT block in an image, $X$, and $(k,l)$ denotes the position of a DCT coefficient within a DCT block. The base threshold, $t_{s-csf}(n,k,l)$, is based on a modification of the formula developed by Ahumada et al. [122], and can be traced back to Van Nes and Boudman's experiments on CSF [64]. The formula is modified to avoid over estimation of the base threshold for coefficients in DCT subband at position $(n,k,l)$. The base threshold is computed as

$$t_{s-csf}(n,k,l) = \frac{G}{\phi_k \phi_l (L_{max} - L_{min})} T^o(n,k,l) \tag{3.64}$$

where

$$log\, T^o(n,k,l) = log \frac{b.T_{min}(n)}{r + (1-r)cos^2\theta(k,l)} + K(n)\cdot(log\, f(k,l) - log\, f_p(n))^2 \tag{3.65}$$

$$f(k,l) = \frac{1}{2N}\sqrt{\frac{k^2}{\omega_x^2} + \frac{l^2}{\omega_y^2}} \tag{3.66}$$

$$\theta(k,l) = sin^{-1} \frac{2 \cdot f(k,0) \cdot f(0,l)}{f^2(k,l)} \tag{3.67}$$

where $L_{max}$ and $L_{min}$ are the maximum and minimum display luminance values, $\omega_x$ and $\omega_y$ are the horizontal and vertical visual angles of a pixel. $f_p$ is the spatial frequency at which the minimum CSF threshold ($T_{min}$) occurs. $K(n)$ is a positive constant that be empirically determined as reported in [136]. $r$ is set to 0.7. The normalizing coefficients $\phi_k$ and $\phi_l$ of equation (3.64) can be determined as follows,

$$\phi_m = \begin{cases} \sqrt{\frac{1}{N}} & ,m = 0 \\ \sqrt{\frac{2}{N}} & ,otherwise \end{cases}, \tag{3.68}$$

where $N = 8$, and $r \in \{k,l\}$.

As reported by Lin, the luminance adaptation in digital images is affected by the ambient illumination on the display and the gamma correction of the display tube. With gamma correction, the luminance adaptation is computed as,

$$\alpha_{lum}(n,0,0) = \begin{cases} k_1\left(1 - \frac{2X(n,0,0)}{G \cdot N}\right)^3, & \frac{X(n,0,0)}{G \cdot N} \le \frac{1}{2} \\ k_2\left(\frac{2X(n,0,0)}{G \cdot N} - 1\right)^2, & otherwise \end{cases} \tag{3.69}$$

where $k_1$ and $k_2$ are constants values associated with $X(n,0,0) = 0$ and $X(n,0,0) = G \cdot N$, respectively. $G$, $N$, and $X(n,0,0)$ are the maximum number of grey-level, the size of DCT block, and DC coefficient at the $n^{th}$ DCT block, respectively. Note that the a constant grey value is not used as it tends to underestimate the visibility threshold at dark region.

The intra-band frequency masking, $\alpha_{intra}$, computed as

$$\alpha_{intra}(n,k,l) = max\left(1, \left|\frac{X(n,k,l)}{t_{s-csf}(n,k,l)\cdot\alpha_{lum}(n,0,0)}\right|^{\varsigma}\right) \tag{3.70}$$

where the exponent, $\varsigma$, varies from 0 to 1.

The current model of the inter-band frequency masking, $\alpha_{inter}$, is determined by,

$$\alpha_{inter}(n) = \begin{cases} 1, & \textit{for Low Masking blocks} \\ \delta_1, & \textit{for Medium Masking blocks and } R_L(n)+R_M(n) \le R_o \\ \delta_2, & \textit{for Medium Masking blocks and } R_L(n)+R_M(n) > R_o \\ \left(1+\dfrac{E_{mh}(n)-\mu_2}{2\cdot\mu_3-\mu_2}\right)\cdot\delta_2, & \textit{otherwise} \end{cases} \tag{3.71}$$

The inter-band frequency masking, $\alpha_{inter}$, for the $n^{th}$ DCT block depends on whether the $n^{th}$ DCT block belongs to a Low, Medium, or High Masking block. Classification of the DCT block as either a Low, Medium or High Masking block is determined by the process outlined below,

Firstly, for the $n^{th}$ DCT block, the medium-frequency (MF) and high-frequency (HF) energy, $E_{mh}(n)$, is defined as

$$E_{mh}(n) = R_M(n) + R_H(n) \tag{3.72}$$

and the relative low-frequency (LF) strength, $\tilde{E}_d(n)$, is defined as

$$\tilde{E}_d(n) = \frac{\overline{R}_L(n)}{\overline{R}_M(n)} \tag{3.73}$$

and the relative LF and MF strength, $\tilde{E}_{dm}(n)$, is defined as

$$\tilde{E}_{dm}(n) = \frac{\overline{R}_L(n)+\overline{R}_M(n)}{\overline{R}_H(n)} \tag{3.74}$$

where $R_L(n)$, $R_M(n)$ and $R_H(n)$ are the sums of the absolute DCT coefficients in the LF, MF, and HF groups, respectively. The LF, MF an HF groups are similar to those in [137]. Their corresponding means are $\overline{R}_L(n)$, $\overline{R}_M(n)$ and $\overline{R}_H(n)$, respectively.

A DCT block is assigned to one of these classes (i.e., Low Masking, Medium Masking, or High Masking Class) according to the following rules:

1. For $E_{mh}(n) \le \mu_1$ : the DCT block belongs to Low-Masking class.

2. For $\mu_1 < E_{mh}(n) \le \mu_2$ : if condition (3.75) or (3.76) is met, the DCT block belongs to Medium Masking class; otherwise it belongs to Low-Masking class.

3. For $\mu_2 < E_{mh}(n) \le \mu_3$ : if condition (3.75) or (3.76) is met, the DCT block belongs to Medium Masking class; otherwise it belongs to High-Masking class.

4. For $E_{mh}(n) > \mu_3$ : if condition (3.75) or (3.76) is met for $\varphi_\tau = \tau \cdot \varphi$ and $\chi_\tau = \tau \cdot \chi$ (where $\tau < 1$), the DCT block belongs to Medium Masking class; otherwise it belongs to High-Masking class.

Conditions:

$$\tilde{E}_{dm}(n) \ge Q \tag{3.75}$$

$$max\{\tilde{E}_d(n), \tilde{E}_{dm}(n)\} \ge \varphi \qquad and \qquad min\{\tilde{E}_d(n), \tilde{E}_{dm}(n)\} \ge \chi \tag{3.76}$$

where the model parameters for determining $\alpha_{inter}$ are set as $\mu_1 = 125$, $\mu_2 = 290$, $\mu_3 = 900$, $\varphi = 7$, $\chi = 5$, $\tau = 0.1$, $Q = 16$, $R_o = 400$, $\delta_1 = 1.125$ and $\delta_2 = 1.25$.

Together with the conditions specified in equations (3.75) and (3.76), the model parameters ($\mu_1$, $\mu_2$, $\mu_3$) are use as for either lower or upper ranges for the medium-frequency and high-frequency energy, $E_{mh}(n)$, so that the $n^{th}$ DCT block can be

classified as either belonging to low-masking, medium-masking or high-masking block. Once the block is classified, the interband frequency masking, $\alpha_{inter}(n)$, for the $n^{th}$ DCT block can be computed as in equation (3.71) according to the block classification and its low-frequency and medium-frequency energies.

### 3.5.9 Perceptual Distortion Metric by Liu et al.

Liu et al. propose a Perceptual Distortion Metric [138] for the JPEG2000 coder with a quality-driven encoding scheme. The distortion metric is computed based on JND threshold, which modelled the HVS with contrast sensitivity function (CSF), luminance masking adaptation and contrast masking adaptation. The JND threshold in this instance is defined as,

$$t_{JND}(l,\theta,m,n) = JND(l,\theta)\cdot M_L(l,\theta,m,n)\cdot M_C(l,\theta,m,n) \tag{3.77}$$

where $JND(l,\theta)$, $M_L(l,\theta,m,n)$ and $M_C(l,\theta,m,n)$ are the base JND detection threshold, luminance masking adjustment, and contrast masking adjustment, respectively for subband $(l,\theta)$ at spatial location ($m$,$n$). Variables $l$ and $\theta$ specify the frequency and orientation (i.e., the LL, LH, HL, HH orientation), respectively. The $JND(l,\theta)$ was acquired through data fitting of experimental data. It is expressed as,

$$JND(l,\theta) = \frac{a}{A(l,\theta)}\cdot 10^{k\left[log\left(\frac{g_\theta f_o \cdot 2^\lambda}{r}\right)\right]^2} \tag{3.78}$$

where $A(l,\theta)$ is the amplitude of the wavelet 9/7 basis functions for subband $(l,\theta)$ (Table 3.1), and $r = dv\cdot tan\left(\frac{\pi}{180}\right) \approx \frac{dv}{57.3}$ is the visual resolution of the display in pixel per degree. The $d$ and $v$ are the display resolution in pixel/cm and viewing distance in cm, respectively. The parameters, $a$, $k$, $g_\theta$, $f_o$, are obtained through data fitting and listed in Table 3.2. The $JND(l,\theta)$ in equation (3.78) is essentially the same model used for $d_c(\theta,f)$ in equation (3.59) (note that one needs to substitute

equation (3.60) into equation (3.59) in order to observe their similarity). Consequently, the values for $a$, $k$, $g_\theta$ and $f_o$ are the same for both WVDP model in section 3.5.7 and the model presented in this section.

| Orientation $\theta$ | Decomposition Level, $l$ | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| LL | 0.62171 | 0.34537 | 0.18004 | 0.09140 | 0.045943 | 0.023013 |
| LH, HL | 0.67234 | 0.41317 | 0.22727 | 0.11792 | 0.059758 | 0.030018 |
| HH | 0.72709 | 0.49428 | 0.28688 | 0.15214 | 0.077727 | 0.039156 |

Table 3.1: $A(l,\theta)$ for wavelet 9/7 basis functions.

| | | | $g_\theta$ | | |
|---|---|---|---|---|---|
| | | | Orientation, $\theta$ | | |
| $a$ | $k$ | $f_o$ | $\theta = LL$ | $\theta = HL, LH$ | $\theta = HH$ |
| 0.495 | 0.466 | 0.401 | 1.501 | 1.0 | 0.534 |

Table 3.2: The constant parameters for the base JND threshold, $JND(l,\theta)$.

The luminance masking adjustment accounts for the HVS response that depends not so much on absolute luminance, but more on the luminance variation relative to the surrounding background. This phenomena can be described by the Weber-Fechner law [139]. The luminance masking adjustment is approximated by,

$$M_L(l,\theta,m,n) = \left( \frac{X_{LL}(m',n')}{\mu_L} \right)^{a_T} \tag{3.79}$$

where $X_{LL}(m',n')$ is the wavelet coefficient in the LL band that corresponds spatially to location $(l,\theta,m,n)$ whereby $m' = \lfloor m / 2^{l_{max}-l} \rfloor$ and $n' = \lfloor n / 2^{l_{max}-l} \rfloor$, and $\mu_L = 128$ is the mean luminance of the display for an unsigned 8-bit image. The exponent, $a_T$, has a value of 0.649.

The contrast masking adjustment accounts for the fact that the visibility of visual signal can be affected (i.e., reduce or enhance) by the presence of other visual patterns. Here the contrast masking adjustment, $M_C(l,\theta,m,n)$, includes two factors, the self masking and masking due to neighbouring visual signals. It is expressed as,

$$M_C(l,\theta,m,n) = M_{self}(l,\theta,m,n) \cdot M_{neighbor}(l,\theta,m,n) \tag{3.80}$$

The self masking, $M_{self}(l,\theta,m,n)$, is expressed as,

$$M_{self}(l,\theta,m,n) = max\left\{1, \left(\frac{|X(l,\theta,m,n)|}{JND(l,\theta) \cdot M_L(l,\theta,m,n)}\right)^{\gamma}\right\} \tag{3.81}$$

where $X(i,\theta,m,n)$ is the wavelet coefficient of subband $(l,\theta)$ at location $(m,n)$, and the exponent, $\gamma$, is set at a value of 0.6. The neighourhood masking adjustment for subband $(l,\theta)$ at location $(m,n)$ is expressed as,

$$M_{neighbor}(l,\theta,m,n) = \max\left\{1, \sum_{u \in neighbors\ of\ X(l,\theta,m,n)} \frac{1}{N_{m,n}} \cdot \left|\frac{X_u}{JND(l,\theta) \cdot M_L(l,\theta,m,n)}\right|^{\varphi}\right\} \tag{3.82}$$

where all the elements specified by $X_\mu$ are neighbourhood coefficients with location $(l,\theta,m,n)$ being at its centre, $\varphi$ is a constant parameter. The total number of neighbourhood coefficients is specified by $N_{m,n}$ for subband $(l,\theta)$ at location $(m,n)$.

For the HVS, the fovea region has the highest cone concentration, and hence has the highest visual acuity. This region covers about two degree of visual angle. Hence the distortion is computed by considering the spatial region, $F(n_1,n_2)$, in the image domain that is covered by the fovea region. Consequently, the number of coefficients in $F(n_1,n_2)$ can be approximated by,

$$N(F(n_1,n_2)) = \left( \left\lfloor 2dv\ tan\left(\frac{2^o}{2}\right) \right\rfloor \right)^2 \approx (\lfloor 2r \rfloor)^2 \tag{3.83}$$

where $r$ is the visual resolution for the display in pixels per degree. The distortion appears in the form of the Minkowski metric as follows,

$$D_{F(n_1,n_2)} = \left( \sum_{(l,\theta,m,n)\in F(n_1,n_2)} \left| \frac{e_q(l,\theta,m,n)}{t_{JND}(l,\theta,m,n)} \right|^{\beta} \right)^{\frac{1}{\beta}} \tag{3.84}$$

Where $e_q(l,\theta,m,n)$ is the quantisation error at location $(l,\theta,m,n)$. The distortion measure, $D$, is determined by considering the highest probability of detecting a distortion over all possible fovea region over the entire image. This corresponds to the expression below,

$$D = \max_{(n_1,n_2)} \{ D_{F(n_1,n_2)} \} \tag{3.85}$$

For a given target distortion, $D_t$, the minimum bitrate can be determined by ensuring all $D_{F(n_1,n_2)} = D_t$ is met.

### 3.5.10 Perceptual Image Distortion Metric by Tan et al.

The Perceptual Image Distortion Metric (PIDM) proposed by Tan et al. [15] is based on the Contrast Gain Control (CGC) model of Watson and Solomon [27], and the model proposed by Teo and Heeger [23]. The PIDM employs CSF, intra-band masking, and inter-orientation masking of similar frequencies to model the HVS. It is adapted into the EBCOT encoding framework [14]. From subjective test results, the PIDM produces better perceived visual quality of digital monochrome images when compared to those that used the MSE measure. The PIDM uses the Daubechies 9/7 biorthogonal filter set for its frequency decomposition in a dyadic structure. There are three stages in the CGC model:

Stage 1: Dyadic transform with Daubechies 9/7 bi-orthogonal filters [140] is used to approximate the frequency and orientation selective nature of the HVS. (Note that cortex transform [141] will produce a more accurate model for the HVS),

Stage 2: The effect of contrast sensitivity is accounted for via a set of weights to adjust the wavelet coefficients according to the sensitivity of the HVS at various spatial frequencies,

Stage 3: Intra-band masking and inter-orientation masking are considered and are represented by inhibition functions.

The neural response, $R_z$, is defined as,

$$R_z(l,\theta,m,n) = k_z . \frac{E_z(l,\theta,m,n)}{I_z(l,\theta,m,n) + \tau_z^q} \tag{3.86}$$

where $z \in \{\Gamma, \Theta\}$, with $\Gamma$ and $\Theta$ denoting intra-band spatial masking domain and inter-orientation masking domain of the similar frequency coefficients, respectively. $E_z$ and $I_z$ are the excitation and inhibition functions for the two domains in $z \in \{\Gamma, \Theta\}$. $k_z$ and $\tau_z$ are the scaling and saturation constants, respectively. The term, $\tau_z > 0$, has been added to provide minimum level of inhibition. $(l,\theta,m,n)$ denotes the location of the wavelet coefficient relative to spatial location ($m$, $n$), resolution ($l$) and orientation ($\theta$) within a codeblock, note that $l = \{1,2,...,5\}$ being the frequency level and $\theta = \{LH, HL, HH\}$ being the orientation band. The $E_z$ and $I_z$ for $z \in \{\Gamma, \Theta\}$ are defined as,

$$E_\Gamma(l,\theta,m,n) = (X_w[l,\theta,m,n])^{p_\Gamma} \tag{3.87}$$

$$E_\Theta(l,\theta,m,n) = (X_w(l,\theta,m,n))^{p_\Theta} \tag{3.88}$$

$$I_\Gamma(l,\theta,m,n) = \frac{8}{N(l)} \cdot \sum_{u=m-l}^{m+l} \sum_{v=n-l}^{n+l} (X_w[l,\theta,u,v])^q + \sigma_{\text{var}}^q(l,\theta,m,n) \tag{3.89}$$

$$I_{\Theta}(l,\theta,m,n) = \sum_{\phi \in \{LH,HL,HH\}} (X_w[l,\phi,m,n])^q \tag{3.90}$$

where

$$\sigma_{\text{var}}^q(l,\theta,m,n) = \frac{1}{N(l)} \cdot \sum_{u=m-l}^{m+l} \sum_{v=n-l}^{n+l} (X_w[l,\theta,u,v] - \mu(m,n))^2 \tag{3.91}$$

$X_w[l,\theta,m,n]$ is the CSF weighted wavelet coefficient, and $q$ is set at 2. The inhibition function, $I_{\Gamma}(l,\theta,m,n)$, consists of two components: (1) spatial masking that is computed based on a square neighbourhood area around the $X_w[l,\theta,m,n]$, with the area being, $N(l) = (2l+1)^2$, and (2) the texture masking that is computed by the neighbourhood variance, $\sigma_{var}^q$, in equation (3.91). $\mu(m,n)$ represents the mean of the square neighbourhood area. At very high activity region of an image, the HVS is more tolerable to noise. Therefore, the texture masking is included in addition to spatial masking to account for the HVS's ability to tolerate higher distortion at very high activity region, where tolerance to higher distortion could not be sufficiently accounted for by spatial masking alone.

At the lowest frequency subband (i.e., the isotropic LL (DC) band) where very little or no masking is envisaged, the response is computed differently and is defined as,

$$R_{dc}(1,LL,m,n) = k_{\Gamma} \cdot \frac{\left(\tilde{X}_w(1,LL,m,n)\right)^{p_{\Gamma}}}{\left(X_w(1,LL,m,n)\right)^q + \tau_{\Gamma}^2} \tag{3.92}$$

where $\tilde{X}_w$ and $X_w$ are the quantised and unquantised DC coefficients, respectively. The distortion for individual neural response is defined as follows,

$$D_z(l.\theta,m,n) = \left| R_{\alpha,z}(l,\theta,m,n) - R_{\beta,z}(l,\theta,m,n) \right|^2 \tag{3.93}$$

where $R_{\alpha,z}$ and $R_{\beta,z}$ are from the reference and processed images, respectively. The final distortion measure for codeblock, $b$, is the sum of all intra-band and inter-orientation maskings,

$$D_{l,\theta}(b) = \sum_i \sum_j \left( g_\Gamma \cdot D_\Gamma(l,\theta,i,j) + g_\Theta \cdot D_\Theta(l,\theta,m,n) \right) \tag{3.94}$$

where $i = \{1,2,..,M_b\}$ and $j = \{1,2,...,N_b\}$ are the row and column positions for the codeblock, $b$. The various model parameter constants are listed in Table 3.3 below.

| Parameters | | Parameters | |
|---|---|---|---|
| CSF (LL-band) | 1.4800 | $k_\Theta$ | 0.9876 |
| CSF $(l=1)$ | 1.5500 | $\tau_\Gamma$ | 5.5550 |
| CSF $(l=2)$ | 1.7700 | $\tau_\Theta$ | 7.6800 |
| CSF $(l=3)$ | 1.6800 | $p_\Gamma$ | 2.5800 |
| CSF $(l=4)$ | 1.2900 | $p_\Theta$ | 2.3950 |
| CSF $(l=5)$ | 0.8050 | $g_\Gamma$ | 0.7588 |
| $k_\Gamma$ | 1.0888 | $g_\Theta$ | 0.4834 |

Table 3.3: Vision Model Parameters.

### 3.5.11 Just Noticeable Colour Difference Model by Chou and Liu

Chou and Liu [142] proposed a visual model for measuring perceptual redundancy inherent in colour images. The proposed model can be adapted in the JPEG-LS and JPEG2000 compliance coders. According to Chou and Liu [142], the perceptual redundancy of a particular colour can be determined by the radius of just noticeable colour difference (JNCD) in all regions of the uniform colour space. The radius of JNCD sphere is scaled by both the chroma of the associated pixel and the local luminance properties, and it is expressed as adaptive JNCD (AJNCD) as,

$$AJNCD = JNCD_{Lab} \cdot \alpha_{lum}(E(L), \Delta L) \cdot \alpha_C(a,b) \tag{3.95}$$

where $L$, $a$, and $b$ are components specified in the CIE-Lab Space. $JNCD_{Lab}$ is the threshold for determining if two colours in the CIE-Lab space are considered perceptually distinguishable if their Euclidean distance between them exceeds this threshold. $\alpha_{lum}$ and $\alpha_C$ are scaling factors which consider the effect of chroma changes and masking effect due to local luminance texture, respectively. The scaling factor, $\alpha_C$, is determined as,

$$\alpha_C(a,b) = 1 + 0.045 \cdot \sqrt{a^2 + b^2} \tag{3.96}$$

The masking factor, $\alpha_{lum}$, due to local luminance texture, is defined as,

$$\alpha_{lum}(E(L), \Delta L) = \tau(E(L)) \cdot \Delta L + 1.0 \tag{3.97}$$

where $E(L)$ and $\Delta L$ are mean background luminance of the target pixel and the maximum luminance gradient across the target pixel, respectively. $\tau(E(L))$ is the slope of the lines that fit the empirical data under different ranges of $E(L)$, and it is determined as,

$$\tau(E(L)) = \begin{cases} 0.09 & ,E(L) \le 20 \\ 0.07 & ,21 < E(L) \le 40 \\ 0.05 & ,41 < E(L) \le 60 \\ 0.08 & ,61 < E(L) \le 100 \end{cases} \tag{3.98}$$

A lower colour bound, $k_1$, and upper colour bound, $k_2$, for colour, $k$ ,have been defined so that only colours within the *AJNCD* sphere and those which have luminance components between the colour bounds $k_1$ and $k_2$ are included as candidates for estimating the perceptual redundancy for colour, *k*. As consideration for all colours within the *AJNCD* sphere can be prohibitively large, only limited numbers of critical colours that are at the verge of being distinguishable from colours are selected for setting the lower and upper bounds, and the critical colour samples are chosen as,

$$\varphi_1(k) = \{L_{k_1}, a_k + Re(E_1 \cdot e^{i\theta}), b_k + Im(E_1 \cdot e^{i\theta})\} \tag{3.99}$$

$$\varphi_2(k) = \{L_{k_2}, a_k + Re(E_2 \cdot e^{i\theta}), b_k + Im(E_2 \cdot e^{i\theta})\} \tag{3.100}$$

where

$$E_1 = \sqrt{AJNCD_k^2 - \left|L_{k_1} - L_k\right|^2} \tag{3.101}$$

$$E_2 = \sqrt{AJNCD_k^2 - \left|L_{k_2} - L_k\right|^2} \tag{3.102}$$

$$\theta = \left\{\gamma \mid \gamma mod\left(\frac{360^\circ}{n}\right) = 0\right\} \tag{3.103}$$

where $L_k$ , $L_{k_1}$ and $L_{k_2}$ are the luminance levels for colours $k$ , $k_1$ and $k_2$ , respectively. The $AJNCD_k$ is adaptive *JNCD* for colour, $k$ . $n$ is the number of critical colour samples for $k_1$ and $k_2$ .

The JND value for each colour component $c \in \{Y, Cb, Cr\}$ for colour *k* is computed as,

$$JND_c(k) = \min_{s \in \varphi_1(k) \cup \varphi_2(k)} \left|c_s - c_k\right| \tag{3.104}$$

To incorporate into the JPEG2000 compliance coder, the distortion measure that is used in the post compression rate distortion optimisation is defined as perceptible distortion,

$$D_{c,i}^{n_{c,i}(\lambda)} = \sum_{(u,v) \in B_{c,i}} \left[\left|X_{c,i}(u,v) - \hat{X}_{c,i}^{n_{c,i}(\lambda)}(u,v)\right| - JND_{c,i}(u,v)\right]^2 \cdot \delta_{c,i}^{n_{c,i}(\lambda)}(u,v) \tag{3.105}$$

$$\delta_{c,i}^{n_{c,i}(\lambda)}(u,v)=\begin{cases}1 & ,\left|X(u,v)-\hat{X}_{c,i}^{n_{c,i}(\lambda)}(u,v)\right|>JND_{c,i}(u,v)\\ 0 & ,\left|X(u,v)-\hat{X}_{c,i}^{n_{c,i}(\lambda)}(u,v)\right|\le JND_{c,i}(u,v)\end{cases} \tag{3.106}$$

where $B_{c,i}$ is the set of sequences in code block $i$ of colour component $c\in\{Y,Cb,Cr\}$, $X_{c,i}(u,v)$ is the wavelet coefficient at location $(u,v)$ within code block $i$ of colour component $c\in\{Y,Cb,Cr\}$, and $\hat{X}_{c,i}^{n_{c,i}(\lambda)}(u,v)$ is the reconstruction of $X_{c,i}(u,v)$ by the bit streams truncated at truncation point $n_{c,i}(\lambda)$ at optimal rate-distortion slope $\lambda$, which is obtained via the rate-distortion optimisation procedure in the JPEG2000 compliance coder. $JND_{c,i}(u,v)$ is the $JND$ value obtained as in equation (3.101) for colour component $c\in\{Y,Cb,Cr\}$ for sample $X_{c,i}(u,v)$ belonging to code block $i$.

### 3.5.12 Comparison of Some Perceptual Coders

In sections 3.5.1 to 3.5.11, some perceptual image coders are discussed in detail. The visual properties and features of different perceptual coders are summarised and tabulated in Table 3.4.

| **Perceptual Models** | **Visual properties considered** | **Feature** |
|---|---|---|
| Watson DCTune | Contrast and Luminance masking | Selection of a quantisation matrix that can yield the best quality given the desired compression ratio. |
| Safranek and Johnson | Luminance variations for the purpose of textual masking | Coding of images with rates of less than one bit per pixel. Achieved using a combination of the following compression method: DPCM, entropy coding, perceptual-threshold calculation and quiescent block rejection. |

Table 3.4 Comparison of Some Perceptual Coders

| Perceptual Models | Visual properties considered | Feature |
|---|---|---|
| Chou & Li | Average luminance difference and mean background around pixel | Proposes a method to estimate the Just-Noticeable-Distortion (JND) and Minimally-Noticeable-Distortion (MND) profiles of a monochromatic image.<br>The decomposed JNDs and MNDs in the subbands are used in encoding to achieve the desired bitrate and quality. |
| Hontsch and Karam | Background-luminance adjusted contrast sensitivity, contrast masking | Uses adaptive quantisation scheme with DPCM coding and JND threshold for DCT coefficients. |
| Taubman | Visual masking | Proposes a spatially varying distortion metric that incorporates masking phenomenon within the distortion function of EBCOT. Masking effects are considered after quantisation. |
| Zeng, Daly and Lei | Intra-band masking, Self-contrast masking and neighbourhood masking effects | Incorporates self-contrasting masking and neighbourhood masking effects by introducing a non-linear function that maps the wavelet coefficients into a perceptual domain. Masking effects are considered by applying a signal (which is subjected to masking) to a power function and followed by a divisible neighbourhood masking factor. |
| Liu | Contrast sensitivity, luminance masking, contrast masking | Proposes a distortion metric based on JND thresholds (which incorporates CSF, luminance and contrast masking adaptation) in the wavelet domain in a dyadic structure with Daubechies 9/7 filters. |
| Tan | Contrast sensitivity, intra-band frequency spatial masking, inter-orientation masking of similar frequencies | Considers the CSF, intra-band frequency spatial masking, inter-orientation masking of similar frequencies within a Contrast Gain Control Mode [27] that is adapted into the EBCOT framework[14]. |
| Chou and Liu | Local luminance masking | Incorporates adaptive JNCD into the distortion function for JPEG2000 compliance coder. Considers the effects of chroma variation and luminance properties on adaptive JNCD. |

Table 3.4 Comparison of Some Perceptual Coders (cont…).

