HUMAN VISION MODELS OF PERCEPTUAL IMAGE DISTORTIONS

By

OSAYAMEN OSARETIN IMADE

Bachelor of Science in Electrical Engineering
Oklahoma State University
Stillwater, Oklahoma, USA
2008

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2010

HUMAN VISION MODELS OF PERCEPTUAL IMAGE DISTORTIONS

Thesis Approved:

Dr. Damon M. Chandler

Thesis Advisor

Dr. Nazanin Rahnavard

Dr. Daqing Piao

Dr. Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

Firstly, I would like to thank my heavenly father for giving me the opportunity to be at Oklahoma State University and completing my degrees in good health. I would like to thank my adviser Dr. Damon Chandler for his immense support and advice throughout my Masters and research. His devotion to understanding the human vison has both challenged and driven me to increase my understanding.

I would also like to thank the my dear parents Mr. and Mrs. Imade and my siblings Osatohanmwen Imade, Osadebamwen Imade and Ikponmwosa Imade for their unconditional love and support.

I would like to thank my fellow Computational Perception and Image Quality (CPIQ) lab members for sitting through my numerous experiments and their humorous nature. Lastly, I would like to thank my friends for their patience and support throughout my Masters at Oklahoma State University.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Thesis Overview and Contributions

It can be irrefutably said that digital media in today's society is both widely available and easily distributable. Our ability to capture and store events digitally has led to significant changes in our lives. By the year 2009, unique user picture uploads to the popular social networking site Facebook, had exceeded 15 billion at a rate of 323 photographs per second [2]. According to [3], over 24 hours of video per minute is uploaded to the video sharing website: YouTube. In addition, the numerous devices used to capture, transmit and store digital media has evoked several standards and driven the demand for algorithms that can greatly reduce bandwidth. The potential benefits include increased efficiency, greater compression ratios and improved information assurance.

Driven by society's current appetite for digital content which include high-definition digital pictures and video, 3D video, volumes of data traffic; research has been led to exploit various methods to meet these requirements. For example, in the case of increasing storage demands, two schools of thought exist in meeting this challenge. Firstly, researchers believe in simply increasing the capacity of the storage media as consumer needs rise [4]. The second develops algorithms that can further reduce compression ratios to fit more data in the same storage capacity. In favor of the latter argument, designing an algorithm that will push the limits of current compression techniques requires exploitation of the human visual system.

The human visual system though powerful is not infallible and is both subjective

and qualitative. Images viewed by a human observer is subject to interpretation, perception and recognition [5] and as such can be exploited. Compression can be *lossy* or *lossless*, in the case of the latter higher compression ratios are hard to achieve as compared to a *lossy* model. In the area of information assurance i.e. digital watermarking, a watermark could be visible or invisible. An understanding of how sensitivities of the *human visual system* (HVS) combine to provide perceptional discrepancies across various regions of an image is required. This knowledge will enable development of algorithms that increase not just ratios of compression algorithms but the bit-rate of a digital watermarking system as well.

This thesis presents an image-adaptive contrast and entropy based model as well as a proven hypothesis on context degree influence; to predict the regions whereby distortions in an image are regarded visible by a human observer and how much influence on a distortions visibility is exerted by the context of the image respectively. With the application of two psychophysical experiments which were performed on natural images, the HVS is incorporated into the proposed algorithm to detect the regions of visible distortions. The psychophysical experiments involved the detection of stimuli provided via wavelet-based coefficient distortions and gabor (Gaussian-windowed-sinusoid) targets. Specifically, the algorithm (1) determines the image *activity* via the statistical properties of the image, (2) defines an *activity* scaling parameter and (3) compares the *activity* and distortions to predict the regions of visible distortions.

The contributions of this thesis are summarized as follows:

1. An analysis of the detectability of wavelet subband coefficient-manipulation distortions masked by natural images (contrast and entropy masking).

2. A psychophysical experiment based derivation relating average contrast, average entropy and average luminance to the activity scaling parameter.

3. An algorithm for predicting the regions of visible distortions.

4. Applications of (3) to a compression scheme.

5. An analysis of the detectability of wavelet subband coefficient-manipulation distortions and detectability of gabor targets via context-masking.

1.2 Thesis Organization

This thesis is organized as follows: Chapter 2 reviews the HVS, image compression and digital watermarking. The results of a psychophysical experiment is presented in Chapter 3 whose aim was to study the effects of contrast and entropy masking. Chapter 4 provides an algorithm which models the human observers' detectability of visible distortions in an image and also details an implementation in a compression scheme. Chapter 5 presents the result of a psychophysical experiment aimed at investigating the influence of context-masking on the detectability of gabor targets presented against various masking textures. Chapter 6 provides general conclusions and future work.

CHAPTER 2

BACKGROUND

2.1 Introduction

What happens when we open our eyelids and let light into our eyes is far from a simple, linear and uniform response. It is widely known that the human visual system exhibits various characteristics that shape our perception to the world around us. The main driving force for this thesis, is to understand, what do we see when we gaze upon an image? How do we see visual distortions? And why do we see certain distortions and lose sight of others? Is it the fact that an image is appealing that makes us scrutinize it better for distortions? Could it be that the image being non-uniform itself interacts with our non-uniform visual system causing artifacts that should be visible to disappear? To demonstrate the fallibility of the HVS, consider the image in Figure 2.1 below.

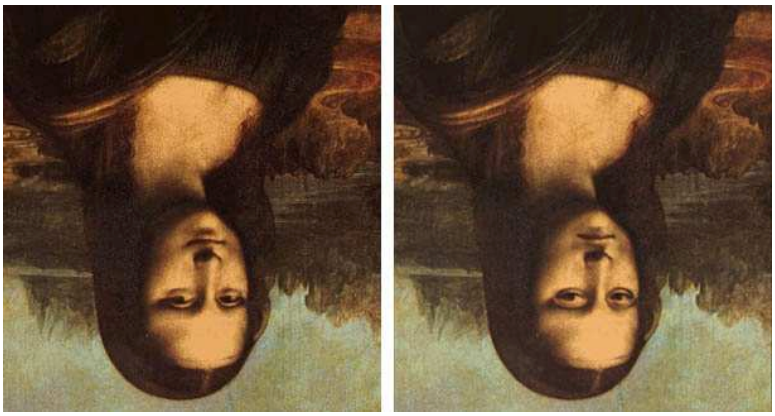

Figure 2.1: An image of the original *mona lisa* on the right and a distorted version on the left

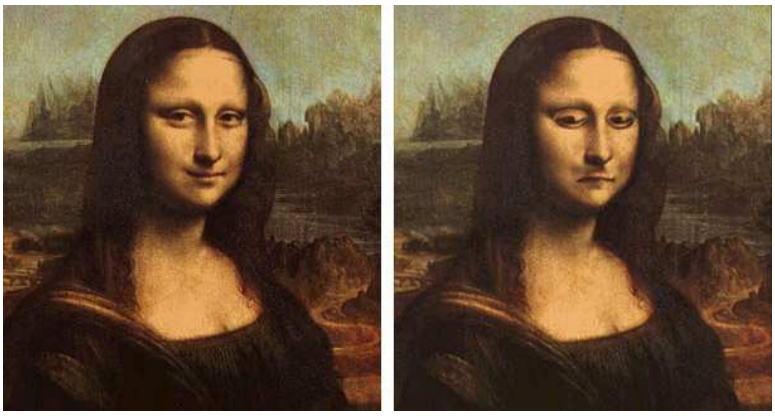

Figure 2.2: An image of the original *mona lisa* on the right and a distorted version on the left, now rotated to reveal the distortion

If you look at the image of the *mona lisa* in Figure 2.1, you may fail to notice the distortion in the face of the left image. Further attention to detail (Figure 2.2) reveals that the structure that compose a regular face such as the eyes and lips are inverted [6]. This demonstrates that the human visual system may develop certain criteria in which we view images and objects in our surroundings and therefore can fail to detect obvious distortions even when placed in plain sight. This also indicates that over time cells in our cortical areas of vision have been tuned to certain orientations that facilitate recognition. Thus it is agreeable to tune image applications towards the human visual system's characteristics. This chapter provides an overview to further understand the aspects of the human visual system that apply to image compression and digital watermarking; namely, contrast and entropy perception and context masking.