## 3.6 Chapter Summary

This Chapter reviews the various coders used for image compression (sections 3.4 & 3.5). Section 3.2 gives a brief overview of information theory which forms the basic foundation of data compression including image coding [10]. Picture compression is categorized into lossy and lossless. Lossy compression allows for some information loss during compression. On the other hand, lossless compression maintains information integrity during the encoding process. Lossless compression systems are centred solely on the removal of statistical redundancies which Shannon refers to as noise [10]. For lossy compression, a balance between information loss and compression ratio must be established. Thus, the rate distortion theory is seen as critical component for mitigating the tradeoff between bitrates and distortion, i.e., picture quality versus file size.

Section 3.3 presents the structure of transform based lossy image compression system, which includes data transformation and quantisation. Section 3.4 presents the concept of hierarchical bitplane coding, specifically the EZW [31], the SPIHT [32], and the EBCOT [14] coders. Apart from improved coding efficiency over the DCT based image coder, i.e., JPEG baseline [11], these coders also offer scalability feature. EBCOT has been adopted as the core of JPEG2000 still image coding standard [12].

A comparison of three wavelet based bitplane image coders have been presented, beginning with the EZW [31], then the SPIHT [32] and finally EBCOT [14] coders. Undoubtedly, EZW represents significant contribution and novelty in the design of hierarchical bitplane coders. Subsequent improvement based on this algorithm can be found in [117]. Its popularity has motivated the development of SPIHT [32] and subsequently the EBCOT [14] coders. Monro et al. [118] has also extended the EZW approach to block-based transform coding, where zero-tree coding for DCT coefficients is proposed. Similar to the EZW, for the SPIHT, encoding can halt at any time once the desired coding rate is achieved. However, empirical studies have shown that SPIHT has achieved better coding results than that of EZW and thus is a more efficient coding tool [1, 12, 32]. In the EBOCT algorithm [14], encoding is performed

on partitioned codeblocks. This involves bitplane quantisation with context arithmetic coding. This is in contrast to EZW and SPIHT, where the dependency nature of the subbands means that coding is carried out across scales without subdivision. Experimental results find that the EBCOT is regarded as superior to EZW and SPIHT in terms of its Signal-to-Noise Ratio (SNR) and resolution scalability. Moreover, the JPEG2000 which is based on the EBCOT structure is now hailed as the current state-of-the-art coder. The JPEG2000 coder is also taken as the benchmark for subsequent image coders developed in Chapters four and five to be measured against.

In an effort to improve the perceived quality of coded images, picture coding systems have been incorporated with HVS based models. A review of some of these perceptual models [14, 15, 119, 123, 126, 127, 132, 134, 136, 138, 142] in section (3.5) highlights the visual properties considered by the various perceptual models. Some of these perceptual coders are either rate or quality driven. A review of these models serves as the backdrop for the development of the PCDM model for colour images and the Perceptual Post Filtering (PPF) algorithm presented in chapters 4 and 5. A comparison of the various perceptual models is shown in Table 3.4.

# Chapter 4 Perceptual Coding based on Intra-band and Inter-orientation Masking

## *4.1 Introduction*

The JPEG2000 standard [12] represents the current state-of-the-art coder for still images. The core coding structure of JPEG2000 is the block-based bitplane coding paradigm adopted from the EBCOT [14] that has demonstrated superior performance over other wavelet-based coders. The EBCOT and, hence, the JPEG2000 generate independent bit-streams for each codeblock which are packed into quality layers. In both coders, the delivery of optimized bit stream is the result of rate-distortion optimisation and context arithmetic coding. While applying the Mean Squared Error (MSE) or masking sensitive distortion measure (i.e., the VDM of EBCOT) as the distortion measure in the R-D optimisation produces good quality performance for the coded images, the MSE has long been recognized as being an inadequate measurement of perceived image quality as reported in [143] and [144]. The MSE only measures the raw mathematical distortion and does not take into account the perceived distortions as seen by the human visual system. It is true that while some aspect of vision modeling design such as the CVIS criteria has been incorporated into JPEG2000 software verification model (VM8) for experimental testing, a more comprehensive vision model can be used to improve the visual quality of the coded images.

## *4.2 The Reference Model – JPEG2000 Coding Structure*

The proposed model that is described in subsequent sections is built into the framework of JPEG2000 [14]. Figure 4.1 depicts a pictorial view of the building block in the JPEG2000 structure. The encoding process involves a tier-1 and tier 2 coding.

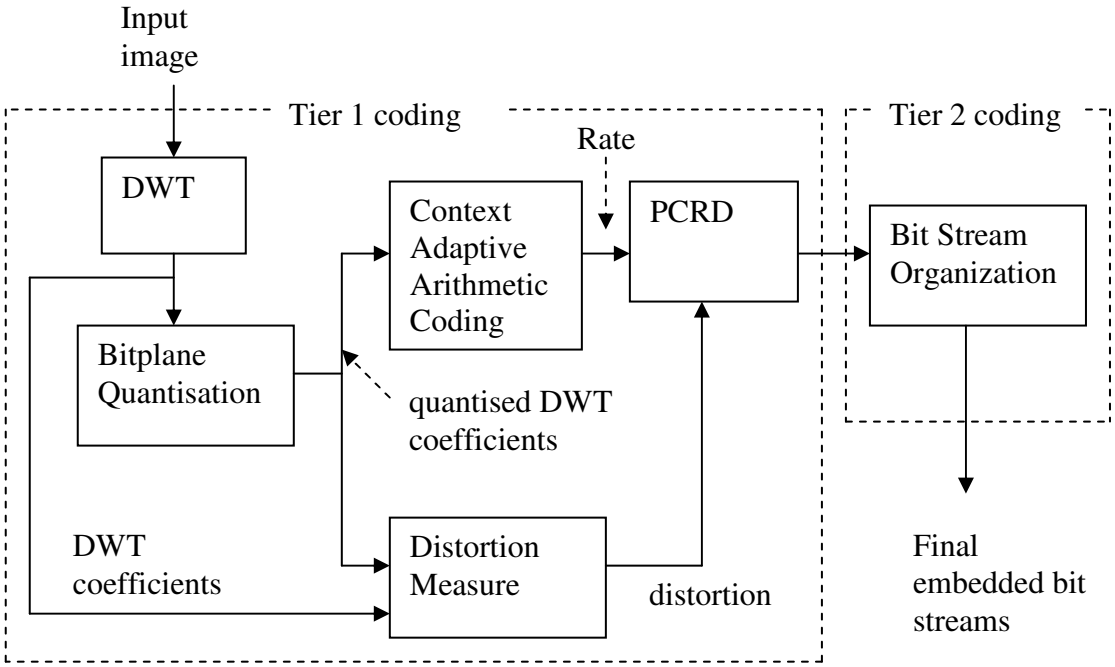



Figure 4.1 Coding Structure of JPEG2000. Tier 1 Coding: The bitplane quantised DWT coefficients and the unquantised coefficients are used to compute the distortions for all coding passes. The bitplane quantized DWT is also entropy coded with context adaptive arithmetic coder. Both distortion reductions and rates for the coding passes are used to generate the embedded bit streams through Post Compression Rate Distortion Optimizer. Tier 2 Coding: The Bit Stream organisation forms the final embedded bit stream.

In the lossy compression mode with irreversible path, the Discrete Wavelet Transform (DWT) or the Lifting Wavelet Transform (LWT) [145, 146] is applied to the image data and decomposes it into a *k*-level multiresolultion representation by Mallat decomposition [103] with the Daubechies 9/7 separable filter set [140], which is the symmetric and linear phase. In both the lossy compression mode with reversible path and the lossless compression mode, a biorthogonal 5/3 integer filter set is used [12] instead. Table 4.1 below and Figure A1 in appendix A show the coefficients and profiles of the 9/7 filter sets, respectively. With relatively short filter lengths, the filters enable relatively fast computational speed. For each decomposition level, each column of a 2-D image is first transformed vertically with a 1-D analysis filter bank, the results of the 1-D transformed coefficients are then transformed horizontally along each row with the same analysis filter bank. For illustration purpose, Figure 4.2 shows the multiresolution of a 5-level DWT decomposition by the Mallat

decomposition [103]. The 5-level decomposition produces one isotropic and 15 oriented subbands at approximately 0 degree orientation for the isotropic band, and 90 and 45/135 degree orientations per level for the other 15 oriented subbands. Note the proposed model operates within the lossy mode with irreversible path of the JPEG2000 structure.

| Filter Taps | Analysis Filter | | Synthesis Filter | |
|---|---|---|---|---|
| | Low Pass, $h$ | High Pass, $g$ | Low Pass, $\overline{h}$ | High Pass, $\overline{g}$ |
| 0 | 0.602949 | -0.557543 | 1.115086 | -1.205898 |
| ±1 | 0.266864 | 0.295636 | 0.591272 | 0.533728 |
| ±2 | -0.078223 | 0.028772 | -0.057544 | 0.156446 |
| ±3 | -0.016864 | -0.045636 | -0.0921272 | -0.033728 |
| ±4 | 0.026749 | 0 | 0 | -0.053498 |

Table 4.1: The Daubechies 9/7 wavelet filter set. (Note: This is the un-normalized version. The normalized version involves a multiplicative factor of $\sqrt{2}$ and $\frac{1}{\sqrt{2}}$ for the analysis filter and synthesis filter, respectively.)

Scalar dead-zone quantisation is applied to the transformed coefficients. In the lossy mode with irreversible path where the Daubechies 9/7 separable filter set is used, the choice of the quantiser step size for each band is relative to the nominal dynamic range of the subband signal.

During tier-1 coding, the quantisation indices produced by the scalar quantisation for each subband are partitioned into code blocks, each of which has typical block size of 64x64. Each code block is then independently coded using bit-plane coding beginning from the most significant bit plane to the least significant bit plane. For each code block, an embedded code is produced, consisting of numerous coding passes. At each bit plane, it involves three coding passes, namely significance pass, refinement pass, and cleanup pass. The samples of each code block are scanned in the same order by the coding passes. In each coding pass, the bit plane encoding process produces a sequence of symbols which may be entropy encoded by context-based adaptive arithmetic coder, specifically, the MQ coder from the JBIG2 standard [147] is used. Each coding pass forms a truncation point. Associated with each coding pass is the rate (in bits) required to generate the coded symbols and the distortion reduction

resulting from encoding the coding pass. The rate increase and the distortion reduction for all truncation points are then used in the Post Compression Rate Distortion (PCRD) optimisation in the tier-2 encoding process to optimize the final bit stream. The distortion criteria used in the JPEG2000 is typically the mean squared error (MSE), or optionally the visual distortion metric (CVIS) in the JPEG2000 software verification model (VM8). However, JPEG2000 standard does not restrict the choice of distortion metric.

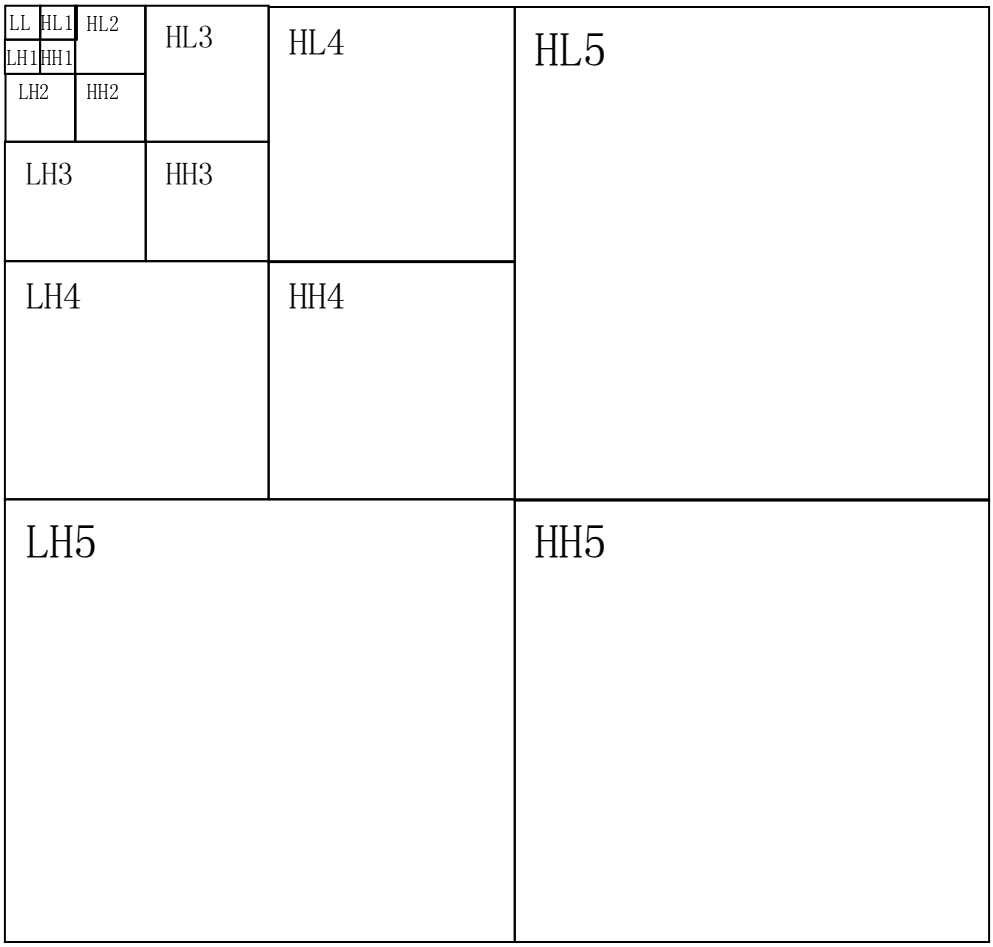



Figure 4.2 A 5-level Multiresolution Mallat decomposition. One Isotropic DC band (LL1), and 15 orientation bands covering 90, 45/135 degrees of orientations, where 1 denotes the lowest frequency level and 5 the highest frequency level.

In tier-2 encoding process, the PCRD optimisation process decides which coding passes to be included or excluded (discarded) from the final bit stream.

The MSE as the distortion metric used in the JPEG2000 coder here (Note it is actually weighted MSE) for a given truncation point $t$ in code block $B_i$ is expressed as,

$$D_{i,MSE}^{(t)} = W_{b_i}^{csf} G_b \sum_{j \in \kappa_t} \left( X_i[j] - X_i^{(t)}[j] \right)^2 \tag{4.1}$$

where $j$ represents the location of the coefficient within the code block $B_i$ for a given truncation point $t$, $\kappa_t$ includes all coefficients within the code block $B_i$ that produces truncation point $t$, $X_i[j]$ is the transform coefficient value, $X_i^{(t)}[j]$ is the bit-plane quantized coefficient value for truncation point $t$, $G_b$ is the squared norm of the synthesis basis vectors for subband $b$ which contains the code block $B_i$, $W_{b_i}^{csf}$ is the CSF energy weighting factor.

The distortion computation according to the CVIS for a given truncation point $t$ in code block $B_i$ is given as,

$$D_{i,CVIS}^{(t)} = \frac{\alpha^2 G_b}{T_b} \frac{\sum_{j \in \kappa_t} \left( X_i[j] - X_i^{(t)}[j] \right)^2}{1 + \left( \frac{1}{T_b} \right)^{2g} \left( \frac{1}{\|N_j\|} \sum_{n \in N_j} |X[n]|^g \right)^2} \tag{4.2}$$

where $T_b$ is the contrast sensitivity thresholds for subband $b$, $N_j$ is the neighbourhood around location $j$ and the neighbourhood is identified with the subblock of size 8x8 that contains location $j$, $\alpha$ is an arbitrary constant, the masking gain , $g$, has a typical value of 0.5.

**Perceptual Colour Distortion Measure**

While both the EBCOT and the JPEG2000 encoding use the mean squared error (MSE) or visual distortion metric (CVIS) as a distortion measure in the R-D optimisation function, the proposed coder uses Perceptual Colour Distortion Measure (PCDM) - mimicking that of the perception of the human visual system (HVS) – as a distortion measure in the R-D optimisation. Specifically, the optical sensitivity at the optical stage of the HVS represented by the response of the contrast sensitivity function (CSF), and the responses of the various masking effects at the cortical stage are considered in the formulation of the PCDM function.

Figure 4.3 gives a pictorial view of the PCDM. Basically, the PCDM is a replacement of the distortion measure used in the JPEG2000 coding structure where the proposed PCDM has been incorporated. From Figure 4.3, both the quantised and raw DWT coefficients are weighted with CSF weights, and the various masking functions are applied to the CSF weighted coefficients to compute the masking responses (i.e., from the raw coefficients and the quantised coefficients). The detection and pooling stage computes the distortion by pooling the error between the two responses.

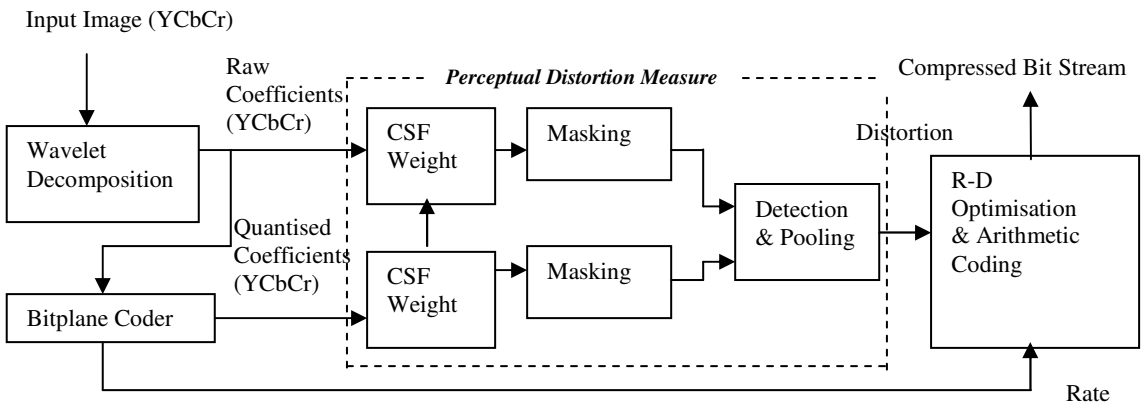



Figure 4.3 The JPEG2000 Coding Structure with the proposed PCDM replacing the MSE criterion.

## *4.3 Proposed Vision Model*

Several HVS based models have gained increasing acceptance as in [148] and [149]. The coverage of HVS perception and some of these HVS model based coders are explained in chapters 2 and 3, respectively. For simplicity, an HVS can be modeled by two successive and separate stages: optical and cortical. The optical stage is concerned with the limitation of the sensitivity of the human optical system relative to background luminance and spatio-temporal frequencies. Discussion of some of the properties of the human optics and the cortical stages of the HVS can be found in chapter 2 of this thesis.

### 4.3.1 The optical stage

The optical sensitivity has been described by Van Nes and Bouman [64] as the "contrast sensitivity function" (CSF). The CSF possesses the characteristic of a band-pass filter. The visual sensitivity described by the CSF is highest at mid-frequencies, and the lowest visual sensitivity is observed at very high frequencies. This implies that visual signal components of high spatial frequencies cannot be easily identified by the human visual system as compared to those of the lower and mid-range frequencies.

Hence, noises at those very high frequencies range produced by quantisation during compression will contribute lesser amount of 'perceived' degradation in the visual quality of reconstructed images than those of lower and mid-range frequencies. The reason for this is due to the weaker ability of the human optics to detect visual signals at very high frequencies. Therefore, there is an obvious advantage for the visual signals to be moderated to reflect this limitation of the sensitivity of the human optics so as to improve the compression system. In the proposed model, the CSF is applied as uniform frequency-specific weights on the visual components in the spectral domain. The values of the weights are calibrated to coarsely address the effect of the band-pass profile of the human optics.

### 4.3.2 Cortical Stage

The cortical stage is represented by the masking or facilitation characteristics of the HVS whereby detection of visual stimulus can be impeded (i.e., masked) or enhanced (i.e., facilitated) in the presence of other visual patterns (i.e., a masker), respectively [22, 25]. Basically, the enhancement or impediment of the visual response is due to the responses of receptive fields in the visual cortex being triggered either positively (excitation) or negatively (inhibition), respectively [19-21]. In the proposed coder, for the purpose of image compression only masking is considered.

### 4.3.3 The Masking Model

The proposed masking model extends the grey scale model of Tan et al. [15] to the *YCbCr* color space within the contrast gain control structure (CGC) described in [27] by Watson and Solomon, and in [23] by Teo and Heeger. Unlike the proposed model that separates masking responses into intra-band and inter-orientation masking domains, Teo and Heeger only considered orientation masking, and Watson and Solomon unified all masking domains into a single response function.

Teo and Heeger used the shift invariant Steerable Pyramid transform [150] to decompose images into different frequencies and orientation bands, thereby avoiding aliasing. Watson and Solomon used either the cortex transform [141] or the Gabor array [27] for signal decomposition. All these transforms are overcomplete, and the basis of their use are due to overlapping nature of receptive fields of the HVS. The receptive fields are likely to be non-orthogonal as observed in [151]. The responses of the receptive fields in the cortex are band-selective. The visual perception is thought to be activated in multiple channels that are each selective in spatial frequency, orientation and temporal frequency. The bandwidths of spatial and orientation channels are found to be around one octave and 40 degrees, respectively. In addition, the data representation in the cortex appears to follow that of multiresolution representation, and it is thought to be covered by about 5 frequency

selective channels and 4 orientation channels. The steerable pyramid transform, cortex transform or Gabor array can provide the choice of tuning filters to specified frequencies and orientations while avoiding aliasing due to down sampling, making them excellent models for approximating the behaviour of the receptive fields in the cortex. They are excellent HVS models that can be used for perceptual quality assessment. However these filters are computationally complex. Also Gabor array has a much higher computational cost than the Cortex transform.

Although the cortex transform or the Gabor filters are better models for representing the receptive fields of the HVS, they are not used as transform kernels in the JPEG2000 framework. Instead, the Bi-orthogonal Daubechies 9/7 filter set as the wavelet transform kernel with dyadic decomposition is used in the proposed coder. The choice of Daubechies 9/7 filter set comes with some problems. Firstly, there are only 3 orientation bands at each frequency level instead of 4 orientation bands (i.e., the HVS needs at least 4 orientation bands). It has only one diagonal band at each frequency level that effectively combines responses from both 45 degrees and 135 degrees. Inaccuracy may arise with insufficient orientation bands. Secondly, the critically sampled wavelet transform can introduce aliasing errors. In spite of the drawbacks, for the purpose of exploiting the existing JPEG2000 framework, and at the same time with reasonable approximation to the modelling of the receptive fields, the coder described here uses the Bi-orthogonal Daubechies 9/7 filters as the wavelet transform kernel with Mallat decomposition [103].

All the above mentioned models and the proposed PCDM here have something in common with Foley's model as described in Chapter 2: the neural response $(R)$ of the cortical stage is modeled in terms of an excitation function ($E$) being 'masked' by a divisible inhibition Function ($I$) as in equation (2.9).

As the PCDM model discussed here is built into the coding structure of JPEG2000, an image in the discrete wavelet transform domain is divided into several codeblocks, each of which is hierarchically bitplane encoded with several coding passes per bitplane, beginning from the most significant bitplane and ending at the lowest bitplane.

We first define a linear transform, $T(.)$, of a natural digital colour image, x , as:

$$X = T(x) \tag{4.3}$$

where $X$ is the frequency and orientation sensitive spectral neural image. In the proposed coder, the image data is decomposed into a 5-level multiresolution spectral representation according to dyadic Mallat decomposition [103]. Transformed coefficients are denoted as either $X[c,l,\theta,m_l,n_l]$ or $X[c,1,LL,m_1,n_1]$ , where $X[c,l,\theta,m_l,n_l]$ is the coefficient at spatial frequency location $[m_l,n_l]$ in the orientation band, $\theta = \{LH, HL, HH\}$, at resolution level, $l = \{1,2,3,4,5\}$, belonging to colour component, $c \in \{Y, Cb, Cr\}$ , and $X[c,1,LL,m_1,n_1]$ refers to the transform coefficient for the lowest $LL$ isotropic (DC) band.

The transform coefficient is then modulated by the CSF weights according to the sensitivity of the human optics. The CSF weights used here are an attempt to roughly reflect the sensitivity of the human optics. The ability of the human optics to detect visual signals at very high frequencies is much weaker than at mid-range and lower frequencies. Note that the technique of CSF weighting for different subbands to account for their relative contributions for the purpose of rate allocation is commonly used. In the proposed coder, CSF weights are assigned according to frequency levels. A more accurate CSF curve is mentioned in Figure 2.12, which is adapted from Wandell [34]. The CSF weighted coefficients are expressed as,

$$X_w[c,l,\theta,m_l,n_l] = C_w[c,l]X[c,l,\theta,m_l,n_l] \tag{4.4}$$

$$X_w[c,1,LL,m_1,n_1] = C_w[c,LL]\cdot X[c,1,LL,m_1,n_1] \tag{4.5}$$

where $X_w[c,l,\theta,m_l,n_l]$ and $X_w[c,1,LL,m_1,n_1]$ are, respectively, the CSF-weighted coefficients of $X[c,l,\theta,m_l,n_l]$ and $X[c,1,LL,m_1,n_1]$. $C_w[c,l]$ is the CSF weights for color component, $c \in \{Y, Cb, Cr\}$, at resolution level, $l = \{1,2,3,4,5\}$. $C_w[c,LL]$ is the

CSF weight for color component, $c \in \{Y, Cb, Cr\}$ for the lowest $LL$ isotropic (DC) band.

The intra-band and inter-orientation maskings are expressed as follows,

$$R_z(c,l,\theta,m_l,n_l) = k_{c,z} \quad \frac{E_z(c,l,\theta,m_l,n_l)}{I_z(c,l,\theta,m_l,n_l) + \sigma_{c,z}^q} \tag{4.6}$$

where $E_z(c,l,\theta,m_l,n_l)$ and $I_z(c,l,\theta,m_l,n_l)$ are the excitation and inhibition functions, respectively, $k_{c,z}$ and $\sigma_{c,z}^q$ are the scaling and saturation coefficients, $z \in \{\Theta, \gamma\}$ with $\Theta$ and $\gamma$ represent the inter-orientation and intra-band masking domains, respectively. Note that the response, $R_z(c,l,\theta,m_l,n_l)$, increases with excitation but diminishes with inhibition. This models the phenomena that the visual pattern can be diminished by the presence of a masking pattern. The excitation and inhibition functions for the inter-orientation masking are defined as,

$$E_\Theta(c,l,\theta,m_l,n_l) = (X_w[c,l,\theta,m_l,n_l])^{p_{c,\Theta}} \tag{4.7}$$

$$I_\Theta(c,l,\theta,m_l,n_l) = \sum_{k=1}^{3} (X_w[c,l,\theta_k,m_l,n_l])^q \tag{4.8}$$

The excitation and inhibition functions of the intra-band domain are defined as,

$$E_\gamma(c,l,\theta,m_l,n_l) = (X_w[c,l,\theta,m_l,n_l])^{p_{c,\gamma}} \tag{4.9}$$

$$\begin{aligned} I_\gamma(c,l,\theta,m_l,n_l) = {} & X_w[c,l,\theta,m_l,n_l] + \\ & \frac{8}{N(l)} \sum_{u=m_l-l}^{m_l+l} \sum_{v=n_l-l}^{n_l+l} (X_w[c,l,\theta,u,v])^q \Big|_{[u,v] \neq [m_l,n_l]} + \sigma_{c,\mathrm{var}}^q(m_l,n_l) \end{aligned} \tag{4.10}$$

In the current model, $q$ is set to 2 with the condition $p_{c,z} > q$. Equation (4.8) represents the inhibition function as a sum of squares of the CSF-weighted transform

coefficients spanning all orientations (i.e., $k$=1,2,3 for LH, HL, HH orientation band, respectively) at spatial location $[m_l, n_l]$. Figure 4.4 depicts the inter-orientation and intra-band masking coefficients at work. The inhibition function in (4.10) comprises of three terms. The second term is the sum of squares of neighbouring CSF-weighted transform coefficients about the centroid, $X_w[c,l,\theta,m_l,n_l]$, and the neighbourhood is defined as a squared region with size of $N(l) = (2l+1)^2 - 1$, and $l = \{1,2,3,4,5\}$ from the lowest to the highest frequency level. The size of the neighbourhood is described pictorially in Figure 4.5. At this stage, little is known about what optimum neighbourhood sizes are for spatial masking. However, we can assume the neighbourhood size to be much smaller than the coverage of 2 degrees visual angle ($\theta$). Assume that an image of size 512x512 pixels (i.e., H=512, W=512) is to be displayed on a monitor with a viewing distance ($D$) at four times the image height ($H$), the vertical coverage that will reach the fovea is $2 \cdot D \cdot tan\left(\frac{\theta}{2}\right) = 2 \cdot (4x512) tan\left(\frac{2^\circ}{2}\right) = 4096 \cdot \tan(1^\circ) = 72$ pixels. For a 5-level Mallat decomposition with downsampling of 2 each at horizontal and vertical directions, the coverage corresponds to area sizes of 36x36, 18x18, 9x9, 5x5 and 3x3 pixels for frequency levels $l$ at 5, 4, 3, 2 and 1, respectively. The neighbourhood size for spatial masking can only be smaller. Based on subjective experiment, the coder is found to achieve excellent visual performance at neighbourhood regions of 11x11, 9x9, 7x7, 5x5 and 3x3 pixels at frequency level $l$ of 5, 4, 3, 2 and 1, respectively, for a 5-level Mallat decomposition. The third term in equation (4.10) is the local variance, $\sigma^2_{c,\mathrm{var}}(m_l, n_l)$, which accounts for the texture masking [123]. It is defined as

$$\sigma^2_{c,\mathrm{var}}(m_l,n_l) = \frac{1}{N(l)} \sum_{u=m_l-l}^{m_l+l} \sum_{v=n_l-l}^{n_l+l} \left(X_w[c,l,\theta,u,v] - \mu(m_l,n_l)\right)^2 \Big|_{[u,v] \neq [m_l,n_l]} \tag{4.11}$$

$$\mu(m_l,n_l) = \frac{1}{N(l)} \sum_{u=m_l-l}^{m_l+l} \sum_{v=n_l-l}^{n_l+1} X_w[c,l,u,v] \mathrm{I}_{[u,v] \neq [m_l,n_l]} \tag{4.12}$$

where $\mu(m_l, n_l)$ is the mean value of the set of neighboring coefficients about $X_w[c,l,\theta,m_l,n_l]$. At very high activity region of an image, the HVS is more tolerable to noise. The texture masking is included in addition to spatial masking to account for

the HVS's ability to tolerate higher distortion at very high activity region, where the tolerance to higher distortion could not be sufficiently accounted for by spatial masking alone.

The response function in (4.6) is applied to all subbands (LH, HL, HH) spanning from all resolution levels except the lowest LL isotropic (DC) band, whereby only intra-band masking is applied. The response function for the LL band is expressed as

$$R_{\gamma}(c,1,LL,m_1,n_1) = k_{c,\gamma} \frac{\left(\tilde{X}_w[c,1,LL,m_1,n_1]\right)^{p_{c,\gamma}}}{\left(X_w[c,1,LL,m_1,n_1]\right)^{q} + \sigma_{c,\gamma}^{q}(m_1,n_1)} \tag{4.13}$$

where $\tilde{X}_w[c,1,LL,m_1,n_1]$ and $X_w[c,1,LL,m_1,n_1]$ are the bitplane quantised and unquantised DC coefficients, respectively.

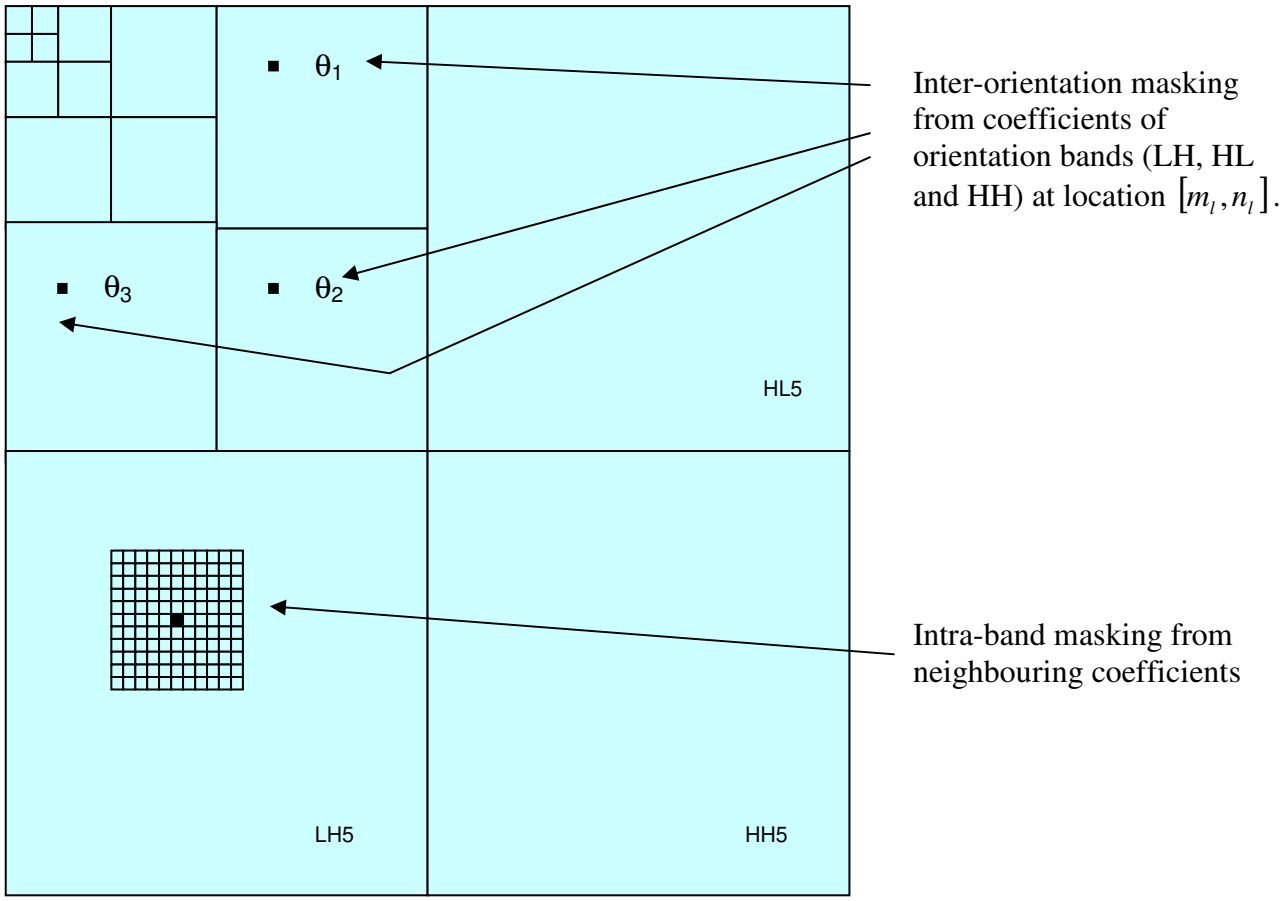



Figure 4.4 Example of 5-level dyadic wavelet decomposition structure. This diagram also gives a pictorial view of how coefficients are used for the inter-orientation masking and intra-band masking.

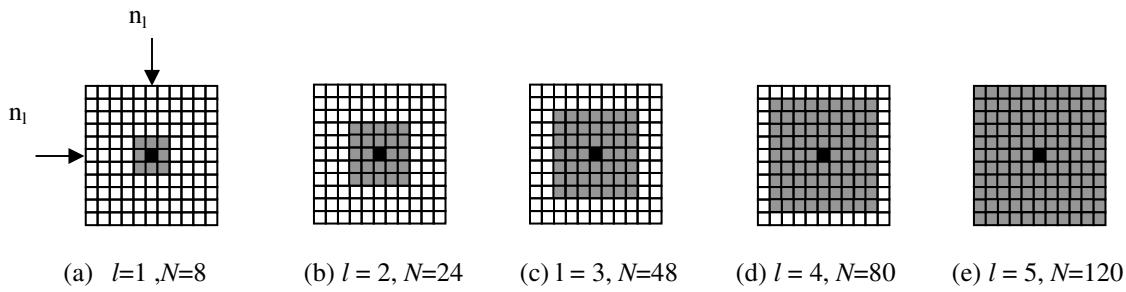



Figure 4.5 Neighbouring coefficients around centroid coefficient $X[c,l,\theta,m_l,n_l]$ for inclusion in computing intra-band masking. The neighbour coefficients are the shaded region excluding the coefficient $X[c,l,\theta,m_l,n_l]$. The size of the square is $N(l)=(2l+1)^2-1$, where $l$ is the resolution level from 1 to 5. Figures (a), (b), (c), (d) and (e) are the neighbouring coefficients for levels 1 to 5 respectively.

The difference of the masking response between the reference image ($\alpha$) and the processed image ($\beta$) (i.e., bitplane quantised image) for each colour component $c \in \{Y,Cb,Cr\}$ is determined by a simple squared-error ( $l_2$ norm) as expressed below,

$$D_z(c,l,\theta,m_l,n_l)=\left|R_{z,\alpha}(c,l,\theta,m_l,n_l)-R_{z,\beta}(c,l,\theta,m_l,n_l)\right|^2 \tag{4.14}$$

In equation (4.14), the $R_{z,\alpha}(c,l,\theta,m_l,n_l)$ is the response due to CSF weighted unquantised coefficient, $X_w[c,l,\theta,m_l,n_l]$ , and $z \in \{\Theta,\gamma\}$ represents the inter-orientation or intra-band masking domain, respectively. The $R_{z,\beta}(c,l,\theta,m_l,n_l)$ is the response due to CSF weighted biplane quantised coefficient, $\tilde{X}_w[c,l,\theta,m_l,n_l]$, at certain bit plane level, $b \in \{B,B-1,B-2,...,2,1\}$ and $B$ is the highest bit plane level. In JPEG2000, the bit plane encoding proceeds from the highest bit plane to the lowest bit plane, and multiple coding passes are involved in each bit plane level. When computing $R_{z,\beta}(c,l,\theta,m_l,n_l)$ , the CSF weighted quantized coefficient, $\tilde{X}_w[c,l,\theta,m_l,n_l]$, is used instead of the use of the CSF weighted unquantised coefficient, $X_w[c,l,\theta,m_l,n_l]$. For a bit plane level, $b \in \{B,B-1,B-2,...,2,1\}$, and $B$ is

the highest bit plane level, the square difference in the masking response of equation (4.14) essentially accounts for the distortion incurred by the bit plane quantisation at bit plane level, $\boldsymbol{b}$, for the coefficient, $X[c,l,\theta,m_l,n_l]$.

Encompassing both the intra-band and inter-orientation masking domains for all subbands, except the LL band which only considers intra-band masking, the final perceptual distortion measure, $D_c$, of each codeblock for each colour component $c \in \{Y,Cb,Cr\}$, is then computed as follows,

$$D_c = \sum_{i=1}^{M_l}\sum_{j=1}^{N_l}\left(g_{c,r}D_\gamma(c,l,\theta,i,j) + g_{c,\Theta}D_\Theta(c,l,\theta,i,j)\right) \tag{4.15}$$

where $g_{c,\gamma}$ and $g_{c,\Theta}$ are the proportional contributing gains for both intra-band and inter-orientation masking, respectively. $M_l$ and $N_l$ represent the actual size for the codeblock, at resolution level, $l$. At LL band, $g_{c,r}$ is set to 1 and the term, $g_{c,\Theta}D_\Theta(c,l,\theta,i,j)$ is omitted. Note that the perceptual distortion measure, $D_c$, is computed separately for each colour component $c \in \{Y,Cb,Cr\}$.

## 4.4 Model Adaptation

The PCDM is built into the coding structure of JPEG2000, where an image in the discrete wavelet transform domain is divided into several codeblocks, each of which is bitplane encoded [12]. In the proposed coder, both the unquantised and bitplane quantised coefficients are weighted according to their respective CSF weights. The masking function described in section 4.3 is applied to these weighted output and the distortion measure is then computed at the final detection stage. The distortion measure and the rate accumulated during bitplane encoding are used as inputs to the R-D optimisation function to generate the compressed bitstreams.

For a rate driven lossy coder, the purpose of the R-D function is to determine the minimum distortion possible for a given bitrate in such a way that any further reduction below the minimum distortion will not be possible without allowing an increase in the specified bitrate. In the JPEG2000 framework, the R-D optimisation uses the rate of reduction of distortion against the rate of increase in the bitrate to obtain the best possible distortion for the least number of bits. Let $R_{c,z}$ be the response of unquantised coefficient, and $R_{c,z,k,p}$ be the response of a coefficient quantised to the $k^{th}$ bitplane at $p^{th}$ coding pass, where $c \in \{Y, Cb, Cr\}$ denotes the colour component and $z \in \{\Theta, \gamma\}$ for inter-orientation and intra-band maskings domains. The perceptual distortion that corresponds to bitplane quantisation for the $k^{th}$ bitplane at $p^{th}$ coding pass of colour component $c \in \{Y, Cb, Cr\}$ is

$$D_{c,k,p} = \sum_{j \in N_{c,k,p}} \sum_{z} g_{c,z} \left| R_{c,z,k,p}(j) - R_{c,z}(j) \right|^2 \tag{4.16}$$

where $g_{c,z} = \{g_{c,\gamma}, g_{c,\Theta}\}$ refers to the proportional contributing gains for both intra-band and inter-orientation masking. $N_{c,k,p}$ denotes the set of coefficients that belong to the coding pass $p^{th}$ at $k^{th}$ bitplane of colour component $c \in \{Y, Cb, Cr\}$. For the JPEG2000-PCDM coder, the perceptual distortion in equation (4.16) is used to replace the MSE distortion described in equation (4.1). The reduction in perceptual distortion between successive bitplanes $k^{th}$ and $(k+1)^{th}$ for colour component $c \in \{Y, Cb, Cr\}$ is

$$\Delta D_{c,k,p} = D_{c,k,p} - D_{c,k+1,p} \tag{4.17}$$

## 4.5 Model Calibration

The CSF weights and model parameters (see Tables 4.2 and 4.3) are calibrated to the perceptual response of the HVS. For each model parameter value estimation, nine natural images are derived as test images from three sets of images (i.e., *barbara*,

*barbara2*, *boats*, see appendix B), each of which is coded at bitrates of 0.5, 0.3 and 0.25 bpp.

### 4.5.1 Test Condition

The calibration of CSF weights and model parameters was conducted on a Sun Ultra 60 Workstation in a dark room with minimum illumination. The test images were viewed by one expert viewer on a 21-inch, 0.24mm dot pitch Sun Colour Monitor with its display set at 1280x1024 pixels resolution. This display setting allows the paired images (512x512 pixels each) to span the entire display horizontally. However, the tradeoff of not having display set at its native resolution is that some internal interpolation does occur. The viewing distance was three times the image height [152]. Between quality assessments of the images of the current estimated parameter set and the next one, a break of at least 10 minutes was observed to avoid the effect of fatigue during the subjective test. The presentation of the test images is depicted in Figure 4.6. Force-choice comparative subject assessment was used to evaluate the quality of the images.

### 4.5.2 Calibration Process

The set of model parameter values are taken from Tan et al. [15] as the set of initial parameter values for the *YCbCr* color space. While no best way has yet been devised for parameterising the 42 parameters ($M_p$), the current approach to optimising the parameters is sequential tuning iteratively. The sequential tuning of parameters may proceed for multiple passes (i.e., an approximation pass and multiple refinement passes) with different step sizes ($\delta_R$). While the approximation pass uses larger step size, the refinement passes use smaller step sizes. The approximation pass aims at achieving the parameter set close to the sub-optimal values with fast convergence, while the refinement passes attempt to calibrate the parameters to the sub-optimal set at a finer resolution.

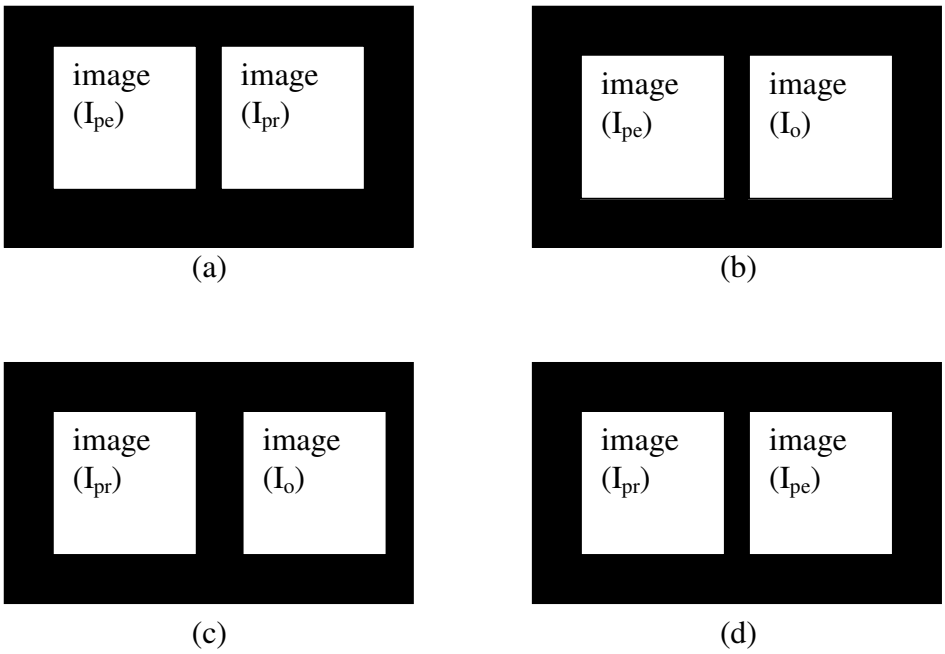



Figure 4.6 Presentation of subjective test images for parameter calibration. $I_{pe}$, $I_{pr}$ and $I_o$ represent the image with estimated parameter set, the image with reference parameter set, and the original uncompressed image respectively. The images are sequentially presented in the order of (a), (b), (c) and (d). For each paired images, the position of an image on either left or right is pseudo-randomised. Each iteration uses four combination assessments, a, b, c, and d. A decision is made after viewing all the images.

The model parameters are calibrated within the context of the coder as shown in Figure 4.7.

Let $I_d(\Gamma)$ represents the complete set of distorted images (9 images) produced by the PCDM with parameter set, $\Gamma = \{P_e(i), P_r(i)\}$. $P_e(i)$ and $P_r(i)$ represent, respectively, the estimate and reference parameter sets of the current iteration, $i$. Consequently, the selection of the reference parameter set is expressed as,

$$P_r(i+1) = f_s(I_d(P_e(i)), I_d(P_r(i))) \tag{4.18}$$

where $f_s(.)$ is the force choice subjective assessment operation. The selection of $P_r(i+1)$ is subjected to the assessment setup as depicted in Figure 4.6. Note that all the 9 distorted images at bitrates of 0.5, 0.3 and 0.25 bpp of $P_e(i)$ are evaluated against those of $P_r(i)$ with their original uncompressed images taken as additional reference set for force choice consideration. The parameter set (i.e., either $P_e(i)$ or $P_r(i)$) is selected as the better parameter set if it scores the higher number of subjective preferences (a value between 0 and 9). The better parameter set is then used in the next iteration ($i+1$) as the reference parameter set, $P_r(i+1)$. The next estimated parameter set $P_e(i+1)$ is determined by the step size, $\delta_R$, which varies from 0.02 to 0.0001 depending on whether it is in approximation pass or refinement pass. The force choice procedure applies to all the model parameters, $M_p$, where,

$$M_p = \{C_w[Y,LL], C_w[Y,l], C_w[Cr,l], C_w[Cb,l], k_{c,z}, \sigma_{c,z}, p_{c,z}, g_{c,z}\} \tag{4.19}$$

where $z \in \{\Theta, \gamma\}$ and $l = \{1,2,...,5\}$.

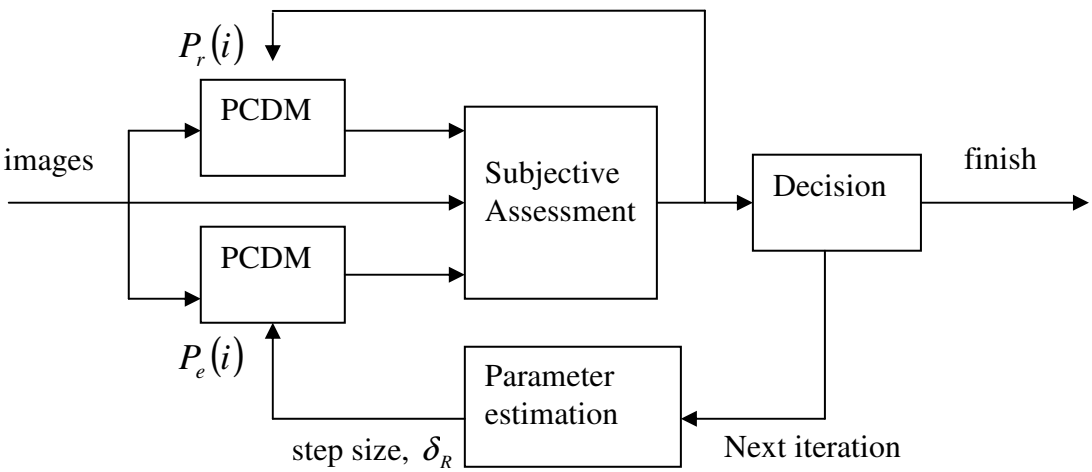



Figure 4.7 Calibration of parameters in the context of coder. (The step size $\delta_R$ for each parameter varies according to the approximation and refinement passes.)

The calibration process is described in details as follows,

a.1. The values of the parameter set $M_p$ for all the *Y*, *Cb* and *Cr* are initialised to the same values of the parameter set from Tan et al. [15]. The step size, $\delta_R$, varies from 0.02 to 0.0001, is expressed as,

$$\delta_R(i) = \frac{1}{i} \tag{4.20}$$

a.2. Calibration begins with *Y* component with step size increment of $\delta_R(50)$ and initial $C_w[Y, LL] = 0.6$.

a.3. Equation (4.18) is used to determine the parameter set, either $P_r(i)$ or $P_e(i)$, that scores the higher subjective preferences. The parameter value is increased by the same step size increment until the visual quality of the images degrades in three consecutive step size increments. The parameter set that gives the best visual quality is chosen as the new parameter set so that it will be used for calibration for the other model parameter as well as in the next iteration *i+1*. When calibrating a model parameter, the calibration always begins by setting that parameter to its initial value while the other parameters use their 'best values' obtained from the previous calibration.

a.4. Similarly, calibrate all $C_w[Y, l]$ with initial value of 0.6 and step size increment of $\delta_R(50)$ with the same procedure as in step a.3.

a.5. Follow the same procedures in a.3 and a.4, calibrate the $C_w[Cb, LL]$ and $C_w[Cb, l]$ with initial value of 0.6 and step size increment of $\delta_R(50)$ for colour component *Cb*.

a.6. Follow the same procedures in a.3 and a.4, calibrate the $C_w[Cr, LL]$ and $C_w[Cr, l]$ with initial value of 0.6 and step size increment of $\delta_R(50)$ for colour component *Cr*. $k_{c,\gamma}, g_{c,\gamma}, k_{c,\Theta}, g_{c,\Theta}$

a.7. Calibrate the $\sigma_{c,\Theta}, p_{c,\Theta}, \sigma_{c,\gamma}, p_{c,\gamma}$ with step size $\delta_R(50)$ by following the step in a.3 in the order of $\sigma_{c,\Theta}$, $p_{c,\Theta}$, $\sigma_{c,\gamma}$, $p_{c,\gamma}$, and colour component, $c \in \{Y, Cb, Cr\}$, in the order of *Y*, *Cb*, and *Cr*. When calibrating a parameter, it is set to its initial value while the other parameters use the new set of values.

The initial values for $\sigma_{c,\Theta}$, $p_{c,\Theta}$, $\sigma_{c,\gamma}$, and $p_{c,\gamma}$ are set at 4.0, 2.0, 1.0, and 2.0, respectively.

a.8. Calibrate the $k_{c,\gamma}$, $g_{c,\gamma}$, $k_{c,\Theta}$, $g_{c,\Theta}$ with step size $\delta_R(50)$ by following the step in a.3 in the order of $k_{c,\gamma}$, $g_{c,\gamma}$, $k_{c,\Theta}$, $g_{c,\Theta}$, and colour component, $c \in \{Y, Cb, Cr\}$, in the order of *Y*, *Cb*, and *Cr*. When calibrating a parameter, it is set to its initial value while the other parameters use the new set of values. The initial values for $k_{c,\gamma}$, $g_{c,\gamma}$, $k_{c,\Theta}$, and $g_{c,\Theta}$ are set to 0.8, 0.3, 0.8, and 0.3, respectively.

a.9. With the new set of parameters, calibrate the all parameters by following steps from a.3 to a.8 with the step size being refined to increment of $\delta_R(1000)$. When calibrating a parameter, it is set to its initial value while the other parameters use the new set of values. The initial values for the $C_w[c, LL]$ and $C_w[c, l]$ are set to 0.6, $k_{c,z}$ to 0.8, $p_{c,z}$ at 2.0, $\sigma_{c,\Theta}$ at the maximum value of new set value minus 4.0 and 0.4, $\sigma_{c,\gamma}$ at the maximum value of new set value minus 4.0 and 1.0, $g_{c,z}$ at the maximum value of new set value minus 3.0 and 0.3.

a.10. With the new set of parameters, the calibration repeats from a.3 to a.9 with final step size of $\delta_R(10000)$ and initial values of those used while calibrating with step size of $\delta_R(1000)$.

It is noted that the calibration of each model parameter ends when the next three successive step size increments do not yield a visual improvement in image quality of any of the test images for each step size setting of $\delta_R(50)$, $\delta_R(1000)$ and $\delta_R(10000)$.

Tables 4.2 and 4.3 are the final output of calibration. The SET-A parameters were calibrated with initial values taken from Tan et al. [15]. The SET-A parameters were used in the subjective assessment I, the result of which is reported in section 4.6.1. In the hope of improving the visual performance of the coder, the parameters were re-calibrated by following the steps from a.3 to a.10 but with SET-A parameters as the initial values. The result is the set of parameters listed in Table 4.3 as SET-B parameters. The SET-B parameters were used in subjective assessment II as

described in section 4.6.2. It must be mentioned that these two sets of parameters are just sub-optimals due to the sequential nature of the calibration process that is used to search through a rather large 42-parameter space. It has not been found that either parameter set yields better visual performance than the other. It is believed that many sub-optimal parameter sets could give rise to comparable visual performance for the coder. The calibration process could produce multiple sets of sub-optimal parameters that could give comparable visual performance.

| CSF weights and Model Parameters | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Y | Cb | Cr | | Y | Cb | Cr |
| $C_w[c, LL]$ | 0.95 | 1.03 | 1.28 | $\sigma_{c,\Theta}$ | 6.925 | 15.02 | 10.11 |
| $C_w[c,1]$ | 1.15 | 1.23 | 1.35 | $p_{c,\Theta}$ | 2.145 | 2.040 | 2.215 |
| $C_w[c,2]$ | 1.33 | 1.39 | 1.40 | $g_{c,\Theta}$ | 0.35 | 0.501 | 0.35 |
| $C_w[c,3]$ | 1.41 | 1.34 | 1.35 | $k_{c,\gamma}$ | 1.09 | 1.11 | 0.98 |
| $C_w[c,4]$ | 1.30 | 1.10 | 1.13 | $\sigma_{c,\gamma}$ | 2.505 | 11.00 | 1.505 |
| $C_w[c,5]$ | 1.02 | 0.65 | 0.85 | $p_{c,\gamma}$ | 2.153 | 2.170 | 2.300 |
| $k_{c,\Theta}$ | 0.9876 | 0.9800 | 0.9300 | $g_{c,\gamma}$ | 0.37 | 0.85 | 0.402 |

Table 4.2 SET-A Sub-optimal CSF weights and model parameters.

| CSF weights and Model Parameters | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Y | Cb | Cr | | Y | Cb | Cr |
| $C_w[c, LL]$ | 0.95 | 1.03 | 1.28 | $\sigma_{c,\Theta}$ | 6.925 | 15.02 | 10.11 |
| $C_w[c,1]$ | 1.15 | 1.23 | 1.35 | $p_{c,\Theta}$ | 2.145 | 2.040 | 2.215 |
| $C_w[c,2]$ | 1.33 | 1.39 | 1.40 | $g_{c,\Theta}$ | 0.346 | 0.490 | 0.338 |
| $C_w[c,3]$ | 1.41 | 1.34 | 1.35 | $k_{c,\gamma}$ | 1.053 | 1.092 | 1.005 |
| $C_w[c,4]$ | 1.30 | 1.10 | 1.13 | $\sigma_{c,\gamma}$ | 2.505 | 11.00 | 1.505 |
| $C_w[c,5]$ | 1.02 | 0.65 | 0.85 | $p_{c,\gamma}$ | 2.153 | 2.170 | 2.300 |
| $k_{c,\Theta}$ | 0.999 | 1.002 | 0.963 | $g_{c,\gamma}$ | 0.383 | 0.864 | 0.392 |

Table 4.3 SET-B Sub-optimal CSF weights and model parameters.