2.2 Human Visual System

2.2.1 Structural Description

This section provides a description of the visual areas that process input signals and their hierarchy. Figure 2.3 illustrates the block diagram of the feed forward con-

nections between visual processing areas. The eye, lateral geniculate nucleus (LGN), striate cortex (V1), and higher-level areas such as V2, V3, V4, medial temporal cortex (V5, MT), inferior temporal cortex (IT), and posterior parietal cortex.

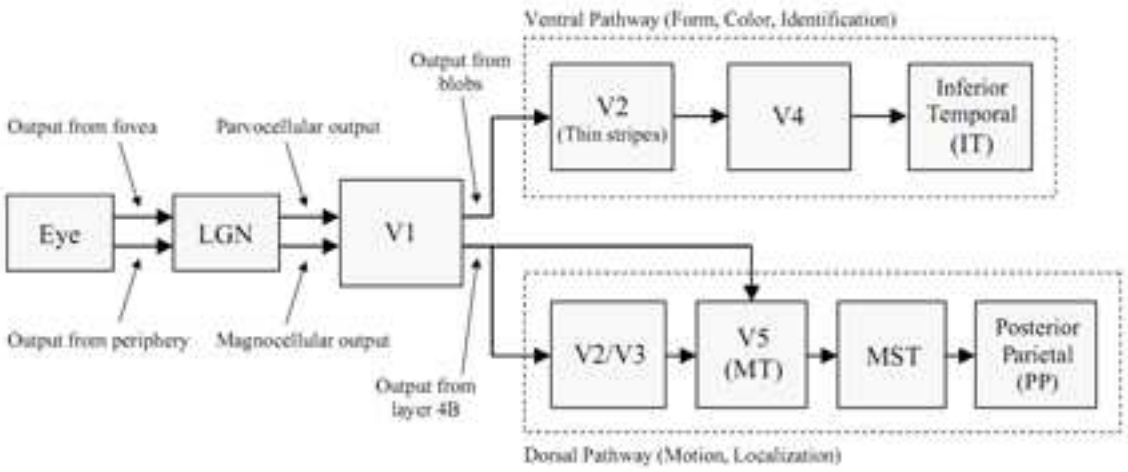

Figure 2.3: Diagram of hierarchy of HVS: Input signals from the eye is relayed via the lateral geniculate nucleus (LGN) to the striate cortex (V1) and then to the inferotemporal cortex (IT) via the V2 and V4 cortex for object recognition and to the dorsal visual pathway(MT, MST) for localization and motion processing.

Eye

The human eye figure 2.4 is dedicated to the focus and capture light and subsequently converting sensory information into neural signals. Light enters the cornea, then through the aqueous humor, then through the lens into the vitreous humor, and finally projected onto the photo receptors (rods and cones) located at the back of the retina. Focus is achieved via accommodation of the lens by the ciliary muscles. The light rays focus onto the fovea, the region of the retina containing the greatest density of cones and thus the highest acuity for spatial and color vision [7][8].

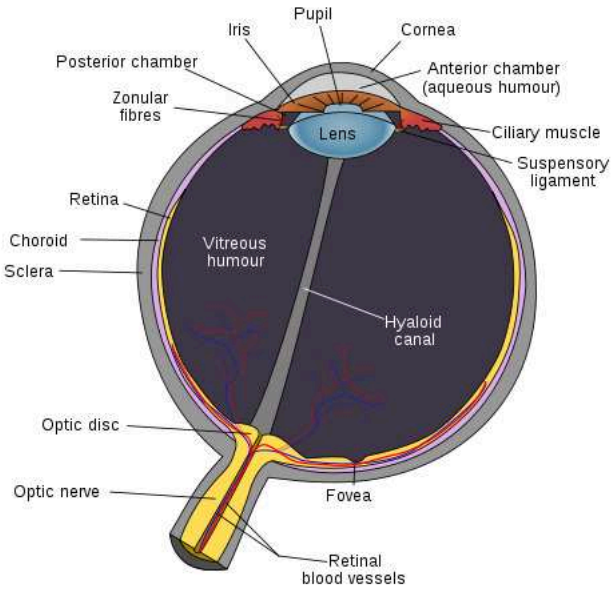

Figure 2.4: Diagram of the human eye depicting the main structure that capture light and convert it to optical neural signals.

Retina

The retina is composed of five layers of neurons: *photoreceptors*, *horizontal cells*, *bipolar cells*, *amacrine cells*, and *ganglion cells* [9]. The photoreceptors consists of the *rods* responsible for vision in low-light (scotopic) conditions and *cones* which function under normal (photopic) lighting and are responsible for color vision. The horizontal, bipolar, and amacrine cells comprise the *plexiform layer* of the retina: Bipolar cells receive input from the photoreceptors and provide output to the ganglion cells; horizontal and amacrine cells provide lateral connections between photoreceptors, and between bipolar cells and ganglion cells, respectively [10].

Retinal ganglion cells constitute the first stage of complex spatial processing. Ganglion cells have also been classified according to the layer in LGN to which their outputs are directed: *M cells*, which receive input from both rods and cones, demonstrate high sensitivity to contrast, but reduced sensitivity to color. *P cells*, which

receive input from only cones, demonstrate high sensitivity to color, but reduced sensitivity to contrast.

LGN

The *lateral geniculate nuclei* (LGN) receives outputs of retinal ganglion cells through the optic nerve. The LGN is a portion of the thalamus composed of six laminar sheets of neurons [7][10]. Two of LGN's six layers, the *magnocellular* layers, contain neurons with large receptive fields that are responsible for coding achromatic contrast; these neurons receive input from the M ganglion cells. Neurons in the other four layers, the *parvocellular* layers, contain smaller receptive fields and accordingly demonstrate a high degree of spatial resolution; these parvocellular neurons, which receive input from the P ganglion cells, are highly sensitive to color, but show reduced contrast sensitivity and lower temporal resolution than magnocellular neurons.

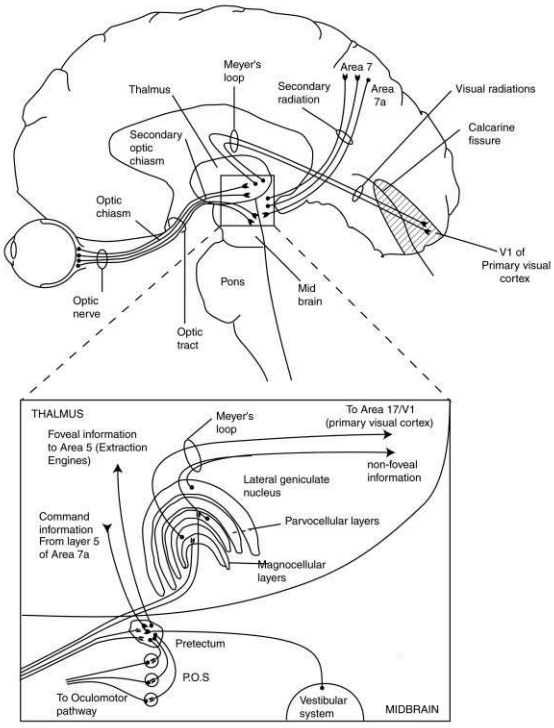

Figure 2.5: Diagram of the anatomical structure of the HVS.

Primary Visual Cortex

Output from LGN is directed to the visual areas of cortex (*visual cortex*), with the majority of the projections synapsing in *primary visual cortex* (V1; also called *striate cortex* or *area 17*). V1, which is located in the posterior region of the occipital lobe, is the largest area of visual cortex that contains approximately 200 million neurons, more than 100 times the amount found in LGN [7].