## 4.6 Experimental Results and Analysis

The performance evaluation of PCDM has been conducted against the two benchmarks metric, the MSE and the CVIS [12] within the JPEG2000 software verification model version 8 (VM8) coder through force-choice comparative subjective tests [153, 154]. The evaluation was carried out in two parts: assessments I and II. For each assessment part, source images were each coded at four different bitrates of 1.0, 0.5, 0.25 and 0.125 bpp by three different coders: JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVIS. Note that the masking gain, $g$, is set at 0.5 for the CVIS criterion (see equation (4.2)). Paired images generated by the JPEG2000-PCDM and benchmarks are arranged side by side for assessment on a monitor as depicted in Figure 4.8. The viewing distance is two and a half times the image height [152]. The position of images displayed either on the left or the right, is pseudo-randomised. Figure 4.9 illustrates the force-choice assessment process.

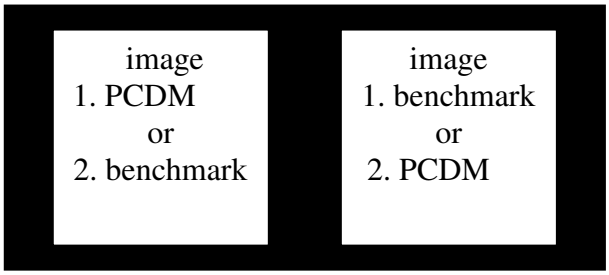



Figure 4.8 Arrangement of paired images on a Monitor.
Left/Right position of images are pseudo-randomised.

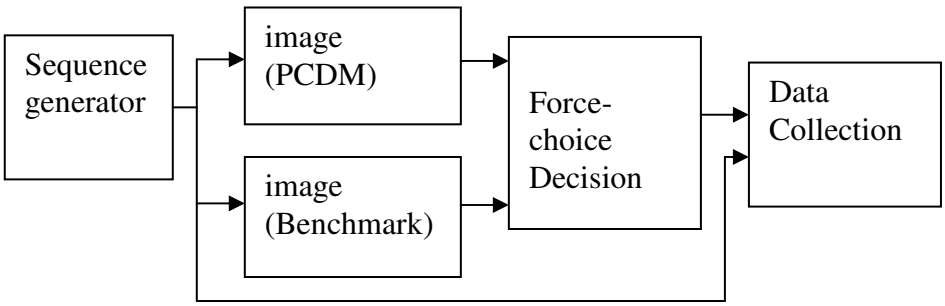



Figure 4.9 Pictorial view of force-choice comparative subjective test. The sequence generator is pseudo-randomised based on both image and bitrate. For each subject, both sequence number will not be re-used after it

The subjective tests were conducted in a dark room with minimum illumination. The sequences of paired images were randomised from 1 to *N*, where *N* was either 20 or 24 for assessment I and assessment II, respectively.

### 4.6.1 Subjective Assessment I

Assessment I involved 6 participants viewing 20 paired images generated from 5 different source images (*goldhill*, *sail*, *pepper*, *lena*, *tulip*). The PCDM in this instance uses SET-A model parameters from Table 4.2. Images (cropped at 512x512 pixels) were viewed on a 21 inch, 0.24mm dot pitch Sun Monitor with display resolution set to 1280×1024 pixels. The images were cropped after compression in such a way that the important image features were included in the cropped images. For example, regions such as the face, the hat, hairs and their immediate neighbourhoods are important features for "*lena*", so they were included in the cropped image of "*lena*". For "*tulip*", several tulip flowers were included. For "*goldhill*", the cropped image contained several adjacent buildings and the backdrop. These are important image contents which were included in the cropped images. This is the policy used for cropped images in all subjective assessments mentioned in chapters 4 and 5 of this thesis. The raw scores of the test results are presented in Table 4.4.

| Image | Bitrate (bpp) | Raw Scores | | | |
|---|---|---|---|---|---|
| | | Test 1 | | Test 2 | |
| | | A | B | A | C |
| *goldhill* | 1.0 | 0 | 6 | 4 | 2 |
| | 0.5 | 5 | 1 | 5 | 1 |
| | 0.25 | 4 | 2 | 6 | 0 |
| | 0.125 | 5 | 1 | 5 | 1 |
| *Sail* | 1.0 | 4 | 2 | 5 | 1 |
| | 0.5 | 5 | 1 | 6 | 0 |
| | 0.25 | 5 | 1 | 6 | 0 |
| | 0.125 | 6 | 0 | 6 | 0 |
| *pepper* | 1.0 | 1 | 5 | 2 | 4 |
| | 0.5 | 5 | 1 | 5 | 1 |
| | 0.25 | 3 | 3 | 5 | 1 |
| | 0.125 | 4 | 2 | 4 | 2 |
| *Lena* | 1.0 | 4 | 2 | 3 | 3 |
| | 0.5 | 4 | 2 | 3 | 3 |
| | 0.25 | 5 | 1 | 6 | 0 |
| | 0.125 | 5 | 1 | 6 | 0 |
| *Tulip* | 1.0 | 2 | 4 | 4 | 2 |
| | 0.5 | 4 | 2 | 5 | 1 |
| | 0.25 | 2 | 4 | 6 | 0 |
| | 0.125 | 5 | 1 | 5 | 1 |

Table 4.4 Comparative Forced-Choice Subjective Test Results. A – JPEG2000-PCDM coder, B – JPEG2000-MSE, C – JPEG2000-CVIS. Test 1 for JPEG2000-PCDM against JPEG2000-MSE. Test 2 for JPEG2000-PCDM against JPEG2000-CVIS.

Evaluation of the test results can be achieved by paired *t*-test [155, 156], and the *t*-value can be computed by,

$$t = \frac{\sum_{i=1}^{N} d_i}{\sqrt{\frac{N\sum_{i=1}^{N} d_i^{\,2} - \left(\sum_{i=1}^{N} d_i\right)^2}{N-1}}} \tag{4.21}$$

where $d_i$ is the difference between raw scores of JPEG2000-PCDM and the benchmark coders, and $i=\{1,2,\ldots, N\}$ is the test sequence number. The critical $t$ for 3 and 4 degrees of freedom (d.f.) at 95%, 99% and 99.5% Confidence Intervals (CI) is

tabulated in Table 4.5. The evaluation is based on comparing the *t*-value and the critical *t* at certain degree of freedom (d.f.) with certain Confidence Interval (CI). If the difference in preference for two coders under measurement has *t*-value higher than the critical *t*, the Null hypothesis is rejected and the Alternate hypothesis is accepted, and vice versa.

| d.f. | $t_{0.05}$ | $t_{0.01}$ | $t_{0.005}$ |
|---|---|---|---|
| 4 | 2.1318 | 3.7469 | 4.6041 |
| 3 | 2.3534 | 4.5407 | 5.8409 |

Table 4.5 Critical *t* [157] at 95% ($t_{0.05}$), 99% ($t_{0.01}$) and 99.5% ($t_{0.005}$) confidence interval

As there were only six participants for assessment I, it will be necessary to combine the data sets before paired *t*-test analysis can be performed. This is to ensure that the data set has reasonable number of sample points for meaningful statistical analysis. This compaction of data also leads to diminished dimensionality, i.e., it cannot measure performance for each image at each bitrate. The data sets from raw scores of Table 4.4 are grouped as follows,

- The scores of bitrate 1.0, 0.5, 0.25 and 0.125 are combined up for each of the five source images. This is tabulated in Table 4.6. The 5 paired sets correspond to 4 degree of freedom (d.f.). This analysis only provides the overall performance according to different source image.
- The scores of the 10 images are summed up for each bitrate (1.0, 0.5, 0.25 and 0.125 bpp), and the data set is tabulated in Table 4.7. The 4 paired sets correspond to 3 degree of freedom (d.f.). This provides overall performance analysis of PCDM for different bitrates only.

| | P | | | |
|---|---|---|---|---|
| Image | Scores | | | |
| | Test 1 | | Test 2 | |
| | A | B | A | C |
| *goldhill* | 14 | 10 | 20 | 4 |
| *sail* | 20 | 4 | 23 | 1 |
| *pepper* | 13 | 11 | 16 | 8 |
| *lena* | 18 | 6 | 18 | 6 |
| *tulip* | 13 | 11 | 20 | 4 |

Table 4.6 Comparative Forced-Choice Subjective Results, categorising according to images. (By summing up the preferences of bitrate 1.0, 0.5, 0.25 and 0.125 for each type of images. Note: A – JPEG2000-PCDM coder, B – JPEG2000-MSE, C – JPEG2000-CVIS. Test 1 for JPEG2000-PCDM against JPEG2000-MSE. Test 2 for JPEG2000-PCDM against JPEG2000-CVIS.)

| | Q | | | |
|---|---|---|---|---|
| Bitrate (bpp) | Scores | | | |
| | Test 1 | | Test 2 | |
| | A | B | A | C |
| 1.0 | 11 | 19 | 18 | 12 |
| 0.5 | 23 | 7 | 24 | 6 |
| 0.25 | 19 | 11 | 29 | 1 |
| 0.125 | 25 | 5 | 26 | 4 |

Table 4.7 Comparative Force-Choice Subjective Test Results, categorising according to bitrates. (By summing up the preferences of 5 images for each of the bitrates. Note: A – JPEG2000-PCDM coder, B – JPEG2000-MSE, C – JPEG2000-CVIS. Test 1 for JPEG2000-PCDM against JPEG2000-MSE. Test 2 for JPEG2000-PCDM against JPEG2000-CVIS.)

The *t*-values are computed based on the group data sets of Tables 4.6 and 4.7. For the paired *t*-test, 5 and 4 paired sets correspond to 4 and 3 degrees of freedom (d.f.), respectively. The *t*-values are tabulated in Table 4.8 for Tests 1 and 2.

The Null hypothesis, $H_0$, of the paired *t*-test here assumes that "The perceived image quality of JPEG2000-PCDM is equivalent to or worse than the benchmarks", while the alternate hypothesis, $H_1$, is "the perceived image quality of the JPEG2000-PCDM is better than the benchmarks."

| | Types of Category | P | Q |
|---|---|---|---|
| | d.f. | 4 | 3 |
| *t*-value | Test 1 | 2.5082 | 1.4536 |
| | Test 2 | 6.3454 | 3.9821 |

Table 4.8 The *t*-values. (P - categorising according to image from Table 4.6. Q - categorising according to bitrates from Table 4.7)

**a. Test 1**

From Table 4.8, in (P), the *t*-value (2.5082) is higher than the critical *t* for 4 d.f. at 95% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, the JPEG2000-PCDM is perceived to be superior to the JPEG2000-MSE for all source images. Based on evaluation of (Q), the quality performance of JPEG2000-PCDM is perceived to be statistically equivalent to or worse than the JPEG2000-MSE according to bitrates category, as the *t*-value (1.4536) is lower than the critical *t*.

**b. Test 2**

In (P), the *t*-value (6.3454) is higher than the critical *t* for 4 d.f. at 95% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, the JPEG2000-PCDM is perceived to be superior to the JPEG2000-CVIS for all source images. Based on evaluation of (Q), the quality performance of JPEG2000-PCDM is perceived to be statistically better than the JPEG2000-CVIS according to bitrates category, as the *t*-value (3.9821) is higher than the critical *t* for 3 d.f. at 95% CI.

From the *t*-test analysis for Tests 1 and 2, overall, JPEG2000-PCDM produces images with better perceived quality improvement than the JPEG2000 benchmarks for all source images. However, it cannot be established that JPEG2000-PCDM produces images better than those of JPEG2000-MSE for all bitrate categories from1.0 to 0.125 bpp. Further subjective assessment with more participants is needed to investigate visual performance of the proposed coder for bitrate categories as in subjective assessment II.

### 4.6.2 Subjective Assessment II

Subjective experiment II involves 30 participants viewing a total of 24 images produced from 6 different images (*goldhill*, *sail*, *pepper*, *lena*, *tulip*, *paintedhouse*), each coded at bitrate of 1.0, 0.5, 0.25 and 0.125 bpp. The PCDM based coder uses SET-B sub-optimal CSF weights and model parameters from Table 4.3. The images (cropped at 500x500 pixels) are assessed on a 19 inch Colour Monitor (Model: Diamond Digital DV997FD) with resolution adjusted at 1280×1024 pixels. Due to unavailability of the 21 inch Sun Monitor at this stage, the 19 inch Monitor is used instead. To avoid displaying the outer region of the images on the slightly curving region along the boundaries of the Monitor, the images are cropped at 500x500 pixels instead of 512x 512 pixels as reported earlier. To ensure the quality of the subjective assessment, the participants were fully voluntary and had to be 18 years and above. They came from a varied range of profession, so that they are not all expert viewers in the field of image processing. It is known that colour perception differs between male and female. Hence a good mix of male and female participants were involved in the subjective assessment. More importantly, all participants are not known to have colour deficiency. For those who did wear glasses, they were asked to view the images with their glasses on. Each participant was presented with the questionnaire set out in Appendix C. Basically, the participants had to choose one of the randomized images according to their preferences. To eliminate the fatigue factor, they were given a break before they were presented with the next sequence of randomized images. The complete set of test images is contained in the CD in Appendix H. The raw scores of the test results are presented in Table 4.9.

Again, the same Null hypothesis, $H_0$, and Alternate hypothesis, $H_1$, from Assessment I were assumed. Evaluation of the test results is based on (a) all the twenty images covering all the four bitrates, and (b) per bitrate category (involving six images per bitrate). For the paired *t*-test, 24 and 6 paired sets correspond to 23 and 5 degrees of freedom (d.f.), respectively. Table 4.10 shows the critical *t* for 23 and 5 d.f. at 95%, 99% and 99.5% confidence intervals (CI), respectively.

| Bitrate (bpp) | Images | Raw scores | | | |
|---|---|---|---|---|---|
| | | Test 1 | | Test 2 | |
| | | A | B | A | C |
| 1.0 | Goldhill | 18 | 12 | 20 | 10 |
| | Sail | 12 | 18 | 17 | 13 |
| | Pepper | 18 | 12 | 17 | 13 |
| | Lena | 18 | 12 | 12 | 18 |
| | Tulip | 11 | 19 | 16 | 14 |
| | paintedhouse | 15 | 15 | 25 | 5 |
| 0.5 | Goldhill | 20 | 10 | 26 | 4 |
| | Sail | 17 | 13 | 26 | 4 |
| | Pepper | 15 | 15 | 21 | 9 |
| | Lena | 19 | 11 | 22 | 8 |
| | Tulip | 17 | 13 | 24 | 6 |
| | paintedhouse | 23 | 7 | 25 | 5 |
| 0.25 | Goldhill | 19 | 11 | 27 | 3 |
| | Sail | 19 | 11 | 27 | 3 |
| | Pepper | 20 | 10 | 25 | 5 |
| | Lena | 21 | 9 | 28 | 2 |
| | Tulip | 23 | 7 | 27 | 3 |
| | paintedhouse | 27 | 3 | 26 | 4 |
| 0.125 | Goldhill | 25 | 5 | 27 | 3 |
| | Sail | 26 | 4 | 29 | 1 |
| | Pepper | 16 | 14 | 29 | 1 |
| | Lena | 28 | 2 | 29 | 1 |
| | Tulip | 23 | 7 | 28 | 2 |
| | paintedhouse | 27 | 3 | 25 | 5 |

Table 4.9 Comparative Forced-Choice Subjective Results.
(A – JPEG2000-PCDM coder, B JPEG2000-MSE, C – JPEG2000-CVIS. Test 1 for JPEG2000-PCDM against JPEG2000-MSE, Test 2 for JPEG2000-PCDM against JPEG2000-CVIS)

| d.f. | $t_{0.05}$ | $t_{0.01}$ | $t_{0.005}$ |
|---|---|---|---|
| 23 | 1.7139 | 2.4999 | 2.8073 |
| 5 | 2.0150 | 3.3649 | 4.0322 |

Table 4.10 Critical $t$ at 95% ($t_{0.05}$), 99% ($t_{0.01}$) and 99.5% ($t_{0.005}$) confidence interval.

The $t$-values are presented in Table 4.11. In the ALL bitrate category, the $t$-values for Tests 1 (5.1500) and 2 (9.6033) are higher than the critical $t$ (2.8073) at 23 d.f. with 99.5% CI. Hence, the Null Hypothesis ($H_0$) is rejected, and the JPEG2000-PCDM is

overall statistically superior to both the JPEG2000-MSE and JPEG2000-CVIS coder with 99.5% CI. At high bitrate (1.0 bpp) category, the JPEG2000-PCDM is equivalent to or worse than the JPEG2000-MSE and JPEG2000-CVIS since the *t*-values (0.2548 for Test 1, and 1.5936 for Test 2) are lower than the critical *t* (2.0150) at 95% CI. At 99.5% CI, from low (0.125 bpp) to intermediate bitrates (0.5 bpp), their *t*-values are higher than the critical *t* except in the case against JPEG2000-MSE. Therefore the perceived quality of the images generated by JPEG2000-PCDM from low to intermediate bitrates are better than the two benchmarks with 99.5% confidence interval in all cases except against JPEG2000-MSE at 0.5 bpp. At 0.5 bpp, the JPEG2000-PCDM is perceived to have better perceived quality improvement than the JPEG2000-MSE with 95% CI.

| | Bitrate (bpp) | 0.125 | 0.25 | 0.5 | 1.0 | ALL |
|---|---|---|---|---|---|---|
| | d.f. | 5 | 5 | 5 | 5 | 23 |
| Computed t-value | Test 1 | 5.1557 | 5.1657 | 3.0502 | 0.2548 | 5.1500 |
| | Test 2 | 19.6214 | 27.6699 | 10.5097 | 1.5936 | 9.6033 |

Table 4.11 Computed *t*-values based on different bitrate categories for subjective assessment II.

In short, the perceived quality improvements are as follows,

- Overall, JPEG2000-PCDM produces images with better perceived image quality than that of JPEG2000-MSE and JPEG2000-CVIS.
- When breaking down into individual bitrate category, JPEG2000-PCDM produces images with better perceived image quality than JPEG2000-MSE and JPEG2000-CVIS from low (0.125 bpp) to intermediate (0.5 bpp) bitrate with 99.5% CI except against JPEG2000-MSE at 0.5 bpp. At 0.5 bpp, JPEG2000-PCDM is better than JPEG2000-MSE with 95% CI.
- At high bitrate of 1.0 bpp, the force-choice subjective assessment does not establish that JPEG2000-PCDM coder produces images with better perceived image quality than both the JPEG2000-MSE and JPEG2000-CVIS.

At high bit rate of 1.0 bpp and above, it is difficult for the human viewers to identify the quality differences of images produced by the various coders: JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVIS.

The objective measure, peak-signal-to-noise-ratio (PSNR), for the JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVIS for the test images is attached in Appendix F. It must be emphasised that images with higher PSNR as in Appendix F do not necessarily possess better perceived visual quality. On the contrary, some images produced by JPEG2000-MSE and JPEG2000-CVIS with higher PSNR than those of JPEG2000-PCDM were rated poorly than the JPEG2000-PCDM during force-choice subjective assessments. This re-affirms that the MSE or PSNR as an objective quality metric does not correlate well as far as perceived quality by HVS is concerned, which is as reported in Girod [143] and Wang et al. [144].

In Figure 4.10, better visual quality can be observed around the eyes of *lena* at 0.125 bpp for JPEG2000-PCDM coder. For *lena*, 'clipped' eye is observed for both JPEG2000-MSE and JPEG2000-CVIS coders while JPEG2000-PCDM coder retains most of the details of *lena*'s eye. Shaper nose area of *lena* is observed for the JPEG2000-PCDM coder than the two JPEG2000 benchmarks. Pattern aliasing is less obvious around the edges of lena's hat for JPEG2000-PCDM coder. In the case of *tulip* in Figure 4.11, the image coded at 0.125 bpp by the JPEG2000-PCDM coder is less blur with shaper details in the centre of *tulip*. Similarly, *sail* coded at 0.25 bpp by the JPEG2000-PCDM coder is able to preserve number details better than the other coders as indicated in Figure 4.12.

Overall, the JPEG2000-MSE criterion somehow achieves better visual performance than the CVIS criterion. This is likely due to visual weighting being used with the MSE in the VM8 version of the JPEG2000.

A complete set of test images with various bit rates is provided in the CD in Appendix J.

JPEG2000-PCDM coder (0.125 bpp)

JPEG2000 with MSE (0.125 bpp)

PCDM coder produces better perceived visual details around the eyes of *lena*.

JPEG2000 with CVIS (0.125 bpp)

Original Uncompressed Image

Figure 4.10 Cropped images of *lena*.

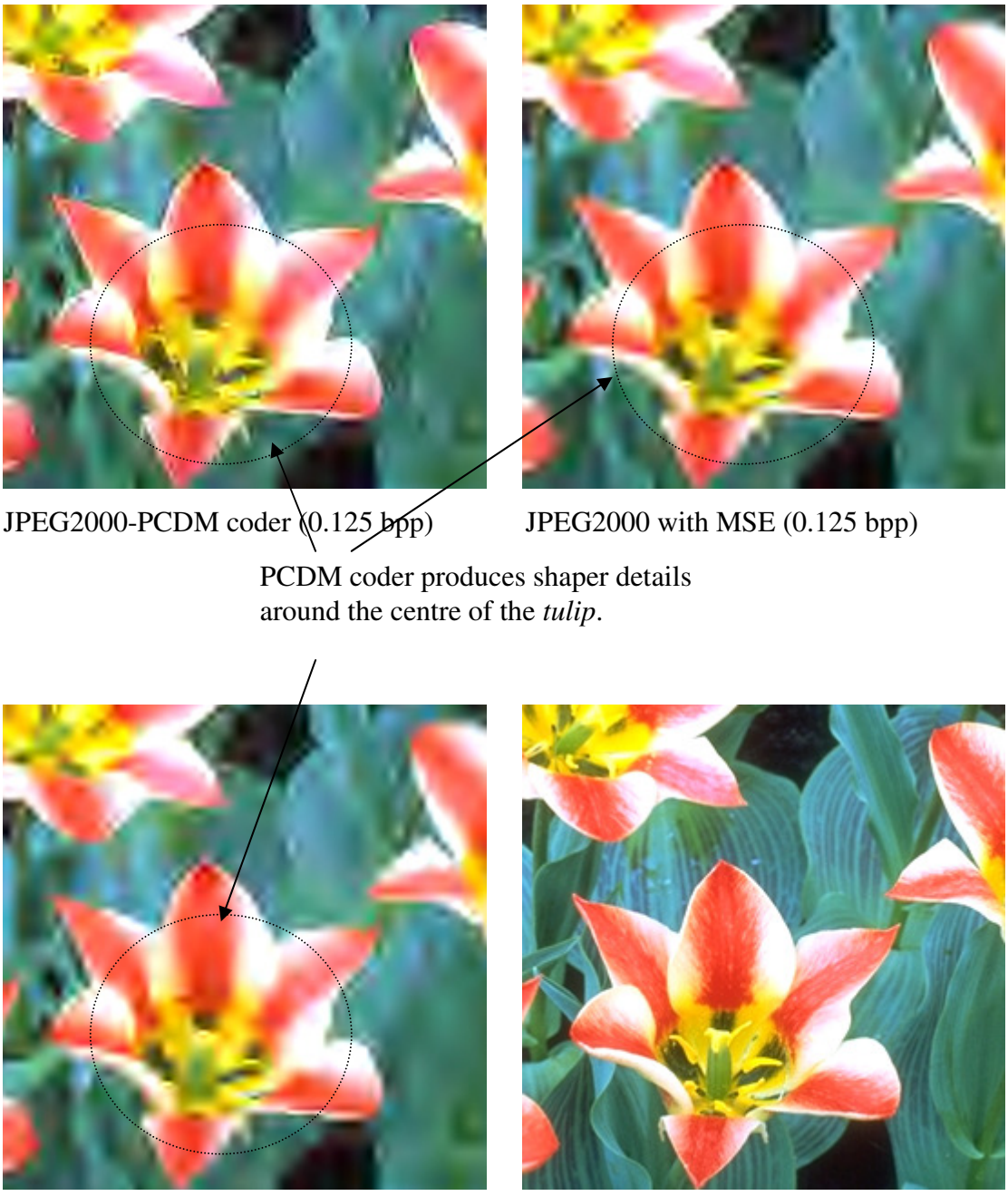



Figure 4.11 Cropped images of *tulip*.

JPEG2000-PCDM coder (0.25 bpp)



JPEG2000 with MSE (0.25 bpp)

PCDM coder preserves number details better than the other coders.

JPEG2000 with CVIS (0.25 bpp)

Original Uncompressed Image

Figure 4.12 Cropped images of *sail*.

## 4.7 Chapter Summary

Applying R-D function ensures that picture quality is optimised relative to bitrate. The MSE is commonly used as the distortion measure. However, the standard MSE has also been shown to be an inadequate measurement of perceived image quality metric [143, 144]. It is true that while some aspects of vision modelling design have been built into the VDM measure of the EBCOT, and also the CVIS of JPEG2000, a more comprehensive vision models based distortion measure can provide better estimation of visual distortion and thus improve the perceived image quality of JPEG2000 coded images.

The PCDM for colour image proposed in this chapter is embedded within the JPEG2000 [12, 158] core structure (Figure 4.1). Instead of using the MSE or the CVIS [12] as distortion measure in the R-D optimisation function, the Perceptual Colour Distortion Measure (PCDM) is employed. The PCDM considers contrast sensitivity and the masking mechanism of the HVS.

The masking model considers intra-band and inter-orientation masking for colour images. The PCDM expands the monochromatic PIDM mentioned in chapter 3 to colour space (*YCbCr*). This involves substantial calibration of the model parameters. While no best way has yet been devised for parameterising all the 42 parameters, the current approach to optimisation is carried out sequentially in an iterative manner in multiple passes. Subjective experiments conducted with 30 participants have shown superior perceived visual performance of the PCDM to that of the MSE or CVIS within the JPEG2000 coder.

# Chapter 5 Vision Model Based Perceptual Post Filtering of JPEG2000 Coded Colour Images

## *5.1 Introduction*

The coding paradigm of the JPEG2000 still image coding standard [12, 159-161] partitions the discrete wavelet transform of image into several codeblocks. Each codeblock is independently bitplane encoded, starting from the most significant bitplane (MSBP) to the least significant bitplane (LSBP) in multiple coding passes (with the exception of the MSBP in only one coding pass) [12]. The distortion reduction and the rate increase are collated and subsequently used to determine what coding passes to be included and/or excluded in the final embedded bitstream for each codeblock through the Post Compression Rate-distortion (PCRD) operation. For rate and quality scalable mode, once decided, those coding passes which are excluded from the PCRD algorithm are simply discarded (i.e., truncated) from the bitstreams. Based on bitrate constraint, the bitplane encoding from the MSBP to the LSBP and the PCRD optimisation as the procedure to subsequently discard coding passes of the bitplanes, will likely result in more bits being truncated (discarded) at the lower bitplanes than those at the higher bitplanes. The truncation of lower bitplanes provides an opportunity of restoring some of the lost visual information through bitplane recovery with a Perceptual Post Filtering (PPF) algorithm. At the heart of the PPF is a vision model that is used to perform the perceptual recovery operation from compressed images in the DWT domain. The PPF operates at the decoding stage and considers the contrast sensitivity, the intra-band masking and inter-orientation masking of the HVS.

The PPF assumes that there must be sufficient amount of information in a compressed image for it to operate effectively. For example, images coded at very low bitrates may not have sufficient information for bitplane recovery. PPF only operates on "significant" coefficients in codeblocks. The vision model used here operates on coded images as a reference set of data for bitplane recovery.

The Wavelet-based Image/Texture Coding Hybrid (WITCH) system proposed by Nadenau [162] works on the principle that most progressive bitplane coders encode bitplane starting from the MSB to the LSB, whereby the lower bitplanes are truncated to zeros under bitrate constraint. Implemented in the JPEG2000 decoder, the WITCH system injects stochastic noise generated based on model parameters from the encoder. The noise essentially synthesises the lost texture information at the decoder, thereby improves the texture quality of the reconstructed image. The stochastic noise injection is limited to the lowest three bitplane layers of all subblocks of typical size of 32 or 16 coefficients each (though other sizes are also applicable), and is applied only to the two highest frequency resolutions. This is in contrast to the PPF based decoder where the vision model is used to inject bits to recover perceived loss of information over the bitplane layers starting from the lowest to the highest bitplane subject to meeting some thresholds set at JND levels over all resolution levels except the isotropic (LL) band. The PPF algorithm is not only limited to texture information recovery alone, but also reconstruct perceived loss of structural details such as edges and lines.

## 5.2 Vision Modelling

The PPF utilises the vision model described in chapter 4 that considers the optical and cortical properties of the HVS as discussed previously in section 4.3. The contrast sensitivity is applied as a set of uniform frequency-specific sensitivity weights to modulate the DWT coefficients. Inter-orientation masking and intra-band spatial masking are taken as ratio operators. Mathematical descriptions are given in equations (4.4), (4.6) to (4.12) of section 4.3.

## 5.3 Coding Adaptation

At the decoding stage, the perceptual post filtering (PPF) algorithm (see Figure 5.1) is applied through progressive bitplane recovery of DWT coefficients for each codeblock, starting from the least significant bit, and then proceeds upwards to the most significant bit.