The cells contained in the V1 exhibit a higher level of information transformation than the LGN as they are both orientation selective and direction selective. The cells are further subdivided into *simple cells* (edge and bar detectors with specific widths and orientations), *complex cells* (edge and bar detectors with preference for specific direction of motion), and *Hypercomplex cells* [11].

Ventral and Dorsal Pathways

The other cortical areas that are associated with the ventral and dorsal pathways are the V2, V3, V4 and V5 regions. Output from the thin stripes of V2 are sent to V4 the majority of whose own output is directed at *inferior temporal cortex* (IT); this path, V1 (blobs) $\rightarrow$ V2 (thin stripes) $\rightarrow$ V4 $\rightarrow$ IT, is often called the "what" system as it is believed to be responsible for form perception and recognition. Output from the thick stripes of V2 project to *medial temporal cortex* (V5, MT; responsible for processing stereo and motion), which in turn projects to *medial superior temporal cortex* (MST; responsible for visual tracking), and then to *posterior parietal cortex* (PP); this path, V1 (layer 4B) $\rightarrow$ V2 (thick stripes) $\rightarrow$ V5 $\rightarrow$ MST $\rightarrow$ PP, is known as the "where" system which is believed to handle localization.

2.2.2 Psychophysical Description

The HVS cannot be completely described by its structure alone, the summation of the whole system provides non-uniform responses that require psychophysical analysis.

Experimental characterization of the HVS include but are not limited to finding *contrast detection thresholds*, *region of interest detection* and *object recognition*. Here we examine the contrast sensitivity of the HVS as well as detail the contrast measures used primarily in the psychophysical experiments conducted in Chapter 3 and Chapter 5.

Contrast Sensitivity Function

The definition of contrast according to Merriam-Webster's Dictionary is the degree of difference between the lightest and darkest parts of a picture. Although, this definition is standard, several quantitative measures exist to define this property. Psychophysical experiments typically employ a *simple contrast*, *Weber contrast*, *Michelson contrast*, and *root-mean-squared contrast* (*RMS contrast*) [12]. In the psychophysical experiments performed the *RMS contrast* and is commonly employed for non-periodic targets (noise, textures, images). Note that the luminance values used in all of these definitions of contrast are defined in standard CIE units of candelas/square meter ($\mathrm{cd/m^2}$). RMS contrastand is defined as:

$$C_{RMS} \equiv \frac{1}{\mu_L}\left(\frac{1}{N}\sum_{i=0}^{N}\left[L_i-\mu_L\right]^2\right)^{1/2} = \frac{\sigma_L}{\mu_L} \tag{2.1}$$

where μ_L denotes the mean luminance, σ_L denotes the standard deviation of the target's luminance, L_i denotes the target's luminance at spatial location i, and N denotes the total number of spatial locations. For cases in which the mean luminance of the target differs from the mean luminance of the background, RMS contrast is given by

$$C_{RMS} \equiv \frac{1}{\mu_{L_{bkgnd}}}\left(\frac{1}{N}\sum_{i=0}^{N}\left[L_i-\mu_{L_{target}}\right]^2\right)^{1/2} = \frac{\sigma_{L_{target}}}{\mu_{L_{bkgnd}}} \tag{2.2}$$

The *contrast sensitivity function* (CSF) is a model of the HVS contrast detection thresholds for sine-wave gratings of various spatial frequencies [1]. The *contrast sensitivity* was defined to be the inverse of the contrast which was measured just as the

sine-wave grating was visible to the human observer. The CSF is pictured in Figure 2.6 It was found that humans have peak sensitivity near 4-6 cycles/degree. This experiment has been the basis for several applications where a HVS model is applied, including compression algorithms and digital watermarking algorithms [11,12,13,14].

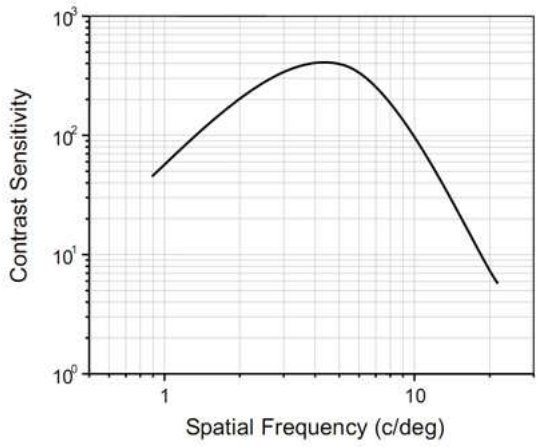

Figure 2.6: Human contrast sensitivity function (CSF; inverse detection thresholds) for targets consisting of sine-wave gratings (after Ref. [1]).

Visual Masking

It has been found that contrast thresholds can vary based on an effect known as Visual masking. Generally, the presence of a masking signal (*mask* or *masker*) may reduce or facilitate a subject's ability to detect a given test signal (*target*). Currently, the visual masking phenomenon can be divided into three types:

1. *Noise masking*, which attributes the increase in detection thresholds to the corruptive effects of the masker on internal decision variables [13];

2. *Contrast masking*, which attributes threshold elevations to contrast gain control [14] (discussed next);

3. *Entropy masking*, which is imposed solely by an observer's unfamiliarity with the masker [15].

2.3 Previous Work

2.3.1 Applications

Watermarking is the practice of imperceptibly embedding information about a work into the work itself. [16]. Early incorporation of HVS sensitivities include works by G. S. Lewis et al. [17] where quantization step sizes for subband wavelet coefficients were calculated as a function of local luminance and texture. G. S. Lewis et al. retrieved sensitivity properties of the HVS to lower frequency wavelet coefficients due to their higher energy content and the inability of detection of quantization error in regions of high luminosity. [18] built on the work [17] by including the inability of the HVS to detect errors in regions of low luminosity. Podilchuk in [19] compare the results of block based DCT watermarking in which the watermark is inserted in mid-range frequency coefficients to wavelet watermarking using Watson's [20] visual model. Podilchuck concluded that watermarking strength could be more aggressive using a perceptually based model rather than a global scaling weight which results in overly aggressive in some regions while being extremely conservative in others. Watson's popular model [20] using sine-wave grating experiments against gray backgrounds for threshold detection does not take the masking properties of a cover work into consideration. This leads to relatively conservative watermark weighting factors.

In the medical community there is an established culture of placing trust in those that handle the sensitive documents related to our ailments and maladies. Over the past three decades, the medical industry has benefited immensely from the rapid increase in the advances in digital imaging technology. However, with this technology a number of issues accompany it that could hamper that trust. Copyright infringement of images via reproduction, ownership dispute and distribution are reg-

ular concerns of the digital imaging arena. Security plays a bigger role in the storage, transmission and retrieval of medical images. According to [21], security is defined by authentication, privacy, immutability, accountability, traceability and origination. We believe securing medical images can be achieved through digital watermarking. Digital watermarks should be unobtrusive, robust, universal and unambiguous [22]. An unobtrusive watermark is imperceptible (visually lossless) to the human eye and as a result we propose this as diagnostically lossless. A robust watermark remains impervious to compression, reproduction and geometric distortions such as rotation, scaling and translation. Universal and unambiguous refer to being able to employ the watermarking scheme across multiple media forms and accuracy of owner identification respectively. Of major concern to the medical community is the intactness of the region of interest (ROI) of an image, complexity of the scheme should prioritize speed of execution in applications such as telediagnosis, ability to be detected in the presence of noise, resistance to filtering as well as capacity [23]. The challenge therefore is to design a watermarking scheme that meets the above stated criteria and DICOM standards [24] [25] As of today the medical industry has yet to implement a standard medical image watermarking scheme. The area of medical image watermarking is an active field of research and this section provides a background on the state of the art algorithms specifically used for medical images. The digital watermarking technologies available for medical images can be segmented based on the domain of embedding the watermark namely; spatial domain watermarking and transform domain watermarking. Further subdivision includes localization of the watermark in Region of Non-Interest (RONI), ROI or spread-spectrum embedding of the watermark.

Spatial Domain Watermarking

Spatial domain watermarking algorithms embed watermarks into an image via modification of intensity values of pixels of the image. Jagadish et al. [26] employed various Error Correcting Codes (ECC) on encrypted Electronic Patient Record (EPR) a watermark embedded by modifying the least significant bit (LSB) of pixels of grayscale medical. Watermarks of signal graphs such as Electrocardiograms, electroencephalograms etc. were compressed using Differential Pulse Code Modulation (DPCM) then embedded into the image. Presence of the watermark was evaluated in the presence of noise. The Bit Error Rate (BER) was higher for un-coded than ECC watermark with increase in noise. Fan and Hongbin in [27] embed a watermark by modifying an images pixel intensity weighted according to the local variance of the image. They relate the BER to the carrying capacity of the image and find that watermarking should be associated with the content of the image. In [28] , [29] and [30] also watermark the LSB bit-plane.

Deepthi and Niranjan [31] present a watermarking scheme encrypting ASCII codes of EPR using a reversible log function embedded in the LSB of the image. Signal graphs were also embedded using DPCM. A Normalize Root Mean Square Error measure ascertained that the diagnostic value of the image was not reduced. Zain et al. [32] proposed a reversible watermark in the LSB of ultrasound image specifically in RONI. The rationale for LSB embedding is its fragility to attack thus enabling detection of tampering. However, an original image can only be verified if the watermark was not distorted. Spatial domain embedding of a digital watermark via techniques such as LSB embedding is essentially fragile. The shortcomings of the above algorithms include the lack of adaptability to the HVS perception of the medical image and reduced carrying capacity of the images as the watermark is no longer obfuscated when higher bit-planes are modified to increase capacity.

Transform Domain Watermarking

Transform domain watermarking refers to algorithms that embed a digital watermark in an image via modification of transform coefficients. The popularity of compression methods such as Discrete Cosine Transforms (as in JPEG) and Discrete Wavelet Transforms (DWT) as in JPEG2000 has increased focus on implementation in these transform domains. Lee et al. [33] tackle forgery of medical images via a robust method of embedding the watermark in RONI. The watermark is the 8-bit plane representation of the coefficients of the LL3 subband of the 3 Level DWT of the ROI and embedded in the RONI of the original image. The algorithm requires the presence of the original image to extract the watermark defeating the purpose of securing the image with a digital watermark. In [25], a block-based method DCT watermarking was employed.

Giakoumaki et al. in [34] exploited the spatial and frequency resolution capability of the DWT to embed multiple watermarks in medical images. Tampering of an image is detected by watermarking wavelet coefficients in preselected areas of the image including the ROI. Digital signatures and EPR are watermarked in RONI of the image. Peak signal-to-noise Ratio (PSNR) was used as the key performance indicator as well as an evaluation by a physician. Badran et al. in [35] combine the work of Giakoumaki et al and Lee et al. by inserting a robust watermark in the RONI of the DWT coefficients of the image for EPR while inserting the tamper detection watermark in the spatial domain in the ROI of the image.

This thesis contributes to the field by presenting psychophysical results and an associated algorithm which allow the consideration of masking properties of an image based on local luminance, local contrast and local entropy to model the HVS perception to visual distortions. We first conducted a psychophysical experiment to determine the masking effect of natural images on detection of wavelet-coefficients distortions in Chapter 3. Next in Chapter 4, we used these results to train a model

to automatically determine those regions in an image where distortions are visible, we also employed an image-quality model to perform the same task. An application of the suggested model was used to design a compresssion scheme to demonstrate that shaping quantization steps could provide a compressed image of higher visual quality than regular compression. Finally, we performed an additional experiment to study the effect of an image's context in the detection of spatially centered targets in Chapter 5.

CHAPTER 3

PSYCHOPHYSICAL EXPERIMENT I

3.1 Introduction

Over the years, the colleciton of a subjective Mean-Opinion-Score through experimentation has been criticized as being both expensive and time-consuming [36]. Arguing otherwise, the benefits of conducting an experiment to determine the feasibility of a model far outweigh the costs. Secondly, to model such a complex system as the HVS should involve a characterization through experimentation and subject feedback. In accordance, the aim of this experiment was to obtain maps illustrating the masking effect of an image to a human observer. A psychophysical experiment was performed in which human subjects were given a set ten images, each with varying levels of distortions. The subjects were required to indicate regions in an image (human visible error maps) where they could visually detect error introduced via wavelet coefficient distortions. The resulting maps were split into two sets, the first set was used for training and the second set was used to validate and evaluate the algorithm's performance.

3.2 Methods

3.2.1 Apparatus and Subjects

Stimuli used in the experiment were displayed on a Sceptre X24WG 24-inch monitor at a resolution of 1900×1200 pixels. The display monitor yielded a minimum luminance of 0.02 cd/m^2 and a maximum luminance of 101 cd/m^2, at a frame rate of 60Hz. The

monitor was calibrated to determine its gamma function by using a Minolta LS-110 photometer to map its digital-pixel-value to luminance response. Subjects viewed the stimuli binocularly through natural pupils in a darkened room at a distance of approximately 60 cm.

A total of eight subjects took part in this study. Subjects were engineering graduate students of varying degree of experience in evaluation and detection of wavelet coefficient distortions in images. Their objective was to mark regions where they could see an error in the distorted image. The two authors of the experiment did not participate in the study due to their familiarity with the image content. Subjects ranged in age from 22 to 28 years. All subjects had either normal or corrected visual acuity.

Each observer had 100 images to evaluate with each image containing various magnitudes of distortion at the LH2 and LH3 subbands. at scales 2 and 3 (spatial frequencies of 2.3 and 4.6 cycles/degree respectively).

3.2.2 Stimuli and Methods

The Stimuli consisted of 10 512×512 Figure 3.1 bitmap gray scale natural images with varying content of activity. The stimulus was presented as a two forced choice experiment by presenting an original (no watermark) and the watermarked image. The distorted images were generated from the original images by applying a five level decomposition discrete wavelet transform (DWT) using 9/7 biorthorgonal DWT filters [37]. The resulting coefficients were distorted by the addition of uniformly distributed noise at spatial frequencies of 2.3 and 4.6 cycles/degree horizontal orientations according to Equation (3.1).

$$\mathbf{x_m} = \mathbf{x} + \alpha\mathbf{n} \tag{3.1}$$

where α represents weighting magnitude of the distortion $\mathbf{n}$. Figure 3.2 is a subset

Figure 3.1: 1st Row: *baby*, *balloon*, *flowers*, *frog*, 2nd Row: *hawaii*, *horse*, *lena*, *lily*, 3rd Row: *seagulls*, *wall*

of the 10 images used in the experiment. The top row of the figure shows natural images and the second row their watermarked versions with the 3rd row representing the results of a subject's regions of the images where the distortions were detected.

The subjects were presented with three images on the screen; the original image $\mathbf{X}$ and the watermarked image $\mathbf{X_m}$ placed adjacent to each other, and a gridded version of the watermarked image placed at the bottom of the monitor to provide location reference. All images were presented on a uniform background of 26.2 cd/m^2. The subjects were required to select via the mouse the regions of the watermarked image whereby they noticed visible errors. The subjects were given a period of adaptation to a 26.2 cd/m^2 blank background. Response time was not limited as a watermarked image could be subject to deep scrutiny. However, subjects were required to refrain from shortening their viewing distance. Ground truth maps of region of visible distortions was then generated by each subject and are discussed in the next section.