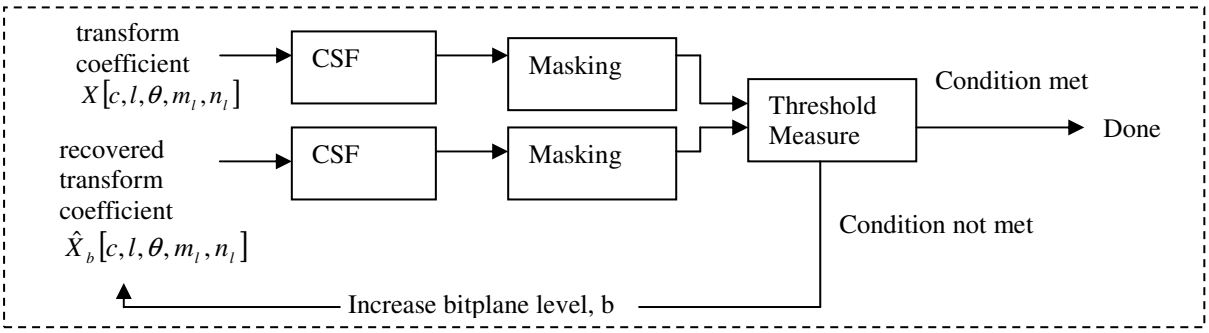



Figure 5.1 Block diagram of the structure of the Perceptual Post Filtering at the decoder. (The condition is met when $DR_{T,b_{\min}}(c,l,\theta,m_l,n_l) > T_D(c,l,\theta)$ and $PR_{p,b_{\min}}(c,l,\theta,m_l,n_l) < T_p(c,l,\theta)$ is satisfied.)

For each decoded transform coefficient, $X[c,l,\theta,m_l,n_l]$, and $X_M[c,l,\theta,m_l,n_l]$ being the magnitude portion of the coefficient, $X[c,l,\theta,m_l,n_l]$, and hereby we call $X_M[c,l,\theta,m_l,n_l]$ as the magnitude coefficient, the recovered bit-plane magnitude coefficient, $\hat{X}_{b,M}[c,l,\theta,m_l,n_l]$, up to bit plane level, b, is expressed as,

$$\hat{X}_{b,M}[c,l,\theta,m_l,n_l] = X_M[c,l,\theta,m_l,n_l] \mid (2^b - 1) \tag{5.1}$$

where $b \in \beta$, and $\beta = \{1,2,...,B\}$ is a set of bitplane level, and $B$ is the most significant bitplane of the magnitude coefficient $X_M[c,l,\theta,m_l,n_l]$. "|" denotes the bit-wise logical OR operator. The variables, $c$, $l$, $\theta$, are defined in section 4.3.3.

Similar to the CSF-weighted transform coefficient, $X_w[c,l,\theta,m_i,n_i]$, in equation (4.4), and the recovered CSF-weighted transform coefficient is expressed as follows,

$$\hat{X}_{w,b}[c,l,\theta,m_l,n_l] = C_w[c,l] \cdot \hat{X}_b[c,l,\theta,m_l,n_l] \tag{5.2}$$

where $C_w[c,l]$ is the CSF weight at frequency level, $l$, for colour component, $c$. $\hat{X}_b[c,l,\theta,m_l,n_l]$ is the recovered transform coefficient whose magnitude coefficient is $\hat{X}_{b,M}[c,l,\theta,m_l,n_l]$ which is computed in equation (5.1). Essentially, the bit plane recovery is applied to the magnitude portion of the transform coefficient only.

The perceptual distortion recovery, $DR_{T,b}$ , of each recovered CSF weighted coefficient for colour component $c \in \{Y, Cb, Cr\}$, is then defined as follows,

$$DR_{T,b}(c,l,\theta,m_l,n_l) = \sum_z g_{c,z} \left| R_{z,b}(c,l,\theta,m_l,n_l) - R_z(c,l,\theta,m_l,n_l) \right|^2 \tag{5.3}$$

where $R_z$ is the masking response of CSF-weighted transform coefficient at the decoder, and $R_{z,b}$ is the masking response of the recovered CSF-weighted transform coefficient at up to bitplane level b, and $z \in \{\Theta, \gamma\}$ with Θ and γ representing the inter-orientation and intra-band masking domains, respectively. $g_{c,z}$ are proportional gain factors which are used to determine the relative amount of contributions from inter-orientation and intra-band masking domains towards perceptual distortion recovery. (Note that the relative amounts of their contributions are not equal.)

The equation for the response $R_z(c,l,\theta,m_l,n_l)$ is taken directly from equation (4.6), and $R_{z,b}(c,l,\theta,m_l,n_l)$ is modified from equation (4.6), and is expressed as,

$$R_{z,b}(c,l,\theta,m_l,n_l) = k_{c,z} \frac{E_{z,b}(c,l,\theta,m_l,n_l)}{I_z(c,l,\theta,m_l,n_l) + \sigma_{c,z}^q} \tag{5.4}$$

Currently, $q$ , set at 2. $I_z(c,l,\theta,m_l,n_l)$, is the inhibition function from equations (4.8) and (4.10). The excitation functions, $E_{z,b}(c,l,\theta,m_l,n_l)$ , due to estimated CSF-weighted transform coefficient, are expressed as follows, respectively,

$$E_{\Theta,b}(c,l,\theta,m_l,n_l) = \left(\hat{X}_{w,b}[c,l,\theta,m_l,n_l]\right)^{p_{c,\Theta}} \tag{5.5}$$

$$E_{\gamma,b}(c,l,\theta,m_l,n_l) = \left(\hat{X}_{w,b}[c,l,\theta,m_l,n_l]\right)^{p_{c,\gamma}} \tag{5.6}$$

where $p_{c,z}$ are the exponents for inter-orientation masking and intra-band masking domains with $z \in \{\Theta, \gamma\}$.

$DR_{T,b}$ in equation (5.3) calculates the amount of perceived distortion recovery when the bits are added to the coefficient to form the recovered coefficient as the bit plane recovery proceeds from the lowest to the highest bit plane level. As the bitplane recovery proceeds from the lower bit plane to the higher bitplane, care must be taken to ensure that recovery process is not overdone. Otherwise, distortion may occur. What mechanism is used by the HVS to determine if the process is overdone is also not clear at this stage. Hence, a hypothetical perceptual percentage response, $PR_{p,b}(c,l,\theta,m_l,n_l)$, is introduced. The $PR_{p,b}(c,l,\theta,m_l,n_l)$ calculates the amount of hypothetical neuron energy response ratio that is altered as a result of adding bits to coefficients along the bitplane layers. The amount allowed for the percentage response cannot be too substantial as over correction may occur. The percentage response, $PR_{p,b}(c,l,\theta,m_l,n_l)$, is defined as,

$$PR_{p,b}(c,l,\theta,m_l,n_l) = \frac{R_{\Theta}(c,l,\theta,m_l,n_l) + R_{\gamma}(c,l,\theta,m_l,n_l)}{R_{\Theta,b}(c,l,\theta,m_l,n_l) + R_{\gamma,b}(c,l,\theta,m_l,n_l)} \tag{5.7}$$

where $R_{\Theta}(c,l,\theta,m_l,n_l)$ and $R_{\gamma}(c,l,\theta,m_l,n_l)$ are the inter-orientation and intra-band masking responses of CSF-weighted DWT coefficient, respectively. Similarly, $R_{\Theta,b}(c,l,\theta,m_l,n_l)$ and $R_{\gamma,b}(c,l,\theta,m_l,n_l)$ are the inter-orientation and intra-band masking responses of the recovered CSF-weighted DWT coefficient, respectively.

For each coefficient, the progressive bitplane recovery is achieved when the minimum bitplane level, $b_{\min}$ , is reached for that coefficient such that the condition

$DR_{T,b_{\min}}(c,l,\theta,m_l,n_l) > T_D(c,l,\theta)$ and $PR_{p,b_{\min}}(c,l,\theta,m_l,n_l) < T_p(c,l,\theta)$ is satisfied. Consequently, the final DWT coefficient is as follows,

$$\overline{X}[c,l,\theta,m_l,n_l] = \begin{cases} \hat{X}_{b_{\min}}[c,l,\theta,m_l,n_l] & ,if\ \vartheta\ is\ true\ and\ 1 \le b_{\min} \le B \\ X[c,l,\theta,m_l,n_l] & ,else \end{cases} \tag{5.8}$$

where $\vartheta = \{DR_{T,b_{min}}(c,l,\theta,m_l,n_l) > T_D(c,l,\theta) \quad and \quad PR_{p,b_{min}}(c,l,\theta,m_l,n_l) < T_p(c,l,\theta)\}$ The perceptual distortion recovery threshold, $T_D(c,l,\theta)$, and the perceptual percentage threshold, $T_p(c,l,\theta)$, are pairs of predetermined thresholds for the perceptual distortion recovery, $DR_{T,b}(c,l,\theta,m_l,n_l)$, and the perceptual percentage response, $PR_{p,b}(c,l,\theta,m_l,n_l)$, respectively, at resolution level $l = \{1,2,3,4,5\}$ and orientation $\theta = \{LH, HL, HH\}$. $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ are obtained through calibration as mentioned in section 5.4. Equation (5.8) ensures the bitplane recovery is achieved up to bit plane level, $b_{min}$, such that the $DR_{T,b_{min}}(c,l,\theta,m_l,n_l)$ is just above the threshod $T_D(c,l,\theta)$ but below the condition where over-correction is reached (i.e., $PR_{p,b_{min}}(c,l,\theta,m_l,n_l)$ is below the threshold, $T_p(c,l,\theta)$). In practice, $T_D(c,l,\theta)$ is very small and $b_{min}$ will usually be reached. Should bit plane recovery arrive beyond the highest bitplane, $B$, no recovery is allowed, and the transform coefficient remains unaltered. If at any time where $PR_{p,b_{min}}(c,l,\theta,m_l,n_l) \ge T_p(c,l,\theta)$ is reached before $DR_{T,b_{min}}(c,l,\theta,m_l,n_l) > T_D(c,l,\theta)$, no recovery is allowed, and the transform coefficient remains unaltered.

The progressive bitplane estimation is applied to all transform coefficients, $\overline{X}[c,l,\theta,m_l,n_l]$, at the decoder spanning all frequencies and orientation bands except the isotropic low pass band (LL) which is too sensitive to be included for bitplane recovery. The inverse DWT is then applied with the recovered transform coefficients and the unaffected coefficients at the isotropic low pass band to reconstruct the compressed image.

Note that the decoded sample values prior to bit plane recovery were obtained using mid-point dequantisation rule. During implementation, buffers are created to keep the samples after dequantisation, so that sufficient sample coefficients were obtained before they were bit plane recovered and then followed by inverse transform.

## 5.4 Model Parameterisation and Thresholding

The PPF utilizes the PCDM parameters in Tables 4.2. However, the set of thresholds for $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ requires some calibration to recover perceptually relevant information. These thresholds were set at the Just Noticeable Difference (JND) levels.

The calibration process involved a total of nine test images generated from three different source images (*barbara2*, *bikes*, *building2*), each at three different bitrates, namely, 1.0, 0.5 and 0.25 bpp. Test images were displayed on a 21-inch, 0.25 mm dot pitch Sun Monitor with a display resolution set to 1280×1024 pixels. The test images are attached in Figures B2, B4, and B5 of appendix B. During calibration, the images were displayed on the Monitor as illustrated in Figure 5.2 below.

The calibration starts with the *Y* colour component by adjusting the value of $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ sequentially while resetting the values of $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ of *Cr* and *Cb* colour components to zero.

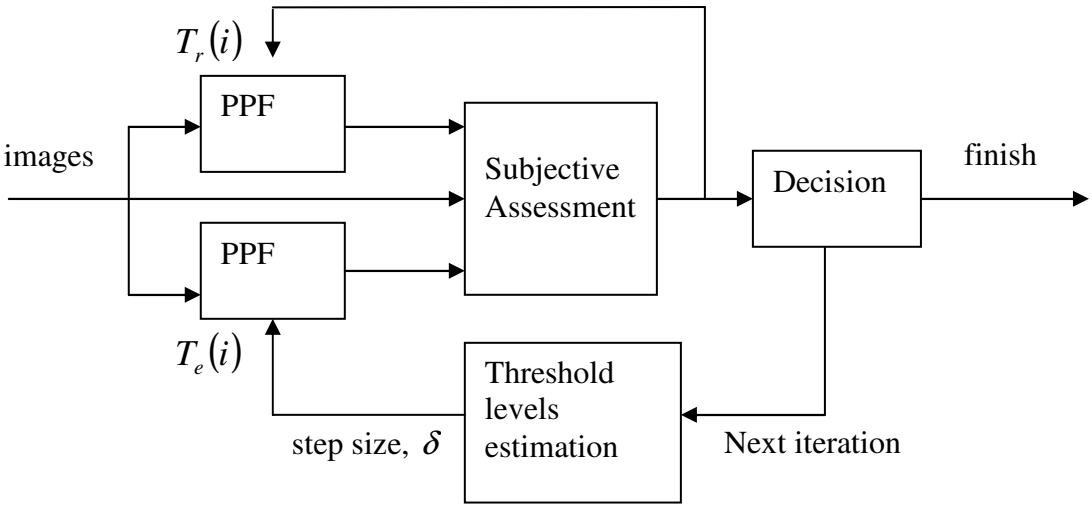



Figure 5.2 Calibration of parameters in the context of coder. (step size, $\delta$, is 0.0001 for $T_D(c,l,\theta)$, and varies from 0.05 to 0.01 for $T_p(c,l,\theta)$.)

Let $I_d(\chi)$ represents the complete set of images (nine images) produced by the PPF with threshold set, $\chi = \{T_e(i), T_\gamma(i)\}$. $T_e(i)$ and $T_r(i)$ represent the estimate and reference threshold sets of the current iteration, *i*, respectively. Consequently, the selection of the reference threshold set is expressed as,

$$T_\gamma(i+1) = f_s\left(I_d(T_e(i)), I_d(T_\gamma(i))\right) \tag{5.8}$$

where $f_s(.)$ is the force choice subjective assessment operation. The selection of $T_\gamma(i+1)$ is subjected to the similar assessment setup as depicted in Figure 4.6. Note that all the nine distorted images at bitrate of 1.0, 0.5, and 0.25 bpp of $T_e(i)$ were evaluated against those of $T_\gamma(i)$ with their original uncompressed images taken as additional reference set for force-choice test. The parameter set (i.e., either $T_e(i)$ or $T_\gamma(i)$) is selected as the better threshold set if it scores the higher number of subjective preferences (a value between 0 and 9) at JND level. The subjective preferred threshold set is then used in the next iteration $(i+1)$ as the reference threshold set, $T_\gamma(i+1)$. The next estimated threshold set $T_e(i+1)$ is determined by

the step size, $\delta$, which is set as 0.0001 for $T_D(c,l,\theta)$, and varies from 0.05 to 0.01 for $T_p(c,l,\theta)$.

The calibration process is described in details as follows,

a.1. All the values of $T_D$ are initialized 0 while all the values of $T_p$ are first initialized to 1. The step size increment, $\delta_D$, is set to 0.0001. Calibration starts with $T_D$ of *Y* component.

a.2. Start with level $l=1$, the $T_D$ for the three orientations $\theta=\{LH,HL,HH\}$ is increased by the step size $\delta_D=0.0001$. With three orientations, there will be seven possible sets of $T_D$ as follows,

| $T_D$ set | HH | HL | LH |
|---|---|---|---|
| Set 1 | No change | No change | Increased by $\delta_D$ |
| Set 2 | No change | Increased by $\delta_D$ | No change |
| Set 3 | No change | Increased by $\delta_D$ | Increased by $\delta_D$ |
| Set 4 | Increased by $\delta_D$ | No change | No change |
| Set 5 | Increased by $\delta_D$ | No change | Increased by $\delta_D$ |
| Set 6 | Increased by $\delta_D$ | Increased by $\delta_D$ | No change |
| Set 7 | Increased by $\delta_D$ | Increased by $\delta_D$ | Increased by $\delta_D$ |

For each set of the $T_D$, equation (5.8) is applied to determine the parameter set, either $T_\gamma(i)$ or $T_e(i)$, that has the higher preference score at JND level in a force-choice test. In the event that JND level has not been observed, the $T_D$ is increased by step size $\delta_D=0.0001$ starting from Sets 1 to 7 again. The increment process of the $T_D$ is repeated until the JND level is reached. The highest preference score of the seven sets of $T_D$ at JND level will be selected as the new parameter set for the next iteration *i+1*. In the event of more than two sets of $T_D$ having the highest preference score at JND level, the $T_D$ set with the highest index is chosen (e.g. Set 7 is chosen if Set 6 and Set 7 are having the same highest preference score).

a.3. With the new $T_D$ set determined in step a.2, calibrate $T_D$ for levels 2,3,4, and 5 in that order with step size increment $\delta_D = 0.0001$ by following the step in a.2.

a.4. Calibrate the $T_D$ for the *Cb* component with step size $\delta_D = 0.0001$ while setting all the $T_D$ values of *Y* component to half their values so as to give allowance for calibrating thresholds for other colour components. Calibrate $T_D$ for *Cb* component by following steps a.2 and a.3.

a.5. Calibrate the $T_D$ for the *Cr* component with step size $\delta_D = 0.0001$ while setting all the $T_D$ values of *Cb* component to half their values so as to give allowance for calibrating thresholds for other colour components. Calibrate $T_D$ for *Cr* component by following steps a.2 and a.3.

a.6. Next set the $T_D$ values of *Cr* component to half their values. Calibrate $T_p$ of *Y* component with step size $\delta_p = 0.05$.

a.7. Start with level $l = 1$, the $T_p$ for the three orientations $\theta = \{LH, HL, HH\}$ is decreased by the step size $\delta_p = 0.05$. With three orientations, there will be seven possible sets of $T_p$ as follows,

| $T_p$ set | HH | HL | LH |
|---|---|---|---|
| Set 1 | No change | No change | Decreased by $\delta_p$ |
| Set 2 | No change | Decreased by $\delta_p$ | No change |
| Set 3 | No change | Decreased by $\delta_p$ | Decreased by $\delta_p$ |
| Set 4 | Decreased by $\delta_p$ | No change | No change |
| Set 5 | Decreased by $\delta_p$ | No change | Decreased by $\delta_p$ |
| Set 6 | Decreased by $\delta_p$ | Decreased by $\delta_p$ | No change |
| Set 7 | Decreased by $\delta_p$ | Decreased by $\delta_p$ | Decreased by $\delta_p$ |

For each set of the $T_p$, equation (5.8) is applied to determine the parameter set, either $T_\gamma(i)$ or $T_e(i)$, that has the higher preference score at JND level in a force-choice test. In the event that JND level has not been observed, the $T_p$ is

increased by step size $\delta_p = 0.05$ starting from Sets 1 to 7 again. The increment process of the $T_p$ is repeated until the JND level is reached. The highest preference score of the seven sets of $T_p$ at JND level will be selected as the new parameter set for the next iteration *i+1*. In the event that more than two sets of $T_p$ having the highest preference score at JND level, $T_p$ set with the highest index is chosen (e.g. Set 7 is chosen if set 6 and set 7 are having the same highest preference score).

a.8. With the new $T_p$ set determined in step a.2, calibrate $T_p$ for levels 2,3,4, and 5 in that order with step size decrement of $\delta_p = 0.05$ by following the step in a.7.

a.9. Calibrate the $T_p$ for the *Cb* component with step size $\delta_p = 0.05$ while setting all of the $T_p$ values of *Y* component to half the sum of 1.0 and their previously calibrated values. Calibrate Tp for *Cb* component by following steps a.7 and a.8.

a.10. Calibrate the $T_p$ for the *Cr* component with step size $\delta_p = 0.05$ while setting all the $T_p$ values of *Cb* component to half sum of 1.0 and their previously calibrated values. Calibrate $T_p$ for *Cr* component by following steps a.7 and a.8.

a.11. Next set the $T_p$ values of *Cr* component to half the sum of 1.0 and their previously calibrated values.

a.12. Finally, beginning with $T_D$ of *Y* component at level *l=1*, recalibrate the $T_D$ and $T_p$ iteratively from steps a.2 to a.12 with increment of $T_D$ by step size of $\delta_D = 0.0001$ and decrement of $T_p$ by step size of $\delta_p = 0.01$, respectively. The manner in which the $T_D$ is set to half their previously calibrated values and $T_p$ is set to half the sum of 1.0 and its previously calibrated value from iteration *i* to *i+1* will ensure convergence of their threshold values at JND level.

When calibrating the value of each $T_D(c,l,\theta)$ or $T_p(c,l,\theta)$ values, the step size increment is applied to that parameter only until the visual difference of the image quality is just recognized. This is to ensure that the JND level is reached.

Once the thresholds of $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ of *Y* colour component are calibrated, their values are then set to half their values before the calibration proceeds to the next $T_D(c,l,\theta)$ or $T_p(c,l,\theta)$ parameter. The reason for setting thresholds of $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ of *Y* colour component to half their values is to prevent over correction of the threshold values as observed in the actual calibration experiment. It is found that simply reversing to the earlier threshold set for *Y* colour component did not allow proper calibration of threshold levels for both *Cb* and *Cr* colour components. The calibration then proceeds sequentially by calibrating $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ for all the colour components according to the same procedure as *Y* component.

The values of thresholds are presented in Tables 5.1 and 5.2. Note that the set of thresholds obtained are at most sub-optimal levels due to the fact that only one expert viewer was involved and only small sample of images were used in the calibration process. Hence, while the perceived visual quality of most images may be improved, It is possible that visual quality of some other images may be degraded by the distortion introduced in bit plane recovery process in the proposed PPF. Therefore, care must be taken to avoid over calibrating the $T_D(c,l,\theta)$ and $T_p(c,l,\theta)$ levels above the JND levels, as higher values may introduce ringing artifacts.

| Colour component | Orientation, $\theta$ | Frequency Level, $l$ | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Y | LH | 0.0004 | 0.0006 | 0.0008 | 0.0010 | 0.0015 |
| | HL | 0.0004 | 0.0006 | 0.0008 | 0.0010 | 0.0015 |
| | HH | 0.0004 | 0.0006 | 0.0008 | 0.0012 | 0.0015 |
| Cb | LH,HL,HH | 0.0002 | 0.0004 | 0.0006 | 0.0008 | 0.0015 |
| Cr | LH,HL,HH | 0.0002 | 0.0004 | 0.0006 | 0.0008 | 0.0015 |

Table 5.1 Predetermined threshold values for $T_D(c,l,\theta)$.

| Colour component | Orientation, $\theta$ | Frequency Level, $l$ | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| Y | LH,HL,HH | 0.90 | 0.85 | 0.75 | 0.5 | 0.35 |
| Cb | LH,HL,HH | 0.90 | 0.85 | 0.75 | 0.5 | 0.35 |
| Cr | LH,HL,HH | 0.95 | 0.90 | 0.85 | 0.80 | 0.75 |

Table 5.2 Predetermined threshold values for $T_p(c,l,\theta)$.

## 5.5 Experiment and Results

The PPF algorithm has been implemented in two ways:

- PPF algorithm at decoder for recovering images generated by JPEG2000 with PCDM coder (as implemented in Chapter 4), is hereby known as JPEG2000-PCDM-PPF,
- PPF algorithm at decoder for recovering images generated by JPEG2000 with MSE or CVIS distortion criterion, is hereby known as JPEG2000-MSE-PPF and JPEG2000-CVIS-PPF, respectively.

For both implementations, Comparative force-choice subjective tests [153, 154] were conducted on a total of 30 paired images generated from 10 different source images coded at three different bitrates 1.0, 0.5, and 0.25 bpp. The images were assessed on a 21-inch, 0.25 mm dot pitch Sun Monitor with a display resolution of 1280×1024 pixels by a group of voluntary viewers. The paired images were left and right pseudo randomised and their sequencing of paired images, numbered from 1 to 30, are also

randomised. The presentation of paired images and the order of presentation are similar to that depicted in Figure 4.8 but with the PPF algorithm, instead of the JPEG2000-PCDM. The viewing distance is set at two and a half times the height of the images [152] which were cropped to 512×512 pixels. The force-choice tests were conducted in a room with low illumination.

### 5.5.1 Implementation I

**Implementation I**: PPF algorithm with separate model parameterisation (SMP) at decoder for recovering images generated by EBCOT/JPEG2000 with PCDM coder, also denoted as JPEG2000-PCDM-PPF.

The 10 images (*goldhill*, *sail*, *pepper*, *lena*, *tulip*, *zelda*, *bikes*, *building2*, *lighthouse2*, and *stream*) were first encoded with the JPEG2000 with PCDM as implemented in chapter 4 at three different bitrates, i.e., 1.0, 0.5 and 0.25 bpp. The compressed bitstreams were then reconstructed with the PPF algorithm with SMP at the JPEG2000 decoder. Separate model parameterisation in PPF refers to the use of three different sets of model parameter values (as shown in Table 4.2) for *Y*, *Cb*, and *Cr*, respectively. The subjective assessment involves three separate rounds of testing, each with 30 pairs of images. There were nine participants for the first and the second rounds and eight participants for the third round. To ensure the quality of the subjective assessment, the participants were fully voluntary and had to be 18 years and above. There was also a good mix of male and female participants. Each participant was presented with the questionnaire set out in Appendix D. Basically, the participants had to choose one of the randomized images according to their preferences. Fifteen minutes interval (or days for some participants) was given between each round of test so as to minimise viewing fatigue. The complete set of test images is contained in the CD in Appendix J.

Rounds 1, 2, and 3 were designed to assess the performance of the images generated by JPEG2000-PCDM-PPF against those generated by (a) JPEG2000-PCDM as in chapter 4 without PPF algorithm, (b) JPEG2000-MSE, and (c) JPEG2000-CVIS,

respectively. With the CVIS criterion of JPEG2000, the images were coded with masking gain, $g$=0.5. The results of the subjective test are tabulated in Table 5.3.

| Image | Bitrate (bpp) | Raw Scores | | | | | |
|---|---|---|---|---|---|---|---|
| | | Round 1 | | Round 2 | | Round 3 | |
| | | A | B | A | C | A | D |
| goldhill | 1.0 | 5 | 4 | 7 | 2 | 7 | 1 |
| | 0.5 | 6 | 3 | 7 | 2 | 8 | 0 |
| | 0.25 | 5 | 4 | 8 | 1 | 6 | 2 |
| Sail | 1.0 | 5 | 4 | 8 | 1 | 7 | 1 |
| | 0.5 | 6 | 3 | 6 | 3 | 6 | 2 |
| | 0.25 | 9 | 0 | 6 | 3 | 5 | 3 |
| pepper | 1 | 4 | 5 | 5 | 4 | 5 | 3 |
| | 0.5 | 7 | 2 | 7 | 2 | 5 | 3 |
| | 0.25 | 6 | 3 | 5 | 4 | 7 | 1 |
| Lena | 1.0 | 6 | 3 | 6 | 3 | 3 | 5 |
| | 0.5 | 7 | 2 | 6 | 3 | 8 | 0 |
| | 0.25 | 4 | 5 | 8 | 1 | 7 | 1 |
| tulip | 1.0 | 7 | 2 | 6 | 3 | 6 | 2 |
| | 0.5 | 2 | 7 | 8 | 1 | 7 | 1 |
| | 0.25 | 6 | 3 | 5 | 4 | 4 | 4 |
| zelda | 1.0 | 3 | 6 | 4 | 5 | 4 | 4 |
| | 0.5 | 4 | 5 | 6 | 3 | 6 | 2 |
| | 0.25 | 3 | 6 | 3 | 6 | 5 | 3 |
| bikes | 1.0 | 4 | 5 | 7 | 2 | 8 | 0 |
| | 0.5 | 6 | 3 | 6 | 3 | 7 | 1 |
| | 0.25 | 8 | 1 | 7 | 2 | 8 | 0 |
| building2 | 1.0 | 8 | 1 | 7 | 2 | 6 | 2 |
| | 0.5 | 9 | 0 | 6 | 3 | 8 | 0 |
| | 0.25 | 9 | 0 | 6 | 3 | 6 | 2 |
| lighthouse2 | 1.0 | 3 | 6 | 6 | 3 | 8 | 0 |
| | 0.5 | 8 | 1 | 4 | 5 | 3 | 5 |
| | 0.25 | 9 | 0 | 7 | 2 | 7 | 1 |
| stream | 1.0 | 6 | 3 | 9 | 0 | 7 | 1 |
| | 0.5 | 7 | 2 | 8 | 1 | 6 | 2 |
| | 0.25 | 7 | 2 | 6 | 3 | 2 | 6 |

Table 5.3: Comparative Force-Choice Subjective Test Results
(A – preference for JPEG2000-PCDM-PPF, B – preference for JPEG2000-PCDM, C – preference for JPEG2000-MSE, D – preference for JPEG2000-CVIS)

The paired *t*-test [155] is used to evaluate the test results. The critical *t* for 9 d.f. and 2 d.f. at 95%, 99%, and 99.5% confidence levels (CI) are tabulated in Table 5.4.

| d.f. | $t_{0.05}$ | $t_{0.01}$ | $t_{0.005}$ |
|---|---|---|---|
| 9 | 1.8331 | 2.8214 | 3.2498 |
| 2 | 2.9200 | 6.9646 | 9.9248 |

Table 5.4 Critical *t* at 95% ($t_{0.05}$), 99% ($t_{0.01}$) and 99.5% ($t_{0.005}$) confidence intervals.