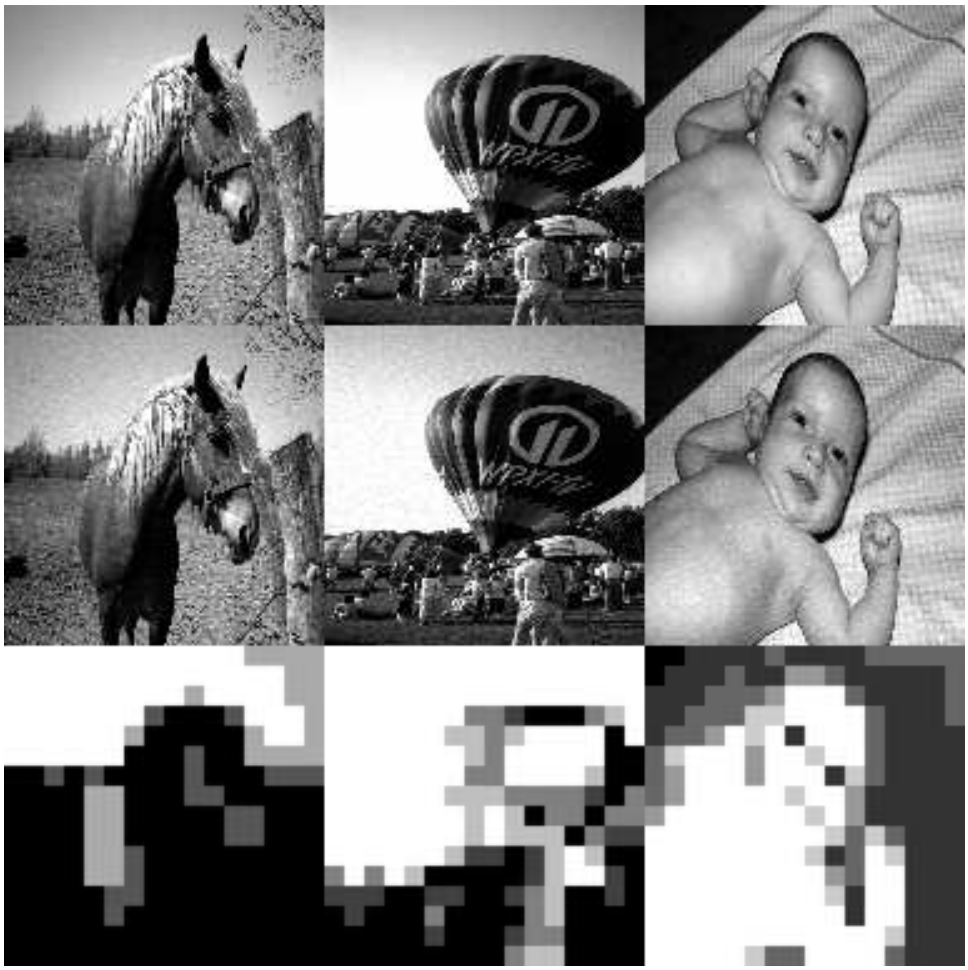

Figure 3.2: *From left to right: a) horse b) balloon c) baby* 1st row source images, 2nd row scale 2 watermarked images, 3rd row ground truth maps of visible distortion

3.2.3 Results and Discussion

The just noticeable difference (JND) of the distortions with locality resolution in the spatial domain were collected by averaging the responses of the eight subjects. These maps are shown in Figures 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12. The regions of the ground truth maps with higher intensities, indicate that the distortions in the image was clearly visible at that location. According to the ground truth maps, darker regions indicate regions were the subjects were unable to detect distortions.

Although, the masking image provided cover for dsistortions in certain areas, the regions where the subjects could detect the distortions increased as a function of

α (See Section 3.2.2. Table 3.1 uses several quality measures to rate the distorted images. The higher the distortion magnitude, the lower the quality of the distorted image. This is important because visual quality of an image can assist in properly tailoring an image application algorithm to a human visual system.

The challenge that was faced having collected the ground truth maps was to design a model that would predict the same regions of visibility of distortions. In Chapter 4, we explore image properties (statistical measures) that can provide clues into where a distortion is visible, we also apply the developed human visual model to a compression algorithm that provides a compressed image with higher visual quality than regular JPEG2000.

Table 3.1: Quality Rating of Distorted Images for *PSNR*, *SSIM* and *VIF* the higher the rating the lower the level of distortion, for *MSE* and *MAD* the lower the score the lower the level of distortion

IMAGE		*PSNR*	*SSIM*	*VIF*	*MSE*	*MAD*
baby	scale2 dist 1	45.9808	0.98705	0.91429	1.6406	0.14219
	scale2 dist 3	38.891	0.93964	0.79045	8.3941	22.2926
	scale2 dist 5	34.8046	0.86565	0.71063	21.5089	63.283
	scale2 dist 7	31.7104	0.773	0.64788	43.8575	107.2259
	scale2 dist 9	29.3281	0.68137	0.59691	75.9049	148.2656
	scale3 dist 1	50.948	0.99721	0.9594	0.52274	0
	scale3 dist 3	44.3547	0.98864	0.88944	2.3857	4.982
	scale3 dist 5	40.4506	0.97281	0.82317	5.8616	24.6697
	scale3 dist 7	37.4068	0.94937	0.75838	11.8142	49.1726
	scale3 dist 9	34.8401	0.91567	0.699	21.3338	82.3647

Continued on Next Page...

IMAGE		*PSNR*	*SSIM*	*VIF*	*MSE*	*MAD*
balloon	scale2 dist 1	46.3662	0.98894	0.9313	1.5013	2.9087
	scale2 dist 3	39.3492	0.95209	0.82275	7.5537	25.9753
	scale2 dist 5	35.3562	0.89989	0.74276	18.9436	62.7189
	scale2 dist 7	32.2932	0.83285	0.67653	38.3495	104.4536
	scale2 dist 9	29.9605	0.76839	0.61988	65.6193	144.6134
	scale3 dist 1	51.4697	0.99701	0.97013	0.46356	0.50236
	scale3 dist 3	44.8138	0.98845	0.91197	2.1464	7.1097
	scale3 dist 5	40.6788	0.97414	0.85079	5.5616	26.8866
	scale3 dist 7	37.779	0.95345	0.79626	10.8437	47.9041
	scale3 dist 9	35.5323	0.93168	0.74801	18.1905	73.9507
flowers	scale2 dist 1	46.4698	0.98639	0.93496	1.4659	1.1354
	scale2 dist 3	39.4964	0.94213	0.83269	7.302	22.0639
	scale2 dist 5	35.4379	0.88202	0.75492	18.5904	56.7766
	scale2 dist 7	32.4568	0.81195	0.69397	36.9317	96.0023
	scale2 dist 9	29.9912	0.74299	0.63609	65.1568	138.1298
	scale3 dist 1	51.5583	0.99569	0.97092	0.4542	0.28395
	scale3 dist 3	45.0143	0.98279	0.9177	2.0495	7.0589
	scale3 dist 5	40.9811	0.96258	0.86198	5.1877	24.7253
	scale3 dist 7	37.9118	0.93859	0.80604	10.5171	50.5493
	scale3 dist 9	35.3533	0.90888	0.75345	18.9562	78.0569
frogs	scale2 dist 1	46.0013	0.98455	0.91201	1.6329	5.7708
	scale2 dist 3	38.9752	0.93044	0.78796	8.233	40.3654
	scale2 dist 5	34.8611	0.84663	0.71305	21.2309	85.3144
	scale2 dist 7	31.8563	0.75005	0.65384	42.4078	131.2891
	scale2 dist 9	29.3114	0.64467	0.6017	76.1967	177.8298

Continued on Next Page...