As there were only nine participants for rounds 1 and 2 tests and eight participants for the round 3 test, it will be necessary to combine the data sets before paired *t*-test analysis is performed. This is to ensure that the data set has reasonable number of sample points for meaningful statistical analysis. The data sets from the raw scores of Table 5.3 are grouped as follows:

- The scores of bitrate 1.0, 0.5 and 0.25 are combined for each type of images (i.e., categorising according to different images), and the data set is tabulated in Table 5.5. The 10 paired sets correspond to 9 degree of freedom (d.f.). This analysis only provides the overall performance according to different source images.
- The scores of the 10 images are combined for each bitrate (1.0, 0.5, 0.25 bpp), (i.e., categorising according to different bitrates), and tabulated in Table 5.6. The three paired sets correspond to 2 degree of freedom (d.f.). This provides overall performance analysis of the PPF according to different bitrates only.

The *t*-values are computed based on the grouped data sets of Table 5.5 and 5.6. For the paired *t*-test, 10 and 3 paired sets correspond to 9 and 2 degrees of freedom (d.f.), respectively. The *t*-values are tabulated in Table 5.7 for all rounds of tests.

| Image | P | | | | | |
|---|---|---|---|---|---|---|
| | Overall Scores | | | | | |
| | Round 1 | | Round 2 | | Round 3 | |
| | A | B | A | C | A | D |
| *Goldhill* | 16 | 11 | 22 | 5 | 21 | 3 |
| *Sail* | 20 | 7 | 20 | 7 | 18 | 6 |
| *Pepper* | 17 | 10 | 17 | 10 | 17 | 7 |
| *Lena* | 17 | 10 | 20 | 7 | 18 | 6 |
| *Tulip* | 15 | 12 | 19 | 8 | 17 | 7 |
| *Zelda* | 10 | 17 | 13 | 14 | 15 | 9 |
| *Bikes* | 18 | 9 | 20 | 7 | 23 | 1 |
| *building2* | 26 | 1 | 19 | 8 | 20 | 4 |
| *lighthouse2* | 20 | 7 | 17 | 10 | 18 | 6 |
| *Stream* | 20 | 7 | 23 | 4 | 15 | 9 |

Table 5.5: Comparative Force-Choice Subjective Test Results, categorized according to images. (By summing up the preferences of bitrate 1.0, 0.5 and 0.25 for each type of images. Note: A – preference for JPEG2000-PCDM-PPF, B – preference for JPEG2000-PCDM, C – preference for JPEG2000-MSE, D – preference for JPEG2000-CVIS)

| Bitrate (bpp) | Q | | | | | |
|---|---|---|---|---|---|---|
| | Overall Preference | | | | | |
| | Round 1 | | Round 2 | | Round 3 | |
| | A | B | A | C | A | D |
| 1.0 | 51 | 39 | 65 | 25 | 61 | 19 |
| 0.5 | 62 | 28 | 64 | 26 | 64 | 16 |
| 0.25 | 66 | 24 | 61 | 29 | 57 | 23 |

Table 5.6: Comparative Force-Choice Subjective Test Results, categorized according to bitrates. (By summing up the preferences of 10 images for each of the bitrates. Note: A – preference for JPEG2000-PCDM-PPF, B – preference for JPEG2000-PCDM, C – preference for JPEG2000-MSE, D – preference for JPEG2000-CVIS)

Evaluation of the test results is based on (a) all the 10 images covering all the bitrates combined, and (b) all the three bitrates (1.0, 0.5 and 0.25 bpp) covering all image types combined. For the paired *t*-test, 10 and 3 paired sets correspond to 9 and 2 degrees of freedom (d.f.), respectively. The *t*-values are tabulated in Table 5.7.

| | Types of Category | P | Q |
|---|---|---|---|
| | d.f. | 9 | 2 |
| Computed t-value | Round 1 | 3.3539 | 3.2705 |
| | Round 2 | 6.1492 | 15.2542 |
| | Round 3 | 7.7500 | 10.1927 |

Table 5.7 The *t*-values. (P) – categorising according to image, computed from Table 5.5. (Q) – categorising according to bitrates, computed from Table 5.6. d.f. denotes degree of freedom.

**a. Evaluation of Round 1 Test Result**

Let the Null Hypothesis ($H_0$) be "the perceived image quality of JPEG2000-PCDM-PPF is equivalent to or worse than the JPEG2000-PCDM", and the Alternate Hypothesis ($H_1$) is "the image quality of JPEG2000-PCDM-PPF is better than the JPEG2000-PCDM."

From Table 5.7, in (P), the *t*-value (3.3539) is higher than the critical *t* (3.2498) for 9 d.f. at 99.5% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, when categorising according to different source images, the perceived image quality produced by JPEG2000-PCDM-PPF based coder is overall statistically superior to the JPEG2000-PCDM based coder at 99.5% CI. For (Q), categorising according to bitrates, the perceived quality performance of JPEG2000-PCDM-PPF is statistically better than the JPEG2000-PCDM for 2 d.f. at 95% CI as the *t*-value (3.2705) is higher than the critical *t* (2.9200).

**b. Evaluation of Round 2 Test Result**

The Null Hypothesis ($H_0$) is assumed to be "the perceived image quality of JPEG2000-PCDM-PPF is equivalent to or worse than the JPEG2000-MSE", and the Alternate Hypothesis ($H_1$) is "the perceived image quality of JPEG2000-PCDM-PPF is better than the JPEG2000-MSE."

In (P), the *t*-value (6.1492) is higher than the critical *t* (3.2498) for 9 d.f. at 99.5% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, when categorising according to different source images, the perceived quality performance of JPEG2000-PCDM-PPF

coder is overall statistically superior to the JPEG2000-MSE at 99.5% CI. In (Q), when categorising according to bitrates, the perceived quality performance of JPEG2000-PCDM-PPF is statistically better than the JPEG2000-MSE for 2 d.f. at 99.5% CI as the *t*-value (15.2542) is higher than the critical *t* (9.9248).

**c. Evaluation of Round 3 Test Result**

The Null Hypothesis ($H_0$) is assumed to be "the perceived image quality of JPEG2000-PCDM-PPF is equivalent to or worse than the JPEG2000-CVIS", and the Alternate Hypothesis ($H_1$) is "the perceived image quality of JPEG2000-PCDM-PPF is better than the JPEG2000-CVIS."

In (P), the *t*-value (7.7500) is higher than the critical *t* (3.2498) for 9 d.f. at 99.5% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, when categorising according to different source images, the perceived quality performance of JPEG2000-PCDM-PPF coder is overall statistically superior to the JPEG2000-CVIS with 99.5% CI. For (Q), when categoring according to different bitrates, the perceived quality performance of PCDM-PPF is statistically better than the JPEG2000-MSE for 2 d.f. at 99.5% CI as the *t*-value (10.1927) is higher than the critical *t* (9.9248).

### 5.5.2 Implementation II

**Implementation II**: PPF algorithm with (a) common model parameterisation (CMP) and (b) separate model parameterisation (SMP) at decoder for recovering images generated by JPEG2000 with MSE or CVIS distortion criterion.

While SMP uses separate sets of parameter values for *Y*, *Cb*, and *Cr* colour components, CMP uses the same set of parameter values for all the three colour components. In CMP, The sets of parameter values for *Cb* and *Cr* colour components are exactly those used in the *Y* component.

The JPEG2000-MSE encoded images (i.e., *goldhill*, *sail*, *pepper*, *lena*, and *tulip*) and JPEG2000-CVIS encoded images (i.e., *zelda*, *bikes*, *building2*, *lighthouse2*, *stream*)

were reconstructed by the JPEG2000-PPF decoder. The qualities of these JPEG2000-PPF decoded images were evaluated against the images generated by JPEG2000-MSE or JPEG2000-CVIS, respectively. In the case of CVIS criterion, masking gain, $g$=0.5, was used. Three different subjective tests as described below were conducted with 5 participants, and their results are tabulated in Table 5.8. Similar to the other subjective assessments, to ensure the quality of the subjective assessment, the participants were fully voluntary and had to be 18 years and above. There was a good mix of male and female participants. Each participant was presented with the questionnaire set out in Appendix E. Basically, the participants had to choose one of the randomized images according to their preferences. To eliminate the fatigue factor, they were given a break before they were presented with the next sequence of randomized images. The complete set of test images is contained in the CD in Appendix H.

**Test #1**

Force-choice Comparative subjective test [153, 154] was conducted between images reconstructed by JPEG2000-PPF algorithm with CMP model against images reconstructed by JPEG2000-PPF with SMP model. The participants were asked to evaluate if the paired images were of similar quality. If they were not of similar quality, the participants had to make a preferred choice of the two. (Please refer to Part 1 of the questionnaire in Appendix E).

**Test #2**

Force-choice Comparative force-choice subjective test [153, 154] was conducted to evaluate the quality of images between those reconstructed by JPEG2000-PPF with CMP model against those generated by JPEG2000-MSE or JPEG2000-CVIS, respectively. The participants had to choose which image is of better quality when they were presented with the left-right randomised paired images. (Please refer to Part 2 of the questionnaire in Appendix E).

**Test #3**

In the third test, paired images between those reconstructed by the JPEG2000-PPF with SMP model and those generated by JPEG2000-MSE or JPEG2000-CVIS, respectively, were presented to the participants. The participants were asked to choose which image is of better quality. (Please refer to Part 3 of the questionnaire in Appendix E).

| Image | Bitrate (bpp) | Score (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Test 1 | | | Test 2 | | Test 3 | |
| | | A | B | N | A | C | B | C |
| *Goldhill* | 1.0 | 1 | 1 | 3 | 5 | 0 | 4 | 1 |
| | 0.5 | 2 | 0 | 3 | 5 | 0 | 5 | 0 |
| | 0.25 | 0 | 2 | 3 | 5 | 0 | 4 | 1 |
| *Sail* | 1.0 | 0 | 1 | 4 | 3 | 2 | 5 | 0 |
| | 0.5 | 1 | 1 | 3 | 5 | 0 | 4 | 1 |
| | 0.25 | 1 | 0 | 4 | 4 | 1 | 4 | 1 |
| *Pepper* | 1.0 | 1 | 0 | 4 | 4 | 1 | 4 | 1 |
| | 0.5 | 0 | 1 | 4 | 4 | 1 | 4 | 1 |
| | 0.25 | 1 | 0 | 4 | 3 | 2 | 3 | 2 |
| *Lena* | 1.0 | 0 | 2 | 3 | 4 | 1 | 4 | 1 |
| | 0.5 | 0 | 3 | 2 | 4 | 1 | 4 | 1 |
| | 0.25 | 1 | 2 | 2 | 3 | 2 | 4 | 1 |
| *Tulip* | 1.0 | 1 | 2 | 2 | 5 | 0 | 5 | 0 |
| | 0.5 | 0 | 1 | 4 | 4 | 1 | 5 | 0 |
| | 0.25 | 0 | 0 | 5 | 4 | 1 | 4 | 1 |
| *Zelda* | 1.0 | 1 | 1 | 3 | 4 | 1 | 5 | 0 |
| | 0.5 | 1 | 1 | 3 | 5 | 0 | 5 | 0 |
| | 0.25 | 2 | 0 | 3 | 5 | 0 | 5 | 0 |
| *Bikes* | 1.0 | 2 | 0 | 3 | 3 | 2 | 3 | 2 |
| | 0.5 | 2 | 0 | 3 | 5 | 0 | 4 | 1 |
| | 0.25 | 1 | 0 | 4 | 4 | 1 | 2 | 3 |
| *building2* | 1.0 | 2 | 1 | 2 | 4 | 1 | 4 | 1 |
| | 0.5 | 0 | 1 | 4 | 5 | 0 | 5 | 0 |
| | 0.25 | 1 | 0 | 4 | 5 | 0 | 5 | 0 |
| *lighthouse2* | 1.0 | 1 | 0 | 4 | 3 | 2 | 2 | 3 |
| | 0.5 | 1 | 0 | 4 | 3 | 2 | 2 | 3 |
| | 0.25 | 2 | 0 | 3 | 3 | 2 | 4 | 1 |
| *Stream* | 1.0 | 1 | 1 | 3 | 3 | 2 | 5 | 0 |
| | 0.5 | 1 | 0 | 4 | 4 | 1 | 4 | 1 |
| | 0.25 | 0 | 2 | 3 | 5 | 0 | 5 | 0 |

Table 5.8: Comparative Subjective Test Result.
(A – preference for JPEG2000-PPF with SMP model, B – preference for JPEG2000-PPF with CMP model, C – preference for JPEG2000, N – preference for neither A nor B. Note that goldhill, sail, pepper, lena, and tulip were encoded by JPEG2000 with MSE, while zelda, bikes, buildings, lighthouse2, and stream were encoded by JPEG2000 with CVIS)

Similar to the argument made in implementation I, as there were only six participants involved in the subjective test for implementation II, grouped data sets for paired *t*-test is statistically more meaningful. The grouped data sets derived from the raw scores of Table 5.3 are grouped as follows:

- The scores of bitrate 1.0, 0.5 and 0.25 are combined for each of the source images, and the data set is tabulated in Table 5.9. This analysis only provides the overall performance, categorised according to the different source image.
- The scores of the 10 images are combined for each bitrate (1.0, 0.5, 0.25 bpp), and the data set is tabulated in Table 5.10. This provides overall performance analysis of the PPF, categorized according to different bitrates.

The *t*-values are computed based on the group data set of Tables 5.9 and 5.10. For the paired *t*-test, 10 and 3 paired sets correspond to 9 and 2 degrees of freedom (d.f.), respectively. The *t*-values are tabulated in Table 5.11 for all Tests 1 to 3.

| Image | P | | | | | | |
|---|---|---|---|---|---|---|---|
| | Overall Preference | | | | | | |
| | Test 1 | | | Test 2 | | Test 3 | |
| | A | B | N | A | C | B | C |
| *goldhill* | 3 | 3 | 9 | 15 | 0 | 13 | 2 |
| *sail* | 2 | 2 | 11 | 12 | 3 | 13 | 2 |
| *pepper* | 2 | 1 | 12 | 11 | 4 | 11 | 4 |
| *lena* | 1 | 7 | 7 | 11 | 4 | 12 | 3 |
| *tulip* | 1 | 3 | 11 | 13 | 2 | 14 | 1 |
| *zelda* | 4 | 2 | 9 | 14 | 1 | 15 | 0 |
| *bikes* | 5 | 0 | 10 | 12 | 3 | 9 | 6 |
| *building2* | 3 | 2 | 10 | 14 | 1 | 14 | 1 |
| *lighthouse2* | 4 | 0 | 11 | 9 | 6 | 8 | 7 |
| *stream* | 2 | 3 | 10 | 12 | 3 | 14 | 1 |

Table 5.9: Comparative Subjective Test Result, categorized according to different source images. (By summing up the preference of bitrate 1.0, 0.5 and 0.25 for each type of images. Note: A – preference for JPEG2000-PPF with SMP model, B – preference for JPEG2000-PPF with CMP model, C – preference for JPEG2000, N – preference for neither A nor B. Note that goldhill, sail, pepper, lena, and tulip were encoded by JPEG2000 with MSE, while zelda, bikes, buildings, lighthouse2, and stream were encoded by JPEG2000 with CVIS.)

| Bitrate (bpp) | Q | | | | | | |
|---|---|---|---|---|---|---|---|
| | Overall Preference | | | | | | |
| | Round 1 | | | Round 2 | | Round 3 | |
| | A | B | N | A | C | B | C |
| 1.0 | 10 | 9 | 31 | 38 | 12 | 41 | 9 |
| 0.5 | 8 | 8 | 34 | 44 | 6 | 42 | 8 |
| 0.25 | 9 | 6 | 35 | 41 | 9 | 40 | 10 |

Table 5.10: Comparative Force-Choice Subjective Test Results, categorized according to bitrates. (By summing up the preferences of 10 images for each of the bitrates. Note: A – preference for JPEG2000-PPF with SMP model, B – preference for JPEG2000-PPF with CMP model, C – preference for JPEG2000, N – preference for neither A nor B. Note that goldhill, sail, pepper, lena, and tulip were encoded by JPEG2000 with MSE, while zelda, bikes, buildings, lighthouse2, and stream were encoded by JPEG2000 with CVIS.)

| | Types of Category | P | Q |
|---|---|---|---|
| | d.f. | 9 | 2 |
| Computed t-value | Test 1 | 0.4082 | 1.5119 |
| | Test 2 | 8.5903 | 9.2376 |
| | Test 3 | 6.5658 | 27.7128 |

Table 5.11 The *t*-values. (P) – categorising according to source images, computed from Table 5.9.
(Q) – categorising according to bitrates, computed from Table 5.10.

**a. Evaluation of Test 1 Result**

Let the Null Hypothesis ($H_0$) be "the perceived image quality of JPEG2000-PPF with SMP is equivalent to or worse than the JPEG2000-PPF with CMP", and the Alternate Hypothesis ($H_1$) be "the perceived image quality of JPEG2000-PPF with SMP is better than PPF with CMP."

From Table 5.11, in (P), the *t*-value (0.4082) is lower than the critical *t* (1.8331) for 9 d.f.. Hence the Null Hypothesis ($H_0$) cannot be rejected at 95% CI. Therefore, when categorising according to source images, the perceived quality performance of JPEG2000-PPF with SMP is overall statistically equivalent to or worse than the JPEG2000-PPF with CMP. Based on evaluation of (Q), when categorising according to bitrates, the *t*-value (1.5119) is lower than the critical *t* (2.9200) for 2 d.f.. Hence

the Null Hypothesis ($H_0$) cannot be rejected at 95% CI. Therefore, the perceived quality performance of JPEG2000-PPF with SMP is also statistically equivalent to or worse than the JPEG2000-PPF with CMP.

However, based on the raw score percentage computation, the overall percentage preferences of JPEG2000-PPF with SMP and JPEG2000-PPF with CMP are 18% and 15.3%, respectively, i.e., a 2.7% preference gain is observed for JPEG2000-PPF with SMP.

**b. Evaluation of Test 2 Result**

The Null Hypothesis ($H_0$) is "the perceived image quality of JPEG2000-PPF with SMP is equivalent to or worse than the JPEG2000 with MSE and CVIS criterion", and the Alternate Hypothesis ($H_1$) is "the perceived image quality of JPEG2000-PPF with SMP is better than JPEG2000 with MSE and CVIS criterion."

For (P), *t*-value (8.5903) is higher than the critical *t* (3.2498) for 9 d.f. at 99.5% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, when categorising according to source images, the perceived quality performance of JPEG2000-PPF with SMP is overall statistically superior to the JPEG2000-MSE and CVIS criteria at 99.5% CI. In (Q), when categorising according to bitrates, the images produced by JPEG2000-PPF with SMP has superior perceived quality to those of JPEG2000 with MSE and CVIS for 2 d.f. at 99% CI as the *t*-value (9.2376) is higher than the critical *t* (6.9646).

**c. Evaluation of Test 3 Result**

The Null Hypothesis ($H_0$) is "the perceived image quality of JPEG2000-PPF with CMP is equivalent to or worse than the JPEG2000 with MSE and CVIS criterion", and the Alternate Hypothesis ($H_1$) is "the perceived image quality of JPEG2000-PPF with CMP is better than JPEG2000 with MSE and CVIS criterion."

In (P), the *t*-value (6.5658) is higher than the critical *t* (3.2498) for 9 d.f. at 99.5% CI. Hence the Null Hypothesis ($H_0$) is rejected. Therefore, when categorising according

to source images, the perceived quality performance of JPEG2000-PPF with CMP is overall statistically superior to the JPEG2000-MSE and CVIS criterion at 99.5% CI. In (Q), the perceived quality performance of JPEG2000-PPF with CMP is statistically better than the JPEG2000-MSE and CVIS for 2 d.f. at 99.5% CI as the *t*-value (27.7128) is higher than the critical *t* (9.9248).

### 5.5.3 Discussion of Subjective Test Results

The subjective test results of implementation I suggests that the images constructed by JPEG2000-PCDM-PPF is overall statistically superior to those of the JPEG2000-PCDM. In comparison to JPEG2000-MSE and JPEG2000-CVIS, the JPEG2000-PCDM-PPF has also shown an overall improvement in perceived quality performance. This result is consistent with the subjective test result presented in chapter 4 for JPEG2000-PCDM. Hence, it can be inferred that JPEG2000-PCDM coded images' perceived quality can be further improved with PPF algorithm at the decoder. When comparing JPEG2000-PCDM-PPF with JPEG200-MSE and JPEG2000-CVIS, the JPEG2000-PCDM-PPF produces better perceived visual quality images at bitrates between 0.25 and 1.0 bpp.

As a reference, the objective measure, PSNR, of the test images produced by the JPEG2000-PCDM-PPF, JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVS is attached in Appendix G. It must be emphasized that images with higher PSNR as in Appendix G do not necessarily imply better perceived visual quality. On the contrary, some images produced by the JPEG2000-PCDM-PPF that possess lower PSNR were rated better perceived image quality than JPEG2000-MSE and/or JPEG2000-CVIS in the force-choice subjective assessments. It re-affirms that the MSE or the PSNR as an objective quality metric does not correlate well with the HVS's perception of image quality as reported by Girod [143] and Wang et al. [144].

The subjective test results of implementation II account for the quality preference between two different model parameterisations of the PPF algorithm: common model parameterisation (CMP) and separate model parameterisation (SMP). From the paired *t*-test analysis at 95% CI, there is no evidence to suggest that the SMP is

superior to the CMP. However, if the test results are calculated by overall percentage preferences of the raw scores, the computation shows that there is a 2.7% gain in preference for the JPEG2000-PPF with SMP over that of the JPEG2000-PPF with CMP. This suggests a very small but thus insignificant preference of images operated on by the JPEG2000-PPF with separate model parameterisation. Given that there is no significant statistical evidence to suggest SMP parameterization to have produced superior results than the CMP parameterization, the model with CMP may be desirable since the optimisation load is significantly reduced as only one third of the model parameters and thresholds are involved in the calibration process for CMP. Tests 2 and 3 results also suggest that the PPF algorithm alone without the PCDM can produce images with improved perceived quality than those of JPEG2000-MSE and JPEG2000-CVIS.

Notwithstanding, both models, PPF with SMP and CMP, consistently produce images with improved visual quality, as perceived by the participants, than both of the JPEG2000-MSE and JPEG2000-CVIS coders. Some examples of coded images by PCDM-PPF are shown in Figures 5.3, 5.4 and 5.5. A complete set of test images with various bit rates is provided in the CD in Appendix H.

In Figures 5.3b, 5.3c, 5.3d and 5.3e where circles are drawn around the region with "WKS", the word "WKS" and the leaves around it are clearer for Figure 5.3b than the others. In addition this region is enhanced for Figure 5.3b. For '*lena*' where an oval is drawn around here eyes, it can be seen that sharper eyes are observed for Figure 5.4b than the others. In the case of '*tulip*', the centre of the flower (i.e., the stigma) is also more visible and enhanced for Figure 5.5b.


Figure 5.3a: *building2* - original uncompressed


Figure 5.3b: *building2* - PPF with JPEG200-PCDM (0.25bpp)


Figure 5.3c: *building2* – JPEG2000-PCDM (0.25bpp)


Figure 5.3d: *building2* - JPEG2000-MSE (0.25bpp)

Figure 5.3e: *building2* - JPEG2000-CVIS (0.25bpp)

Figure 5.4a: *lena* - original uncompressed

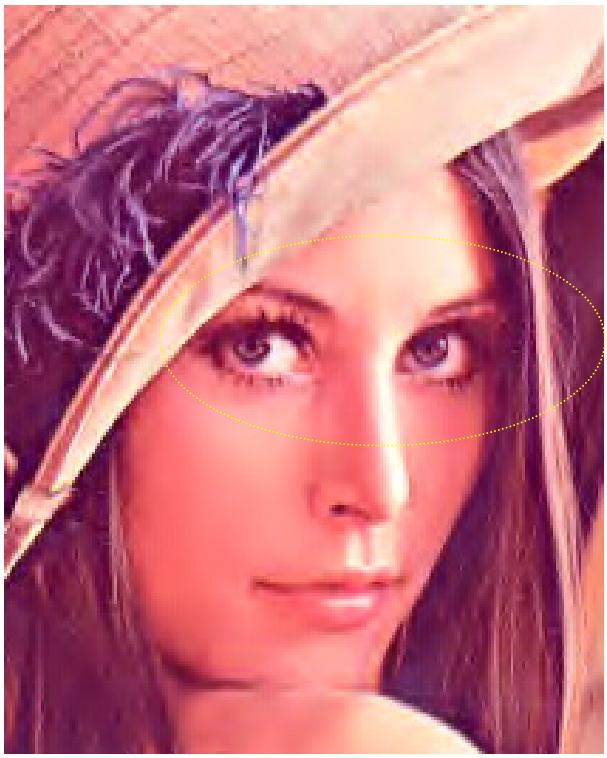


Figure 5.4b: *lena* - PPF with JPEG2000-PCDM (0.5bpp)

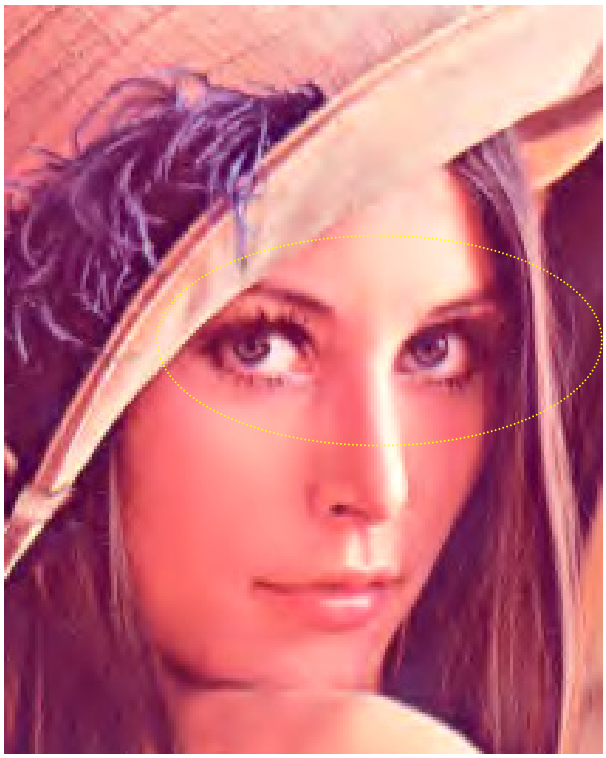


Figure 5.4c: *lena* – JPEG2000-PCDM (0.5bpp)

Figure 5.4d: *lena* - JPEG2000-MSE (0.5bpp)

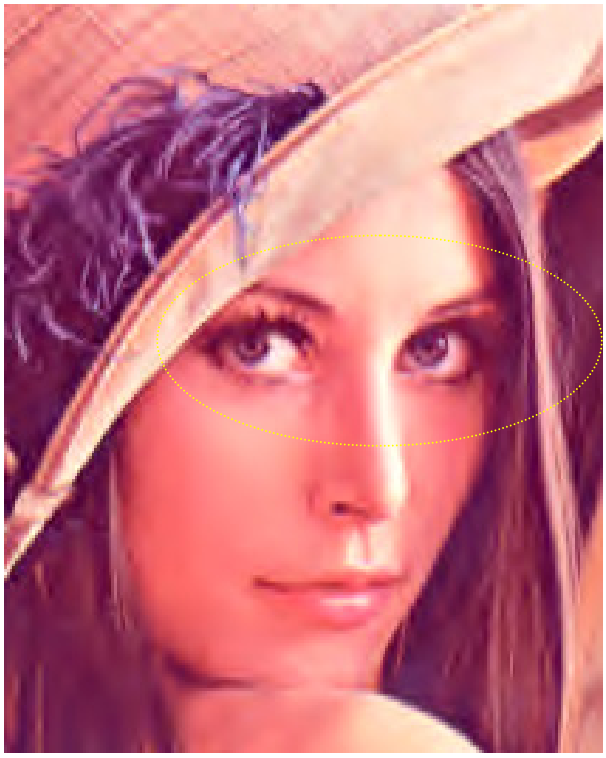


Figure 5.4e*: lena* - JPEG2000-CVIS (0.5bpp)

Figure 5.5a: *tulip* - original uncompressed

Figure 5.5b: *tulip* - PPF with JPEG2000-PCDM (1.0 bpp)

Figure 5.5c: *tulip* – JPEG2000-PCDM (1.0 bpp)

Figure 5.5d: *tulip* - JPEG2000-MSE (1.0 bpp)

Figure 5.5e: *tulip* - JPEG2000-CVIS (1.0 bpp)

## 5.6 Chapter summary

In this Chapter, a Perceptual Post Filtering (PPF) algorithm is proposed. This algorithm is used for perceptual recovery of bitplane information from the compressed images in the DWT domain at the JPEG2000 decoding stage. The visual properties of the HVS considered are the effects of contrast sensitivity, the intra-band masking and inter-orientation masking. At the decoding stage, the PPF is applied in progressive bitplane recovery manner on the transform coefficients for each code block, beginning with the LSB and proceeding upwards to the MSB (refer to Figure 5.1). With the exception of the isotropic low pass band, the PPF algorithm is applied to all transform coefficients of all frequency and orientation bands. Thereafter, an inverse DWT is applied to all these coefficients to reconstruct the compressed image.

In the calibration process, the PPF thresholds, perceptual distortion recovery and perceptual percentage thresholds are set to the JND level. The vision model parameters for the PCDM are taken directly from chapter 4.

Subjective test results of the PPF show that JPEG2000-PCDM-PPF offers visible improvement over JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVIS. Further subjective tests were undertaken to evaluate the PPF algorithm with common model parameterisation (CMP) and separate model parameterisation (SMP). The results showed that there is no statistical advantage of using SMP over CMP parameterization in delivering better visual performance. However, since the CMP is less complex than the SMP in terms of calibration, CMP has the implementation advantage. Without PCDM, the PPF algorithm implemented alone at the decoder has demonstrated better perceived visual quality of images than JPEG2000 without PPF.

# Chapter 6 Conclusion

## *6.1 Research Findings*

As technology becomes intertwined with every aspect of daily lives, and the use of images to convey information and knowledge in this fast paced modern world has increased, the demand for transmitting images quickly with the highest possible resolution and at an affordable cost and given infrastructure has heightened. Along with this surge, a large body of research has been carried out to deal with the all-important issue of data and image compression.

It must be acknowledged that much research has been undertaken out in the areas of the removal of statistical redundancies or "noise" in data as first mentioned by Shannon [10]. Some aspects of removal of statistical redundancies deal with the use of MSE (Mean Square Error) as a distortion measure, PSNR (Peak Signal Noise Ratio) or MAE. This body of research has seen the emergence of various imaged coders or image compression systems. The elements of an image coder are explained in Chapter 3 with a specific focus on transform based image coding and the elements involved in that system. These elements include spectral transformation, quantisation, and entropy encoding. Examples of transform based bit-plane image coders are the EZW [31], SPIHT [32] and EBCOT [14] which are also discussed in greater lengths in Chapter 3. The JPEG2000 standard [12] has also been ear marked as the new state-of-the-art standard for still image coding.

Along with this, some studies have also been carried out for image coding based on the human visual system, in particular, the effect of physiological characteristics of the human eye on the perception of visual signals. These perceptual image coders researched into the removal of other redundancies which are imperceptible to the human visual system. In simpler terms, some redundancies which are not noticeable by the human visual system could be eliminated to produce images with high compression ratios.

To gain a better understanding of these imperceptible characteristics of the human visual system, chapter 2 reviews the human visual system in some details. It covers the physiology of the human eye and neural connections associated with the HVS. The three aspects of the HVS are the optics, the visual pathway and the visual cortex. Extensive experimental studies have been carried out by various researchers to model the behaviours of these components. In particular, Watson and Solomon [27] have incorporated some crucial characteristics of the HVS in the modelling process, and they proposed the Contrast Gain Control model. This includes: 1) optical sensitivity of the human eye with contrast sensitivity function (CSF), 2) spectral decomposition to approximate frequency and orientation sensitivity of cortical neurons, and 3) masking phenomenon of the HVS by incorporating a normalised masking function.

The contribution of this thesis is the proposal of two perceptual image models based on the human visual system -- the Perceptual Colour Distortion Measure (PCDM) and Perceptual Post Filtering (PPF), both based on the human visual system, (in chapters 4 and 5 respectively). Both models exploit the inter-orientation masking and intra-band masking mechanism of the HVS.

The PCDM proposed in this thesis is a perceptual image coder and is an adaptation of the monochromatic based PIDM (Perceptual Image Distortion Metric) into colour based PCDM in the Y*C*b*C*r colour space. The resulting PCDM model is then adapted to the JPEG2000 encoder. Essentially, the proposed PCDM model incorporates a distortion measure that considers the effect of inter-orientation masking and intra-band masking mechanism of the HVS into the JPEG2000 coding system. This is in contrast to the widely used MSE distortion measure which is inaccurate in regard to perception by the HVS. In comparison to the CVIS, the PCDM is more elaborate and comprehensive as it includes inter-orientation masking. The PCDM model requires the calibration of 42 model parameters. Two sub-optimal values were obtained through a labourious and tedious process. Basically, it adopts the current approach to optimise the parameters -- sequential tuning iteratively. The sequential tuning of parameters may proceed for multiple passes (i.e., an approximation pass and multiple refinement passes) with different step sizes ($\delta_R$). This process has been explained in

greater detail in section 4.5 of chapter four. It appears that more than one set of sub-optimal values can be obtained to produce comparable performance in image quality.

The Perceptual Post Filtering (PPF) algorithm presented in chapter 5 is embedded into the JPEG2000 decoder to recover the perceived loss of information, and hence enhanced the perceived image quality. This is carried out through approximated bit-plane reconstruction. The core structure of vision model used in the PCDM is extended to the PPF algorithm, and it is used to achieve approximate bit-plane reconstruction in the PPF by considering the effects of the inter-orientation masking and intra-band masking of the HVS. The calibration of PPF thresholds is also undertaken at the Just-Noticeable-Difference (JND) level. The calibration process involves the use of nine test images generated from three different source images (*barbara2*, *bikes*, *building2*), each at three different bitrates, namely, 1.0, 0.5 and 0.25 bpp. A detailed description of the calibration process is presented in section 5.4 of chapter 5.

It is noted that while both PCDM and PPF employ the same vision model, the PCDM is embedded in the JPEG2000 encoder, whereas PPF is embedded in the JPEG2000 decoder. As JPEG2000 is being regarded as the state-of-the-art standard, some researchers have incorporated their proposed perceptual models in the JPEG2000 coding structure. Due to logistical restraints (i.e., software codes of other perceptual coders proposed by other researchers and their coded images are not made available in the public domain), it is uncertain to accurately compare and validate the performance results of these perceptual models through subjective assessment against the PCDM and PPF based coder proposed in this thesis. However, evaluation of the two models (PCDM and PPF) against the JPEG2000 benchmarks through subjective assessments indicated their performance improvement in the perceived image quality over the JPEG2000 with MSE and CVIS criteria. Moreover, as a reference, the objective measure, PSNR, is also investigated for the PCDM, PPF and JPEG2000 benchmarks. The findings re-affirm that the MSE or the PSNR as an objective quality metric does not correlate well with the HVS's perception of image quality as reported by Girod [143] and Wang et al. [144].

For the PCDM model, subjective assessments have been carried out with 30 viewers and the experimental results showed that the PCDM provided improved visual performance over the JPEG2000 with MSE and CVIS criteria, especially for the low (0.125bpp) and intermediate bitrates (0.5bpp). This improvement of image quality at low and intermediate bit rates is a promising result if its potential to be applied to software applications, file transfer applications can be explored further.

Two separate assessments have been undertaken to evaluate the PPF algorithm. Assessment one involves a performance evaluation of the JPEG2000-PCDM coder with the PPF algorithm with separate model parameterisation (SMP) against the JPEG2000-PCDM, the JPEG2000-MSE and JPEG2000-CVIS, all without the PPF algorithm. Assessment two involves a performance evaluation of the PPF algorithm with common model parameterisation (CMP) against JPEG2000-MSE and JPEG-CVIS. Test results have shown that both the PPF alone and PPF with PCDM improved performance over these JPEG2000 benchmarks. However, further subjective assessments of the PPF algorithm do not suggest any difference between the use of CMP or SMP for the PPF model.

The subjective assessment also suggests that the use of both the PCDM in the encoder and the PPF in the decoder in the JPEG2000 framework improves the visual performance as compared to when PCDM is used alone.

## *6.2 Further Research*

Thus far, the PCDM model has shown promising results in lossy perceptual compression. Attempts to test its performance for perceptually lossless compression for colour image are on-going. The proposed approach is through a bit-plane truncation of the samples with a vision model similar to that proposed in PCDM and PPF, with the bit-plane truncation achieved at JND level. This approach has been reported in Wu [16] for medical images. Hence there is definitely scope for this model to be further developed for colour images. Furthermore, not all the psychovisual characteristics of the HVS have been fully incorporated into the vision

model, e.g. inter-band masking between subbands of different frequency levels. The vision model could be developed along these lines for both the monochromatic and colour images.

The calibration of the model parameters currently produces sub-optimal values. More extensive calibration could lead to more accurate model parameters and yield more favourable results in terms of image quality and compression ratios. In addition, the present calibration algorithm is both tedious and slow; further research is required to develop a better and faster calibration algorithm for optimizing the model parameters of the proposed PCDM model and PPF algorithm.

Having said that, the proposed PCDM models and PPF algorithm, having produced improved image quality as compared with the JPEG2000-MSE and JPEG2000-CVIS, especially at low (0.125bpp) and intermediate bit rates (0.5bpp) is a promising result. Further research could be undertaken to assess its potential to be used in software applications and data transfer and storage purposes.

# Bibliography


[1] D. Salomon, *Data Compression*, 4th ed, California State University, 2007.

[2] D. Wu, D. M. Tan, and H. R. Wu, "Vision Model Based Approach to Medical Image Compression," International Symposium on Consumer Electronics, Sydney, Australia, 2003.

[3] M. Weingberger, G. Seroussi, and G. Sapiro, "The Loco-I Lossless Image Compression Algorithm: Principles and Standardisation into JPEG-LS," *IEEE Transaction of Image Processing*, 2000.

[4] ACR-NEMA, "DICOM Standard," Available:http://medical.nema.org/, 2006.

[5] B. J. Erickson, A. Manduca, P. Palisson, K. R. Persons, F. Earnest, V. Savcemko, and N. J. Hangiandreou, "Wavelet Compression of Medical Images," *Radiology*, vol. 206, pp. 599-607, 1998.

[6] N. Lin, T. Yu, and A. K. Chan, "Perceptually Lossless Wavelet-based Compression for Medical Images," Proceedings of SPIE on Medical Imaging, pp. 763-770, 1997.

[7] Y. Chee and K. Park, "Medical Image Compression Using the Characteristics of Human Visual System," Proceedings of 16th Annual International Conference of the IEEE, Nov 1994.

[8] J. Bradley, C. Brislawn, and T. Hopper, "The FBI Wavelet/Scalar Quantization Standard for Gray-Scale Fingerprint Image Compression," Proc. of SPIE, pp. 293-304, 1993.

[9] C. Brislawn, J. Bradley, R. Onyshczak, and T. Hopper, "The FBI Compression Standard for Digitized Fingerprint Images," Proc. of SPIE, pp. 344-355, 1996.

[10] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379-424 and 623-656, Jul and Oct 1948.

[11] G. K. Wallace, "The JPEG Still Picture Compression Standard," *Communication ACM*, vol. 34, 1991.

[12] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Boston, Kluwer Academic Publishers, Nov 2001.

[13] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 Still Image Coding System: An Overview," *IEEE Transactions on Consumer Electronics*, vol. 46, pp. 1103-1127, Nov 2000.

[14] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, pp. 1158-1170, Jul 2000.

[15] D. M. Tan, H. R. Wu, and Z. Yu, "Perceptual Coding of Digital Monochrome Images," *IEEE Signal Processing Letters*, vol. 11, pp. 239-242, Feb 2004.

[16] D. Wu, D. M. Tan, M. Baird, J. DeCampo, C. White, and H. R. Wu, "Perceptually Lossless Medical Image Coding," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 335-344, March 2006.

[17] D. H. Hubel, *Eye, Brain and Vision*, W.H. Freeman & Company, 1995.

[18] M. J. Farah, *The Cognitive Neuroscience of Vision*, Blackwell Publishers Inc., 2000.

[19] R. L. De Valois, E. W. Yund, and N. Hepler, "The Orientation and Direction Selectivity of Cells in Macaque Visual Cortex," *Vision Research*, vol. 22, pp. 531-544, 1982.

[20] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial Frequency Selectivity of Cells in Macaque Visual Cortex," *Vision Research*, vol. 22, pp. 545-559, 1982.

[21] D. H. Hubel and T. N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex," *Journal of Physiology*, vol. 160, pp. 106-154, Jan 1962.

[22] G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *Journal of the Optical Society of America*, vol. 70, pp. 1458-1471, 1980.

[23] P. C. Teo and D. J. Heeger, "Perceptual Image Distortion," Proceedings of IEEE International Conference on Image Processing, Austin, Texas, pp. 982-986, 13-16 Nov 1994.

[24] G. M. Boynton and J. M. Foley, "A New Model of Human Luminace Pattern Vision Mechanism : Analysis of the Effects of Pattern Orientation, Spatial Phase and Temporal Frequency " Proceeding Of SPIE: Computational Vision Based on Neurology, pp. 32-42, 1994.

[25] J. M. Foley, "Human Luminance Pattern-Vision Mechanisms: Masking Experiments Require a New Model," *Journal of the Optical Society of America*, vol. 11, pp. 1710-1719, 1994.

[26] J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg, "A Model of Visual Masking for Computer Graphics," in *Computer Graphics*, vol. 31, Annual Conference Series, 1997.

[27] A. B. Watson and J. A. Solomon, "A model of Visual Contrast Gain Control and Pattern Masking," *Journal of the Optical Society of America*, vol. 14, pp. 2379-2391, Sep 1997.

[28] D. H. Hubel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *Journal of Physiology*, vol. 195, pp. 215-243, 1968.

[29] D. H. Hubel, "Single Unit Activity in Lateral Geniculate Body and Optic Tract of Unrestrained Cats," *Journal of Physiology*, vol. 150, pp. 91-107, 1959.

[30] D. H. Hubel and T. N. Wiesel, "Receptive Fields of Single Neurones in the Cat's Striate Cortex," *Journal of Physiology*, vol. 148, pp. 574-591, Oct 1959.

[31] J. M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3445-3462, Dec 1993.

[32] 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*, vol. 6, Jun 1996.

[33] A. M. Burt, *Textbook of Neuroanatomy*, W. B. Saunders Company, 1993.

[34] B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.

[35] P. K. Kaiser and R. M. Boynton, *Human Color Vision*, 2 ed, Optical Society of America, Wasington DC, 1996.

[36] E. D. Montag and M. D. Fairchild, "Fundamentals of Human Vision and Vision Modeling," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds., CRC Press, 2006.

[37] J. V. Forrester, A. D. Dick, P. G. McMenamin, and F. Roberts, *The Eye: Basic Sciences in Practice*, Third ed, Elsevier Limited 2008.

[38] G. H. Bell, D. Emslie-smith, and C. R. Paterson, *Textbook of Physiology and Biochemistry*, 9 ed, Churchill Livingstone, 1976.

[39] D. Malacara, *Color Vision and Colorimetry*. Bellingham, Washington USA, SPIE Press, 2002.

[40] J. P. C. Southall and editor, *Helmholtz's Treatise on Physiological Optics*, vol. 1. New York, Dover Publications Inc., 1962.
[41] K. N. Ogle, *Optics*. Springfield Illinois, USA, Charles C. Thomas Publisher, 1961.
[42] J. A. Ferwerda, "Fundamentals of Spatial Vision," pp. 1-27, 1998.
[43] S. W. Kuffler, "Discharge Pattern and Functional Organization of Mammalian Retina," *Journal of Neurophsiology*, vol. 16, pp. 37-68, 1953.
[44] B. B. Boycott and H. Wasslen, "The morphological types of ganglion cells of the domestic cat's retina," *The Journal of Physiology*, vol. 240, pp. 397-419, 1974.
[45] Y. Fukuda and J. Stone, "Retinal distribution and central projections of Y, X, and W cells of the cat's retina," *Journal Neurophysiology*, vol. 37, pp. 749-772, 1974.
[46] A. Parent, *Carpenter's Human Neuroanatomy*, 9th ed, Williams & Wilkins, 1996.
[47] J. E. Dowling, *The Retina: An Approachable Part of the Brain*, Cambridge Harvard University Press, 1987.
[48] P. H. Schiller and N. K. Logothetis, "The Color-opponent and Broad-band Channels of the Primate Visual System," *Trends in Neurosciences*, vol. 13, pp. 392-398, 1990.
[49] S. H. C. Hendry and R. C. Reid, "The Koniocellular Pathway in Primate Vision," *Annual Review of Neuroscience*, vol. 23, pp. 127-153, 2000.
[50] L. C. Sincich and J. C. Horton, "The circuitry of V1 and V2: Integration of Color, Form and Motion," *Annual Review of Neuroscience*, vol. 28, pp. 303-326, 2005.
[51] L. Maffei and A. Fiorentini, "Retinogeniculate Convergence and Analysis of Contrast," *Journal of Neurophyciology*, vol. 35, pp. 65-72, 1972.
[52] S. Zeki, "Colour Vision and Functional specialisation in the visual cortex " *Discussions in Neuroscience* vol. VI, 1990.
[53] C. J. Tinsley, B. S. Webb, N. E. Barraclough, C. J. Vincent, A. Parker, and A. M. Derrington, "The Nature of V1 Neural Responses to 2D Moving Patterns Depends on Receptive-field Structure in the Marmoset Monkey," *Journal of Neurophsiology*, vol. 90, pp. 930-937, 2003.
[54] M. S. Livingstone and D. H. Hubel, "Segregation of Form, Colour, Movement, and Depth: Anatomy, Physiology and Perception," *Science*, vol. 240, 1988.
[55] D. H. Hubel, *Eye, Brain and Vision*, W.H. Freeman & Company, 1998.
[56] D. H. Hubel, *Eye, Brain, and Vision*. New York, Scientific American Library, 1988.
[57] F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *Journal of Physiology*, vol. 186, pp. 558-578, 1966.
[58] G. Westheimer and F. W. Campbell, "Light Distribution in the Image formed by the Living Human Eye," *Journal Optical Society of America*, vol. 52, pp. 1040-1045, 1962.
[59] G. Westheimer, "Optical and Motor Factors in the Formation of the Retinal Image," *Journal Optical Society of America*, vol. 53, pp. 86-93, 1963.
[60] J. Krauskopf, "Light Distribution in Human Retinal Images," *Journal Optical Society of America*, vol. 52, pp. 1046-1050, 1962.
[61] G. Westheimer, "IV: The Eye as an Optical Instrument," in *Handbook of Perception on Human Performance*, vol. 1, K. R. Boff, L. Kaufman, and J. P. Thomas, Eds., John Wiley & Sons, 1986.

[62] J. L. Mannos and D. J. Sakrison, "The Effects of a Visual Fidelity Criterion on the Encoding of Images," *IEEE Transactions on Information Theory*, vol. IT-20, pp. 525-536, July, 1974.
[63] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the Visibility of Grating," *Journal of Physiology*, vol. 197, pp. 551-566, 1968.
[64] F. L. Van Nes and M. A. Bournan, "Spatial Modulation Transfer Function in the Human Eye," *Journal of the Optical Society of America*, vol. 57, pp. 401-406, 1967.
[65] E. G. Boring, *A History of Experimental Psychology*, 2 ed, Appleton Century Crofts, Inc., 1957.
[66] A. A. Michelson, *Studies in Optics*, 3rd ed, Pheonix Books, 1962.
[67] S. Winkler, "Issues in Vision Modeling for Perceptual Video Quality Assessment," *Signal Processing*, vol. 78, pp. 231-252, Oct 1999.
[68] E. Peli, "Contrast in Complex Images," *Journal of the Optical Society of America*, vol. 7, pp. 2032-2040, Oct 1990.
[69] P. C. Teo and D. J. Heeger, "Perceptual Image Distortion," Proceedings of SPIE, pp. 127-141, 1994.
[70] J. M. Foley and G. M. Boynton, "A new model of human luminance pattern vision mechanisms: Analysis of the effects of pattern orientation, spatial phase, and temporal frequency," Proceedings of SPIE, pp. 32-42, 1993.
[71] D. J. Heeger, "Normalization of cell responses in cat simple cells," *Visual Neuroscience*, vol. 9, pp. 181-198, 1992.
[72] E. P. Simoncelli and E. H. Adelson, "Non-separable extensions of quadrature mirror filters to multiple dimensions," Proceedings of IEEE, pp. 652-664, 1990.
[73] S. J. Daly, W. Zeng, and S. Lei, "Visual Masking in Wavelet Compression for JPEG2000," IS&T/SPIE Conf. Image and Video Communications and Processing, pp. 66-80, 2000.
[74] C. W. Thomas, G. C. Gilmore, and F. L. Royer, "Models of Contrast Sensitivity in Human Vision," *IEEE Transactions on Systems, Man, Cybernetics*, vol. 23, pp. 857-864, 1993.
[75] K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann Series, 2000.
[76] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed, Prentice Hall Inc., 2002.
[77] J. J. Hwang, H. R. Wu, and M. K. Rao, "Digital Picture Compression and Coding Structure," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds., CRC Press, pp. 1-43, 2005.
[78] P. Elias, "Universal Codeword Sets and Representations of the Integers," *IEEE Transactions on Information Theory*, vol. IT-21, pp. 194-203, Mar 1975.
[79] P. Fenwick, "Punctured Elias Codes for variable-length coding of the integers," 137, Department of Computer Science, University of Auckland Dec 1996.
[80] S. W. Golomb, "Run-Length Encodings," *IEEE Transactions on Information Theory*, vol. IT-12, pp. 399-401, 1966.
[81] D. Huffman, "A Method for the Construction of Minimum Redundancy Codes," *Proceedings of IRE*, vol. 40, pp. 1098-1101, 1952.
[82] D. E. Knuth, "Dynamic Huffman Coding," *Journal of Algorithms*, vol. 6, pp. 163-180, 1985.

[83] N. Abramson, *Information Theory and Coding*. New York, McGraw-Hill, 1963.
[84] I. H. Witten, M. N. Radford, and G. C. John, "Arithmetic Coding for Data Compression," *Communication of ACM*, vol. 30, pp. 520-540, 1987.
[85] A. Moffat, R. Neal, and I. Witten, "Arithmetic Coding Revisited," *ACM Transactions on Information Systems*, vol. 16, pp. 256-294, 1998.
[86] B. P. Tunstall, "Synthesis of Noiseless Compression Codes," vol. Ph.D. dissertation. Atlanta, GA, Georgia Institute of Technology, 1967.
[87] R. Clarke, *Transform Coding of Images*. Orlando FL, Academic Press, Inc., 1985.
[88] P. A. Wintz, "Transform Picture Coding," *Proceedings of IEEE*, vol. 60, pp. 809-820, Jul 1972.
[89] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey, Prentice-Hall, 1989.
[90] M. Vetterli and J. Kovacevic, *Wavelets and Suband Coding*. Eaglewood Cliffs, New Jersey 07632, Prentice-Hall, 1995.
[91] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*. New Jersey, Prentice Hall, 1998.
[92] I. Daubechies, "The Wavelet Transform, Time-Frequency Localisation and Signal Analysis," *IEEE Transactions on Information Theory*, vol. 36, pp. 961-1005, Sep 1990.
[93] D. Esteban and C. Galand, "Application of Quadrature Mirror Filters to Split Voice Coding System," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 191-195, 1977.
[94] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. ASSP*, vol. ASSP-34, pp. 1278-1288, Oct 1986.
[95] C. K. Chui, *An Introduction to Wavelets*, Academic Press, 1992.
[96] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE TRansactions on Computers*, vol. C-23, pp. 90-93, Jan 1974.
[97] P. S. Addison, *The Illustrated Wavelet Transform Handbook*, Institute of Physics Publishing, 2002.
[98] V. R. Algazi and D. J. Sakrison, "On the Optimality of the Karhunen-Loeve Expension," *IEEE Transactions on Information Theory*, vol. IT-15, pp. 319-321, Mar 1969.
[99] O. Rioul, "Regular Wavelets : A Discrete Time Approach," *IEEE Transactions on Signal Processing*, vol. 41, pp. 3572-3579, Dec 1993.
[100] I. Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," *Communcation of Pure and Applied Math.*, vol. 41, pp. 909-996, 1988.
[101] I. Daubechies, *Ten Lectures on Wavelets*. Montpelier, Vermont, Capital City Press, 1992.
[102] M. Vetterli and C. Herley, "Wavelet and Filter Banks: Theory and Design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207-2232, Sep 1992.
[103] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.
[104] E. J. Stollnitz, A. D. DeRose, and D. H. Salesin, "Wavelets for Computer Graphics: A Primer Part 2," *IEEE Computer Graphics and Applications*, vol. 15, pp. 75-85, Jul 1995.

[105] E. J. Stollnitz, A. D. DerRose, and D. H. Salesin, "Wavelets for Computer Graphics: A Primer Part 1," *IEEE Computer Graphics and Applications*, vol. 15, pp. 76-84, May 1995.
[106] M. K. Rao and P. Yip, *Discrete Cosine Transforms: Algorithms, Advantages and Application*, Academic Press, 1990.
[107] W. Watkinson, *The MPEG Handbook*, 2nd ed, Focal Press, Nov 2004.
[108] A. Croisier, D. Esteban, and C. Galand, "Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques," Int. Conf. on Inform. Sciences and Systems, pp. 443-446, Aug 1976.
[109] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.
[110] A. Gersho, "Quantization," *IEEE Communications Magazine*, vol. 15, pp. 16-29, Sep 1977.
[111] J. Max, "Quantizing for Minimum Distortion," *IRE Trans. Info. Theory*, vol. IT-6, pp. 7-12, 1960.
[112] S. P. Lloyd, "Least Square Quantisation in PCM," *IEEE Transaction of Information Theory*, vol. IT-28, pp. 129-137, Mar 1982.
[113] Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantization Design," *IEEE Transactions on Communications*, vol. 28, pp. 84-95, 1980.
[114] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*, Kulwer Academic Publishers, 1992.
[115] J. W. Schwartz and R. C. Barker, "Bit-Plane Encoding: A Technique for Source Encoding," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-2, pp. 385-392, Jul 1966.
[116] W. H. R. Equitz and T. M. Cover, "Successive Refinement of Information," *IEEE Transactions on Information Theory*, vol. 37, pp. 269-275, Mar 1991.
[117] A. Said and W. A. Pearlman, "Image Compression Using the Spatial Orientation Tree," International Sympossium on Circuits and Systems, pp. 279-282, 1993.
[118] D. M. Monro and G. J. Dickso, "Zerotree Coding of DCT Coefficients," Proceedings of IEEE International Conference on Image Processing, pp. 625-628, 1997.
[119] A. B. Watson, "DCTune: A technique for Visual Optimization of DCT Quantization Matrices for Individual Images," *Society for Information Display Digest of Technical Papers*, vol. XXIV, pp. 946-999, 1993.
[120] H. A. Peterson, "DCT basis function visibility in RGB space," in *Society for Information Display Digest of Technical Papers*, J. Morreale, Ed., Society for information Display, Playa del Rey, CA, 1992.
[121] H. A. Peterson, H. Peng, J. H. Morgan, and W. B. Pennebaker, "Quantization of color image components in the DCT domain," Human Vision, Visual Processing, and Digital Display, pp. 210-222, 1991.
[122] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," *SPIE Proc. Human Vision, Visual Processing, and Digital Display III*, pp. 365-374, 1992.
[123] R. Safranek and J. Johnston, "A Perceptually Tuned Sub-band Image Coder with Image Dependent Quantization and Post-quantization Data Compression," *Proceeding of IEEE ICASSP*, pp. 1945-1948, 1989.
[124] R. V. Cox, "The design of uniformly and nonuniformly spaced pseudo quadrature mirror filters," *IEEE Trans. ASSP*, vol. ASSP-34, pp. 1090-1096, Oct 1986.