IMAGE		*PSNR*	*SSIM*	*VIF*	*MSE*	*MAD*
	scale3 dist 1	51.2075	0.99677	0.95953	0.49242	1.7137
	scale3 dist 3	44.3972	0.98605	0.88312	2.3624	17.8125
	scale3 dist 5	40.3276	0.967	0.81699	6.03	44.8826
	scale3 dist 7	37.3375	0.93996	0.75549	12.0041	77.4991
	scale3 dist 9	34.8034	0.90224	0.69781	21.5151	113.6745
hawaii	scale2 dist 1	45.9856	0.98939	0.92704	1.6388	0.22852
	scale2 dist 3	39.0486	0.9536	0.81508	8.095	21.4036
	scale2 dist 5	34.907	0.89652	0.73296	21.0079	59.6314
	scale2 dist 7	31.8753	0.82988	0.66825	42.2231	98.9113
	scale2 dist 9	29.5066	0.76091	0.61643	72.8481	139.6708
	scale3 dist 1	51.0405	0.99745	0.96765	0.51171	0
	scale3 dist 3	44.5129	0.98968	0.90879	2.3003	6.8784
	scale3 dist 5	40.3823	0.9764	0.84728	5.9546	24.3228
	scale3 dist 7	37.5011	0.95745	0.79426	11.5602	50.1493
	scale3 dist 9	34.9554	0.93218	0.73882	20.775	78.4249
lena	scale2 dist 1	45.9464	0.98793	0.91974	1.6536	1.4183
	scale2 dist 3	38.935	0.94438	0.81047	8.3096	29.5936
	scale2 dist 5	34.8052	0.87562	0.73799	21.506	69.2729
	scale2 dist 7	31.7159	0.7915	0.6804	43.8019	110.6313
	scale2 dist 9	29.2562	0.70426	0.63204	77.1717	149.8306
	scale3 dist 1	51.0874	0.99752	0.96431	0.50622	0.15669
	scale3 dist 3	44.6399	0.98988	0.90245	2.234	9.3021
	scale3 dist 5	40.3007	0.97449	0.83839	6.0675	29.5431
	scale3 dist 7	37.1553	0.95058	0.77993	12.5186	59.4614

Continued on Next Page...

IMAGE		*PSNR*	*SSIM*	*VIF*	*MSE*	*MAD*
	scale3 dist 9	34.772	0.92034	0.72764	21.6712	88.8075
lily	scale2 dist 1	46.1177	0.9876	0.91846	1.5897	6.963
	scale2 dist 3	39.0754	0.94346	0.78803	8.0452	40.0085
	scale2 dist 5	34.9474	0.87487	0.69608	20.8132	80.3103
	scale2 dist 7	31.926	0.7921	0.62505	41.7327	125.4613
	scale2 dist 9	29.5057	0.70329	0.56537	72.8634	168.6822
	scale3 dist 1	51.3131	0.99722	0.96346	0.48059	2.3621
	scale3 dist 3	44.6706	0.98839	0.89428	2.2183	12.6681
	scale3 dist 5	40.5074	0.97148	0.82272	5.7855	32.6108
	scale3 dist 7	37.4638	0.94802	0.75878	11.6602	61.01
	scale3 dist 9	35.0091	0.91674	0.69866	20.5197	91.9464
seagulls	scale2 dist 1	45.885	0.9883	0.92047	1.6772	1.6481
	scale2 dist 3	38.8876	0.94807	0.80322	8.4007	38.3768
	scale2 dist 5	34.7678	0.8862	0.72037	21.6923	85.903
	scale2 dist 7	31.6944	0.81265	0.65859	44.0194	134.2094
	scale2 dist 9	29.2639	0.73783	0.60601	77.0354	180.3489
	scale3 dist 1	51.0375	0.99755	0.96722	0.51207	0.1269
	scale3 dist 3	44.4442	0.98987	0.90395	2.337	15.7475
	scale3 dist 5	40.2473	0.97553	0.83929	6.1426	41.6893
	scale3 dist 7	37.2816	0.95489	0.78175	12.1597	72.1544
	scale3 dist 9	34.9823	0.92839	0.72835	20.6465	101.5148
wall	scale2 dist 1	45.9538	0.98883	0.92515	1.6508	0
	scale2 dist 3	38.9007	0.94913	0.81102	8.3756	23.9761
	scale2 dist 5	34.8431	0.88809	0.73575	21.3193	65.2258
	scale2 dist 7	31.7517	0.81235	0.67518	43.4421	108.4775

Continued on Next Page...

IMAGE		*PSNR*	*SSIM*	*VIF*	*MSE*	*MAD*
	scale2 dist 9	29.3097	0.73192	0.62346	76.2281	149.9899
	scale3 dist 1	51.0551	0.99764	0.96795	0.50999	0
	scale3 dist 3	44.3484	0.99014	0.90741	2.3891	6.7927
	scale3 dist 5	40.3235	0.97621	0.84887	6.0357	27.868
	scale3 dist 7	37.2794	0.95486	0.78931	12.1658	55.6175
	scale3 dist 9	34.7449	0.92743	0.73451	21.8065	86.1695

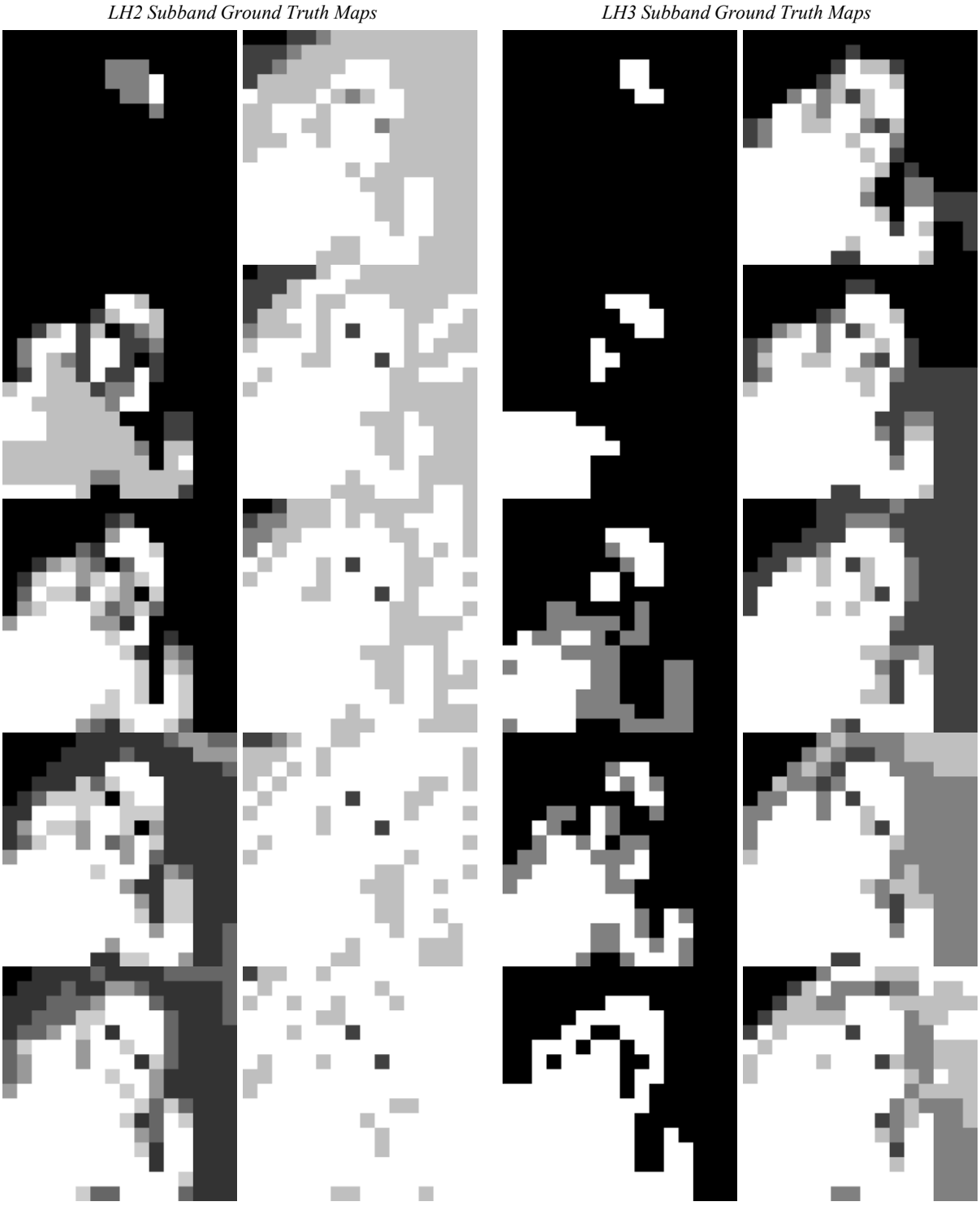

Figure 3.3: *Ground truth maps for baby generated by averaging subjects response to increasing α weights of distortions accross each image*

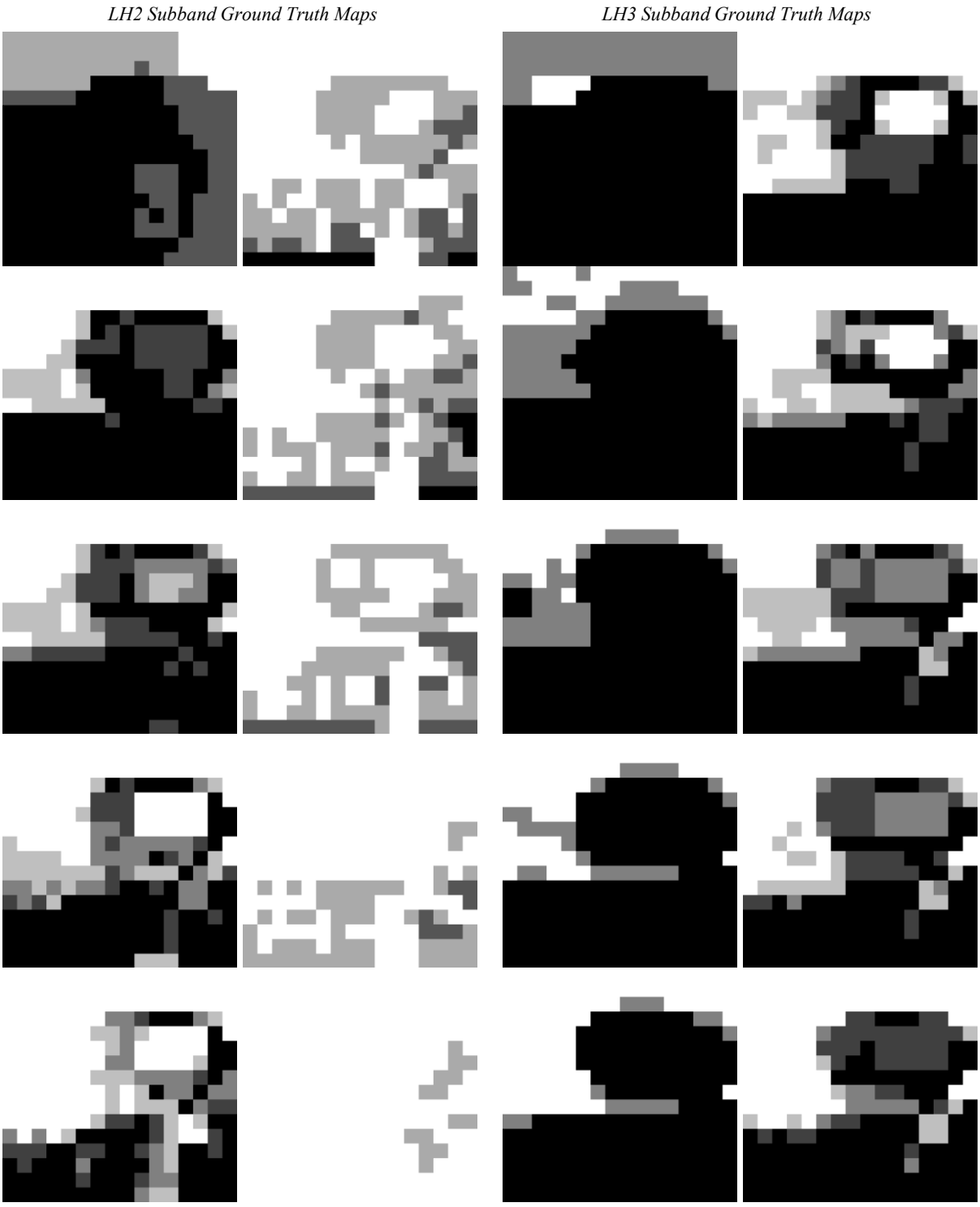

Figure 3.4: *Ground truth maps for balloon generated by averaging subjects response to increasing α weights of distortions accross each image*

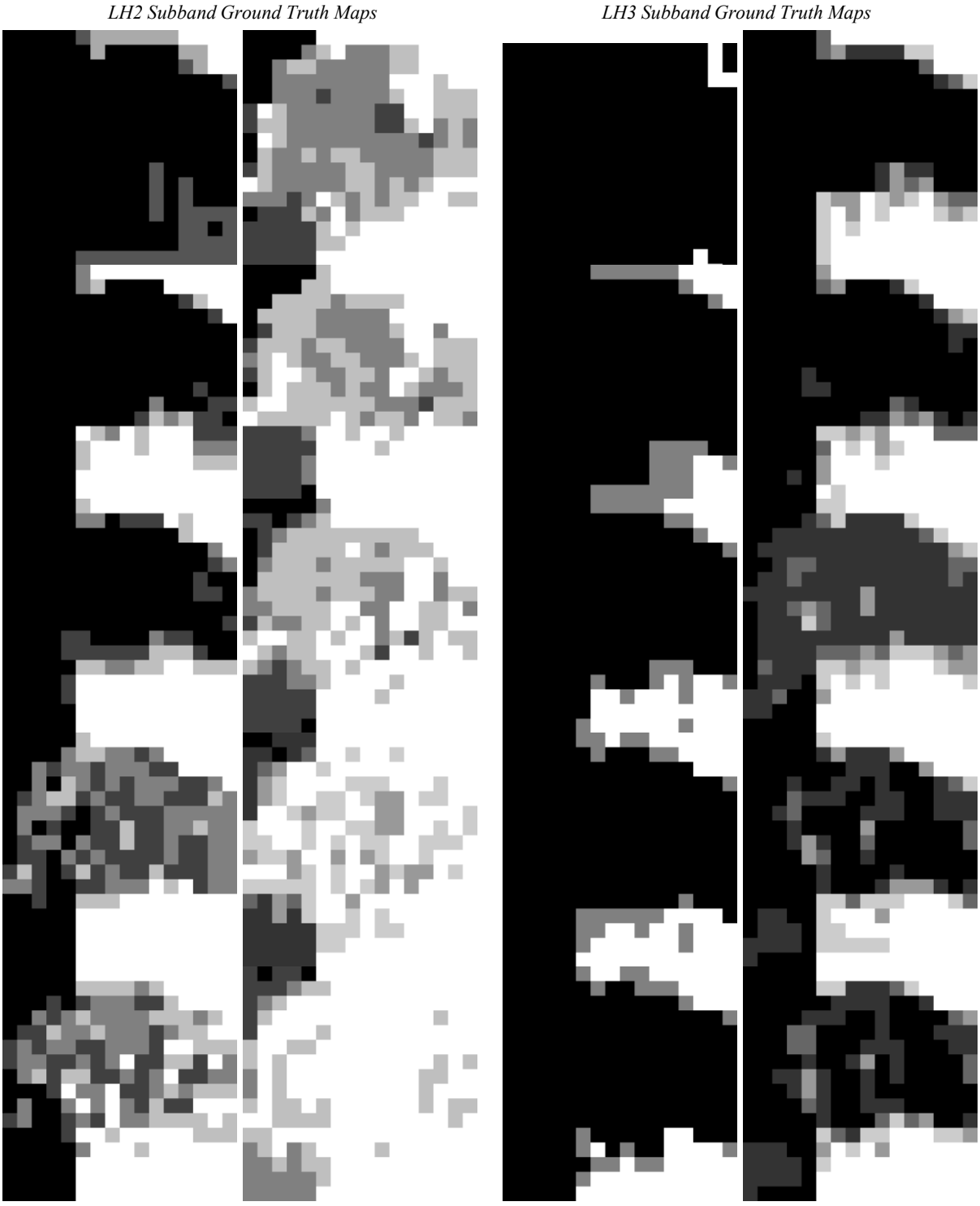

Figure 3.5: *Ground truth maps for flowers generated by averaging subjects response to increasing α weights of distortions accross each image*