[125] T. N. Cornsweet, *Visual Perception*. Orlando, FL, Academic Press, Inc., 1970.
[126] C. Chou and Y. Li, "A Perceptually Tuned Subband Image Coder Based on the Measure of Just-Noticeable-Distortion Profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 467-476, Dec 1995.
[127] I. Hontsch and L. J. Karam, "Adaptive Image Coding with Perceptual Distortion Control," *IEEE Transactions on Image Processing*, vol. 11, pp. 213-222, Mar 2002.
[128] I. Hontsch and L. J. Karam, "APIC: Adaptive Perceptual Image Coding Based on Subband Decomposition with Locally Adaptive Perceptual Weighting," Proceedings of IEEE International Conference on Image Processing, Santa Barbara, pp. 37-40, Oct 1997.
[129] I. Hontsch and L. J. Karam, "Locally Adaptive Perceptual Image Coding," *IEEE Transactions on Image Processing*, vol. 9, pp. 1472-1483, 2000.
[130] N. Jayant, "Signal Compression: Technology Targets and Research Directions," *IEEE J. Select. Areas Commun.*, vol. 10, pp. 314-323, Jun 1992.
[131] ISO/IEC, "Information Technology - JPEG 2000 Image Coding System - Part 2: Extension, 2000.," ISO/IEC 15444-2:2000, 2001.
[132] W. Zeng, S. Daly, and S. Lei, "Point-wise Extended Visual masking for JPEG2000 Image Compression," IEEE International Conference Image Protocol, 2000.
[133] S. Daly, "The Visible Difference Predictor: An Algorithm for the Assessment of Image Fidelity," in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA, MIT Press, pp. 179-206, 1993.
[134] A. P. Bradley, "A Wavelet Visible Difference Predictor," *IEEE Transactions on Image Processing*, vol. 8, pp. 717-730, May 1999.
[135] A. Cohen, I. Daubechies, and J. C. Feauveau, "Biorthogonal Bases of Compactly Supported Wavelets," *Commun. Pure Appl. Math.*, vol. 45, pp. 485-560, 1992.
[136] W. Lin, "Computational Models for Just-Noticeable-Difference," in *Digital Video Image Quality and Perceiptual Coding*, H. R. Wu and K. R. Rao, Eds., CRC Press, 2006.
[137] H. T. Tong and A. N. Venetsanopoulos, "A perceptual model for JPEG applications based on block classification, texture masking, and luminance masking," Proceedings of IEEE International Conference on Image Processing, pp. 3:428-432, 1998.
[138] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 Encoding With Perceptual Distortion Control," *IEEE Transactions on Image Processing*, vol. 15, July 2006.
[139] F. Kingdom and P. Whittle, "Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing," *Vision Research*, vol. 36, pp. 817-829, 1996.
[140] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image Coding Using Wavelet Transform," *IEEE Transactions on Image Processing*, vol. 1, April 1992.
[141] A. B. Watson, "The Cortex Transform: Rapid computation of simulated neural images," *Computer Vision, Graphics, and Image Processing*, vol. 39, pp. 311-327, 1987.
[142] C. H. Chou and K. C. Liu, "Colour Image Compression based on the Measure of Just Noticeable Colour Difference," *IET Image Processing*, vol. 2, pp. 304-322, 2008.

[143] B. Girod, "What's Wrong with Mean-Squared Error?," in *Digital Images and Human Vision*, A. B. Watson, Ed. MIT, MIT Press, pp. 207-220, 1993.
[144] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. IV-3313 - IV-3316, 13-17 May 2002.
[145] W. Sweldens, "The Lifting Scheme: A Custom-design Construction of Biorthogonal Wavelets," *Appl. Comput. Harmon. Anal.*, vol. 3, pp. 186-200, 1996.
[146] I. Daubechies and W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps," *J. Fourier Anal. Appl.*, vol. 4, pp. 245-267, 1998.
[147] ISO/IEC, "Lossy/lossless coding of bi-level images," ISO/IEC 14492-1, 2000.
[148] C. J. Van Den Branden Lambrecht, "Testing Digitial Video System and Quality Metrics based on Perceptual Models and Architecture, CH -1015." CH-1015 Lausanne, Switzerland, EPFL, May 1996.
[149] S. Winkler, "Perceptual Video Quality Metrics - A review," in *Digital Video Image Quality and Perceptual Coding*, H. R. Wu and K. R. Rao, Eds., CRC Press (ISBN: 0-8247-2777-0), pp. 155-179, Nov 2005.
[150] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multiscale Transform," *IEEE Transactions on Image Processing*, vol. 38, pp. 587-607, 1992.
[151] J. G. Daugman, "Self-Similar Oriented Wavelet Pyramids: Conjections About Neural Non-Orthogonality," in *Representation of Vision*, A. Gorea, Ed., Cambridge University Press, 1991.
[152] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures," Rec. ITU-R BT.500-11, 2002.
[153] L. Friedenberg, *Psychological Testing: Design, Analysis and Use.*, Allyn & Bacon, A Simon & Shcuster Company, 1995.
[154] R. M. Kaplan and D. P. Saccuzzo, *Psychological Testing: Principles Applications and Issues.*, 5th ed, Thomson Learning Inc., 2001.
[155] H. Coolican, *Research Methods and Statistics in Psychology*, Hodder & Stoughton, 2004.
[156] R. A. Johnson and G. K. Bhattacharyya, *Statistics: Principles and Methods*, 2nd ed, John Wiley & Sons, Inc., 1992.
[157] M. L. Berenson and D. M. Levine, *Basic Business Statistics: Concepts and Applications*, 4th ed, Prentice-Hall International Editions, 1989.
[158] M. Boliek, C. Christopoulos, and E. Majani, "JPEG2000 Part I Final Committee Draft Version 1.0," *Technical Rep, ISO/IEC JTC1/S9/WG1 N1646*, 2000.
[159] M. Rabbani and R. Joshi, "An Overview of the JPEG 2000 Still Image Compression Standard," *Signal Processing: Image Communication*, vol. 17, pp. 3-48, 2002.
[160] ISO/IEC, "JPEG 2000 Image Coding System -- Part I : Core Coding System, 2000," ISO/IEC 15444-1, 2000.
[161] C. Christopoulos, "JEPG 2000 Verification Model 8.5," *Technincal rep, ISO/IEC JTC1/SC9/WG1 N1878*, Sep 2000.
[162] M. Nadenau, "Integration of Human Color Vision Models into High Quality Image Compression." Lausanne, EPFL, 2000.

## Appendix A

Table A1: The Daubechies 9/7 wavelet filter set

| Filter Taps | Analysis Filter | | Synthesis Filter | |
|---|---|---|---|---|
| | Low Pass, $h$ | High Pass, $g$ | Low Pass, $\overline{h}$ | High Pass, $\overline{g}$ |
| 0 | 0.602949 | -0.557543 | 1.115086 | -1.205898 |
| ±1 | 0.266864 | 0.295636 | 0.591272 | 0.533728 |
| ±2 | -0.078223 | 0.028772 | -0.057544 | 0.156446 |
| ±3 | -0.016864 | -0.045636 | -0.0921272 | -0.033728 |
| ±4 | 0.026749 | 0 | 0 | -0.053498 |

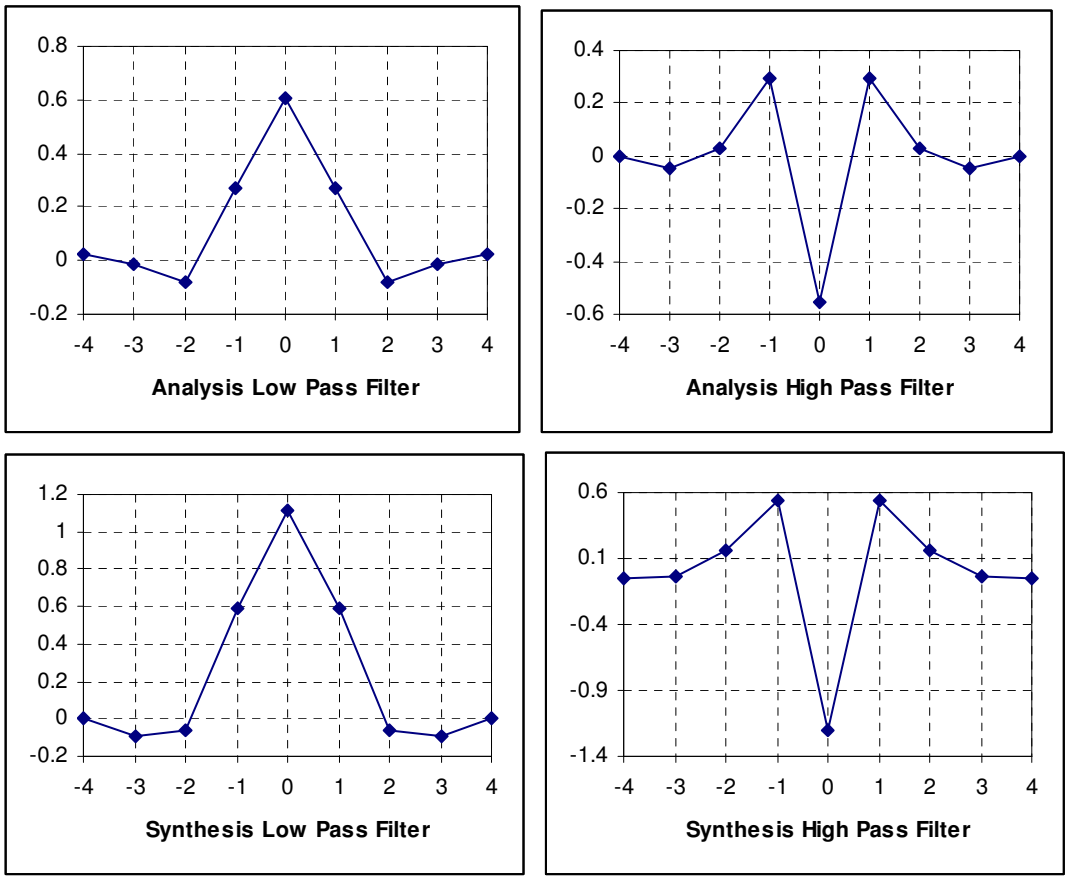



Figure A1: Profiles of Daubechies 9/7 filter taps.

## Appendix B

Figure B1 Original uncompressed image of *barbara*. The size of this image is reduced to 60% to fit within A4 size paper.

Figure B2 Original uncompressed image of *barbara2*. The size of this image is reduced to 60% to fit within A4 size paper.


Figure B3 Original uncompressed image of *boats*. The size of this image is reduced to 60% to fit within A4 size paper.


Figure B4 Original uncompressed image of *bikes*. The size of this image is reduced to 56% to fit within A4 size paper.


Figure B5 Original uncompressed image of *building2*. The size of this image is reduced to 68% to fit within A4 size paper.

## Appendix C: Subjective Assessment Questionnaire for Chapter 4

Digital Image Quality Analysis Form for Digital Colour Images

Venue: Room 87-03-06, RMIT City Campus

Important Information

| |
|---|
| Thank you for your participation.<br>To participate, you must be at least 18 years old.<br>You may withdraw at any time without completing it.<br>Data and methods will be fully published. However, no personal identifiable data and no data identifying an individual will be published. |

Participant Details

| | | | |
|---|---|---|---|
| Name: | | Sex: | Female / Male |
| Do you normally wear glasses? Yes / No | | | |
| Are you colour blind? Yes / No | | | |

Official Use:

| | |
|---|---|
| Serial No: | |

**Part 1: AB/BA sequence** Date:_____ Time: _____

**Context**

You have to spend $2400 on the purchase of 24 pictures either as a gift for someone special or for your personal collection. The pictures are displayed on the left and right. Your task is as follows:
Tick on the box showing your preferred picture ( i.e., " L" for Left image, "R" for the Right image).

| Image Number | L | R |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |

| Image Number | L | R |
|---|---|---|
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |

Legend: L – Left, R - Right

**Part 2: AC/CA sequence** Date:_____ Time: _____

**Context**

You have to spend $2400 on the purchase of 24 pictures either as a gift for someone special or for your personal collection. The pictures are displayed on the left and right. Your task is as follows:
Tick on the box showing your preferred picture ( i.e., " L" for Left image, "R" for the Right image).

| Image Number | L | R |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |
| 11 | | |
| 12 | | |

| Image Number | L | R |
|---|---|---|
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |

Legend: L – Left, R - Right

_____ End of Test _____

## Appendix D: Subjective Assessment Questionnaire for Chapter 5 (Implementation I)

Digital Image Quality Analysis for Digital Colour Images

Venue: Room 1103, Building 75 (Strip), Clayton Campus, Monash University

Important Information

| |
|---|
| Thank you for your participation.<br>To participate, you must be at least 18 years old.<br>You may withdraw at any time without completing it.<br>Data and methods will be fully published. However, no personal identifiable data and no data identifying an individual will be published. |

Participant Details

| | | | |
|---|---|---|---|
| Name: | | Sex: | Female / Male |
| Do you normally wear glasses? Yes / No | | | |
| Are you colour blind? Yes / No | | | |

Official Use:

| | |
|---|---|
| Serial No: | |

**Context**

You have to spend $3000 on the purchase of 30 pictures either as a gift for someone special or for your personal collection. The pictures are displayed on the left and right. Your task is to choose the picture you prefer.

**Task:** Tick on the box indicating your preferred choice.

**Part 1: AB/BA sequence** Date: _____ Time: _____

| Image Number | Left | Right |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |

| Image Number | Left | Right |
|---|---|---|
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |

| Image Number | Left | Right |
|---|---|---|
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |
| 25 | | |
| 26 | | |
| 27 | | |
| 28 | | |
| 29 | | |
| 30 | | |

**Part 2: AC/CA sequence** Date: _____ Time: _____

| Image Number | Left | Right | Image Number | Left | Right | Image Number | Left | Right |
|---|---|---|---|---|---|---|---|---|
| 1 | | | 11 | | | 21 | | |
| 2 | | | 12 | | | 22 | | |
| 3 | | | 13 | | | 23 | | |
| 4 | | | 14 | | | 24 | | |
| 5 | | | 15 | | | 25 | | |
| 6 | | | 16 | | | 26 | | |
| 7 | | | 17 | | | 27 | | |
| 8 | | | 18 | | | 28 | | |
| 9 | | | 19 | | | 29 | | |
| 10 | | | 20 | | | 30 | | |

**Part 3: AD/DA sequence** Date: _____ Time: _____

| Image Number | Left | Right | Image Number | Left | Right | Image Number | Left | Right |
|---|---|---|---|---|---|---|---|---|
| 1 | | | 11 | | | 21 | | |
| 2 | | | 12 | | | 22 | | |
| 3 | | | 13 | | | 23 | | |
| 4 | | | 14 | | | 24 | | |
| 5 | | | 15 | | | 25 | | |
| 6 | | | 16 | | | 26 | | |
| 7 | | | 17 | | | 27 | | |
| 8 | | | 18 | | | 28 | | |
| 9 | | | 19 | | | 29 | | |
| 10 | | | 20 | | | 30 | | |

_____ End of Test _____

## Appendix E: Subjective Assessment Questionnaire for Chapter 5 (Implementation II)

Digital Image Quality Analysis for Digital Colour Images

Venue: Room 87-03-06, RMIT City Campus

Important Information

| |
|---|
| Thank you for your participation.<br>To participate, you must be at least 18 years old.<br>You may withdraw at any time without completing it.<br>Data and methods will be fully published. However, no personal identifiable data and no data identifying an individual will be published. |

Participant Details

| | | | |
|---|---|---|---|
| Name: | | Sex: | Female / Male |
| Do you normally wear glasses? Yes / No | | | |
| Are you colour blind? Yes / No | | | |

Official Use:

| | |
|---|---|
| Serial No: | |

**Part 1: AB/BA sequence** Date: _____ Time: _____

**Context**

You have to spend $3000 on the purchase of 30 pictures either as a gift for someone special or for your personal collection. The pictures are displayed on the left and right. Your task is to choose the picture you prefer.

**Task:** Tick on the box indicating 'N' if both images are of the similar quality. Otherwise tick on the box indicating your preferred choice (either Left or Right).

| Image Number | N | Left | Right |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 13 | | | |
| 14 | | | |
| 15 | | | |

| Image Number | N | Left | Right |
|---|---|---|---|
| 16 | | | |
| 17 | | | |
| 18 | | | |
| 19 | | | |
| 20 | | | |
| 21 | | | |
| 22 | | | |
| 23 | | | |
| 24 | | | |
| 25 | | | |
| 26 | | | |
| 27 | | | |
| 28 | | | |
| 29 | | | |
| 30 | | | |

**Part 2: AC/CA sequence** Date: _____ Time: _____

**Context**

You have to spend $3000 on the purchase of 30 pictures either as a gift for someone special or for your personal collections. The pictures are displayed on the left and right. Your task is to choose the picture you prefer.

**Task:** Tick on the box indicating your preferred choice.

| Image Number | Left | Right |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |

| Image Number | Left | Right |
|---|---|---|
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |

| Image Number | Left | Right |
|---|---|---|
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |
| 25 | | |
| 26 | | |
| 27 | | |
| 28 | | |
| 29 | | |
| 30 | | |

**Part 3: AD/DA sequence** Date: _____ Time: _____

**Context**

You have to spend $3000 on the purchase of 30 pictures either as a gift for someone special or for your personal collections. The pictures are displayed on the left and right. Your task is to choose the picture you prefer.

**Task:** Tick on the box indicating your preferred choice.

| Image Number | Left | Right |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |
| 8 | | |
| 9 | | |
| 10 | | |

| Image Number | Left | Right |
|---|---|---|
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |
| 17 | | |
| 18 | | |
| 19 | | |
| 20 | | |

| Image Number | Left | Right |
|---|---|---|
| 21 | | |
| 22 | | |
| 23 | | |
| 24 | | |
| 25 | | |
| 26 | | |
| 27 | | |
| 28 | | |
| 29 | | |
| 30 | | |

_____ End of Test _____

## Appendix F:
## MSE for JPEG2000-PCDM, JPEG2000-MSE and JPEG2000-CVIS

| | | Average PSNR (db) | | |
|---|---|---|---|---|
| Bit rate (bpp) | Images | JPEG2000-PCDM | JPEG2000-MSE | JPEG2000-CVIS |
| 1.0 | goldhill | 38.43 | 38.52 | 38.47 |
| | Sail | 37.77 | 37.74 | 39.12 |
| | Pepper | 42.26 | 42.47 | 42.42 |
| | Lena | 39.07 | 39.15 | 38.88 |
| | Tulip | 39.65 | 39.93 | 39.95 |
| | Paintedhouse | 39.31 | 39.31 | 39.76 |
| 0.5 | goldhill | 36.63 | 36.80 | 36.78 |
| | Sail | 35.11 | 34.97 | 36.20 |
| | Pepper | 39.70 | 39.78 | 39.74 |
| | Lena | 37.28 | 37.30 | 37.08 |
| | Tulip | 35.72 | 36.02 | 36.22 |
| | Paintedhouse | 36.77 | 36.64 | 37.22 |
| 0.25 | goldhill | 35.17 | 35.19 | 35.51 |
| | Sail | 32.57 | 32.53 | 33.90 |
| | Pepper | 36.51 | 36.67 | 36.58 |
| | Lena | 35.29 | 35.36 | 35.30 |
| | Tulip | 32.23 | 32.72 | 32.99 |
| | Paintedhouse | 34.77 | 34.85 | 35.26 |
| 0.125 | goldhill | 33.58 | 33.88 | 34.18 |
| | Sail | 29.98 | 30.48 | 31.67 |
| | Pepper | 32.90 | 33.27 | 33.19 |
| | Lena | 33.19 | 33.29 | 33.41 |
| | Tulip | 28.87 | 29.86 | 30.07 |
| | Paintedhouse | 33.45 | 33.37 | 34.04 |

The average PSNR is computed based on the expressions below,

$$MSE(c) = \frac{\sum_{i \in N} (\hat{x}_c[i] - x_c[i])^2}{N}$$

$$PSNR(c) = 10 \cdot log_{10}\left(\frac{255}{MSE(c)}\right)$$

$$Average\ PSNR = \frac{PSNR(Y) + PSNR(C_b) + PSNR(C_r)}{3}$$

Where $\hat{x}_c[i]$ and $x_c[i]$ are the sample data of the compressed and original images of $N$ samples, and $c \in \{Y, C_b, C_r\}$ is the colour component.

## Appendix G:
## MSE for JPEG2000-PCDM-PPF, JPEG2000-PCDM, JPEG2000-MSE, and JPEG2000-CVIS

| | | Average PSNR (db) | | | |
|---|---|---|---|---|---|
| Image | Bit Rate (bpp) | JPEG2000-PCDM-PPF | JPEG2000-PCDM | JPEG2000-MSE | JPEG2000-CVIS |
| goldhill | 1.0 | 37.06 | 38.49 | 38.58 | 38.54 |
| | 0.5 | 35.69 | 36.69 | 36.86 | 36.86 |
| | 0.25 | 34.51 | 35.25 | 35.27 | 35.59 |
| sail | 1.0 | 36.18 | 37.72 | 37.74 | 39.11 |
| | 0.5 | 34.19 | 35.05 | 34.97 | 36.21 |
| | 0.25 | 32.05 | 32.51 | 32.54 | 33.91 |
| pepper | 1.0 | 39.69 | 42.23 | 42.45 | 42.38 |
| | 0.5 | 38.02 | 39.59 | 39.68 | 39.65 |
| | 0.25 | 35.52 | 36.31 | 36.48 | 36.41 |
| lena | 1.0 | 37.75 | 39.10 | 39.17 | 38.92 |
| | 0.5 | 36.39 | 37.31 | 37.34 | 37.13 |
| | 0.25 | 34.73 | 35.34 | 35.38 | 35.34 |
| tulip | 1.0 | 37.88 | 39.60 | 39.89 | 39.92 |
| | 0.5 | 34.93 | 35.72 | 35.99 | 36.19 |
| | 0.25 | 31.82 | 32.19 | 32.70 | 32.96 |
| zelda | 1.0 | 40.75 | 42.84 | 43.02 | 42.85 |
| | 0.5 | 39.83 | 41.58 | 41.74 | 41.47 |
| | 0.25 | 38.71 | 40.07 | 40.10 | 40.07 |
| bikes | 1.0 | 36.08 | 37.66 | 37.63 | 39.00 |
| | 0.5 | 33.66 | 34.51 | 34.61 | 36.06 |
| | 0.25 | 31.88 | 32.39 | 32.27 | 33.76 |
| building2 | 1.0 | 32.84 | 33.68 | 33.45 | 34.59 |
| | 0.5 | 30.86 | 31.28 | 31.23 | 32.34 |
| | 0.25 | 29.20 | 29.44 | 29.76 | 30.65 |
| lighthouse2 | 1.0 | 39.18 | 41.82 | 41.88 | 42.42 |
| | 0.5 | 37.29 | 38.96 | 38.94 | 40.05 |
| | 0.25 | 35.51 | 36.53 | 36.80 | 38.07 |
| stream | 1.0 | 35.63 | 37.04 | 37.10 | 38.46 |
| | 0.5 | 34.10 | 35.04 | 35.21 | 36.46 |
| | 0.25 | 33.28 | 34.03 | 33.98 | 35.20 |

The average PSNR is computed based on the expressions below,

$$MSE(c) = \frac{\sum_{i \in N} (\hat{x}_c[i] - x_c[i])^2}{N}$$

$$PSNR(c) = 10 \cdot log_{10}\left(\frac{255}{MSE(c)}\right)$$

$$Average\ PSNR = \frac{PSNR(Y) + PSNR(C_b) + PSNR(C_r)}{3}$$

where $\hat{x}_c[i]$ and $x_c[i]$ are the sample data of the compressed and original images of $N$ samples, and $c \in \{Y, C_b, C_r\}$ is the colour component.

## Appendix H: Bandlimited Contrast by Peli

The image is filtered by a pyramidal structure of 1-octave wide bandwidth bandpass filters centred at different levels that are 1-octave apart. At every level, a local average luminance, $l_i(x, y)$, containing all energy at bands lower than the current band, is computed. The bandlimited contrast is obtained by dividing the bandpass-filtered image point-by-point (i.e., $a_i(x, y)$) by the corresponding local average luminance.

We consider an image $f(x, y)$ that can be represented in the frequency domain as,

$$F(u,v) = F(r,\theta) = L_0(r,\theta) + \sum_{i=1}^{n-1} A_i(r,\theta) + K_n(r,\theta) \tag{h1}$$

where $u$ and $v$ are the horizontal and vertical spatial frequency coordinates, $r = \sqrt{u^2 + v^2}$ and $\theta = tan^{-1}\left(u/v\right)\theta$ are the polar spatial coordinates, $L_0(r,\theta)$ and $K_n(r,\theta)$ are the low and high residual terms. $A_i(r,\theta)$ can be obtained by multiplying the fourier transform of image $f(x, y)$ with a cosine log bandpass filter in equation (h2) which is of 1-octave wide bandwidth centred at frequency 1-octave apart at different levels. The cosine log filter is as follows;

$$G_i(r) = \frac{1}{2}\left[1 + cos(\pi\, log_2\, r - \pi i)\right] \tag{h2}$$

The filtered image is transformed back to space domain via inverse fourier transform. The image in the space domain can be expressed as;

$$f(x, y) = l_0(x, y) + \sum_{i=1}^{n-1} a_i(x, y) + h_n(x, y) \tag{h3}$$

The bandlimited contrast, $C_i^{blc}(x, y)$, is computed as;

$$C_i^{blc}(x, y) = \frac{a_i(x, y)}{l_i(x, y)} = \frac{a_i(x, y)}{l_0(x, y) + \sum_{k=1}^{i-1} a_k(x, y)} \tag{h4}$$

## Appendix I: The Cortex Transform

The cortex transform is modelled with separate class of filters: the dom and fan filters. The dom filters are used to model the spatial frequency channels while the fan filters models the orientation channels of the HVS.

The cortex filter is defined as.

$$cortex_{k,i}(u,v) = d_k(u,v) \cdot h_i(u.v) \tag{i1}$$

Where $d_k(u,v)$ and $h_\alpha(u.v)$ are the dom filter at $k^{\text{th}}$ scale and $j^{\text{th}}$ fan filter for orientation band at $\frac{j \cdot \pi}{K}$ radians (or $\frac{180 \cdot j}{K}$ degrees) with $K$ being the total of number of fan filters at each scale.

The dom filter $d_k(u,v)$ is computed as the difference of mesa filters as follows,

$$d_k(u,v) = m_k(u,v) - m_{k+1}(u,v) \tag{i2}$$

where $m_k(u,v)$ and $m_{k+1}(u,v)$ are the mesa filters at scale k and k+1, respectively. The kth scale mesa filter is defined as,

$$m_k(u,v) = m(s^k u, s^k v) \tag{i3}$$

where $m(u,v)$ is defined as the convolution of a Gaussian function with a cylinder of radius $f_0$ . At every successive resolution, the image is reduced by a factor of *s*.

$$m(u,v) = \left(\frac{\lambda}{f_0}\right)^2 e^{-\pi\left(\frac{\lambda r}{f_0}\right)^2} * \prod\left(\frac{r}{2f_0}\right) \tag{i4}$$

where $r = \sqrt{u^2 + v^2}$ and $\prod\left(\frac{r}{2f_0}\right)$ is a rectangular pulse with unity height centred at the origin. $f_0$ is the corner frequency at which the Gaussian falls off to 0.5 of its height. $\lambda$ is the parameter defining the sharpness of the response.

For the fan filter $h_j(u.v)$, it is computed as,

$$h_j(u.v) = b_{j\cdot w}(u,v)\left[1 - b_{(j+1)\cdot w}(u,v)\right] + b_{(j+K+1)\cdot w}(u,v)\left[1 - b_{(j+K+1)\cdot w}(u,v)\right] \qquad \text{(i5)}$$

Where $w = \frac{\pi}{K}$ is the orientation bandwidth for $K$ fan filters, and the index, $j$, to the orientation band is $j \in \{1,2,...K-1\}$. The bisection filter, $b_\beta(u,v)$, is defined as the cumulative Gaussian as follows.

$$b_\beta(u,v) = g(w(v\cos\beta - u\sin\beta)) \qquad \text{(i6)}$$

where

$$g(wv) = \int_{-\infty}^{v} w \cdot e^{-\pi w^2 r^2} dr \qquad \text{(i7)}$$

The β in radians is the angle of rotation for the orientation band. For example, $\beta = j \cdot w$ refers to the j$^{th}$ orientation band which corresponds to $\frac{j \cdot \pi}{K}$ radians (or $\frac{180 \cdot j}{K}$ degrees). The 3$^{rd}$ orientation band of a 4-orrientation band filter corresponds to the 135 degrees band.

For a two dimensional image, the filtered images are computed by multiplying the discrete Fourier transform of the input image by each filter defined in equation (i1), followed by applying the inverse discrete Fourier transform. To reconstruct the image, discrete Fourier transform is applied to each of the filtered images at each layer, the

DFT of the layer are embedded in a null DFT to the size of the original image, followed by applying the inverse discrete Fourier transform.

## Appendix J

This CD contains test images that were used in the subjective evaluations for

(1) PCDM based coder introduced in chapter 4,
(2) PCDM-PPF based algorithm as introduced in implementation I of chapter 5, and
(3) PPF algorithm as introduced in implementation II of chapter 5.

All images are in PPM format. The images can be viewed by a PPM compatible image viewer.