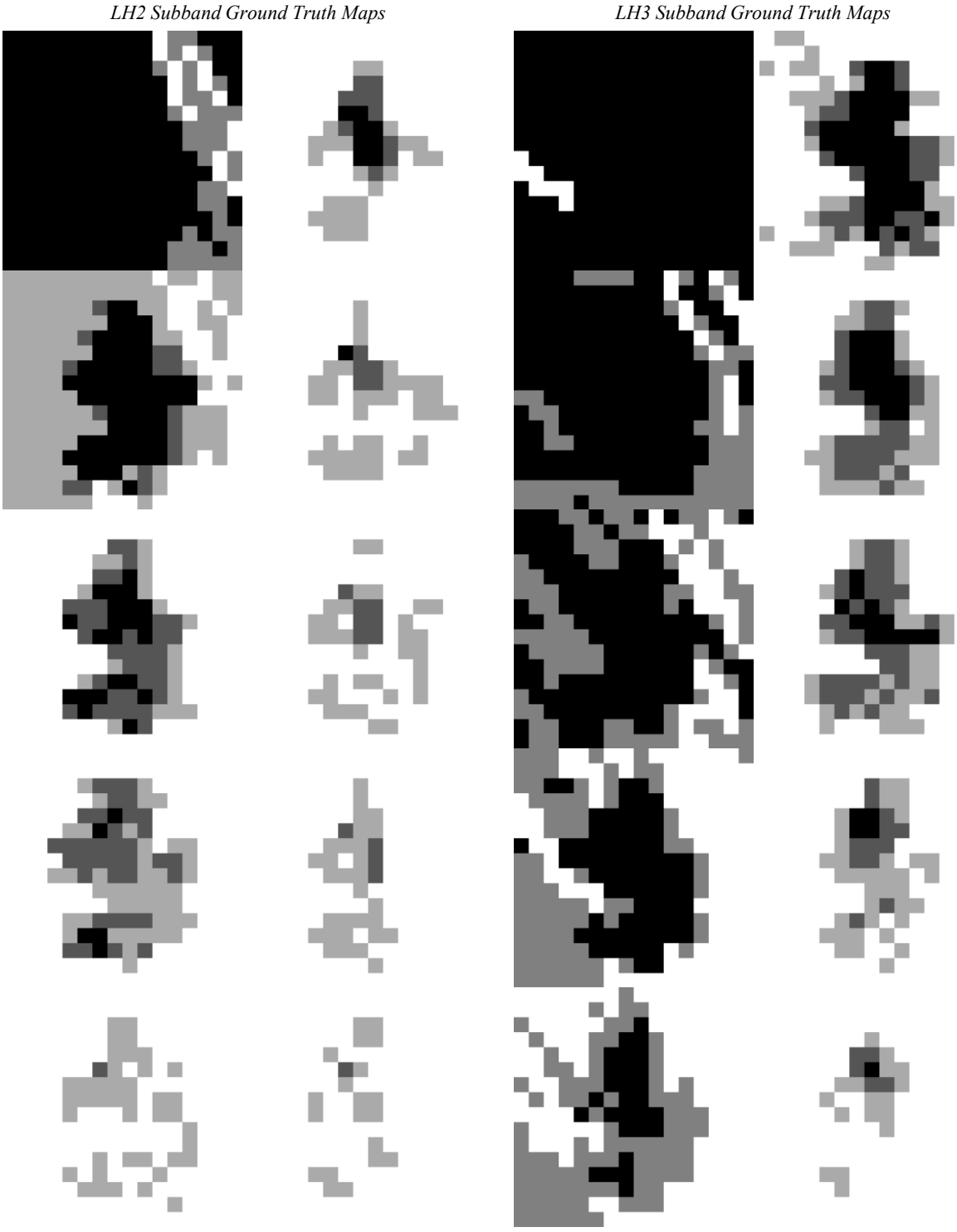

Figure 3.6: *Ground truth maps for frog generated by averaging subjects response to increasing α weights of distortions accross each image*

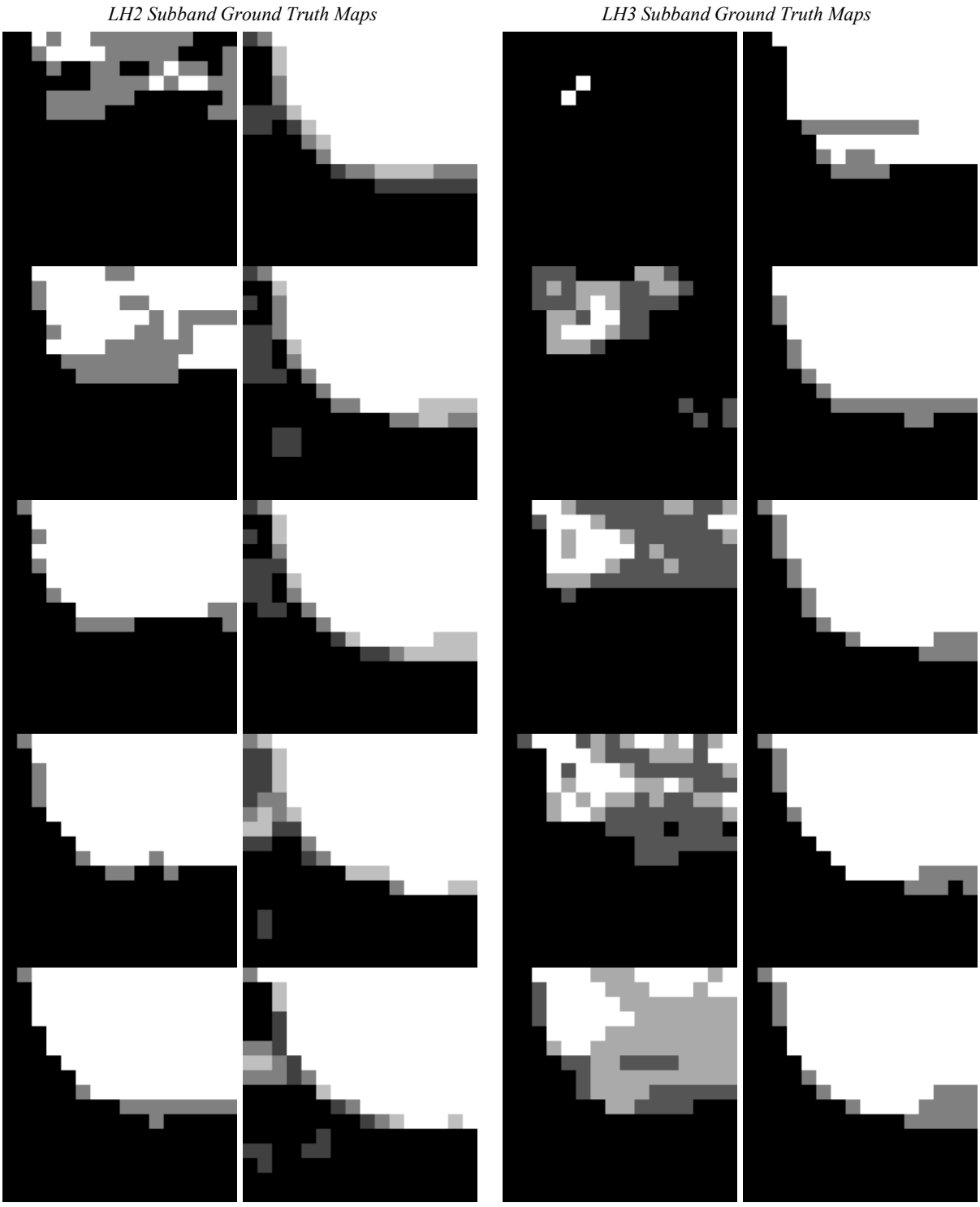

Figure 3.7: *Ground truth maps for hawaii generated by averaging subjects response to increasing α weights of distortions accross each image*

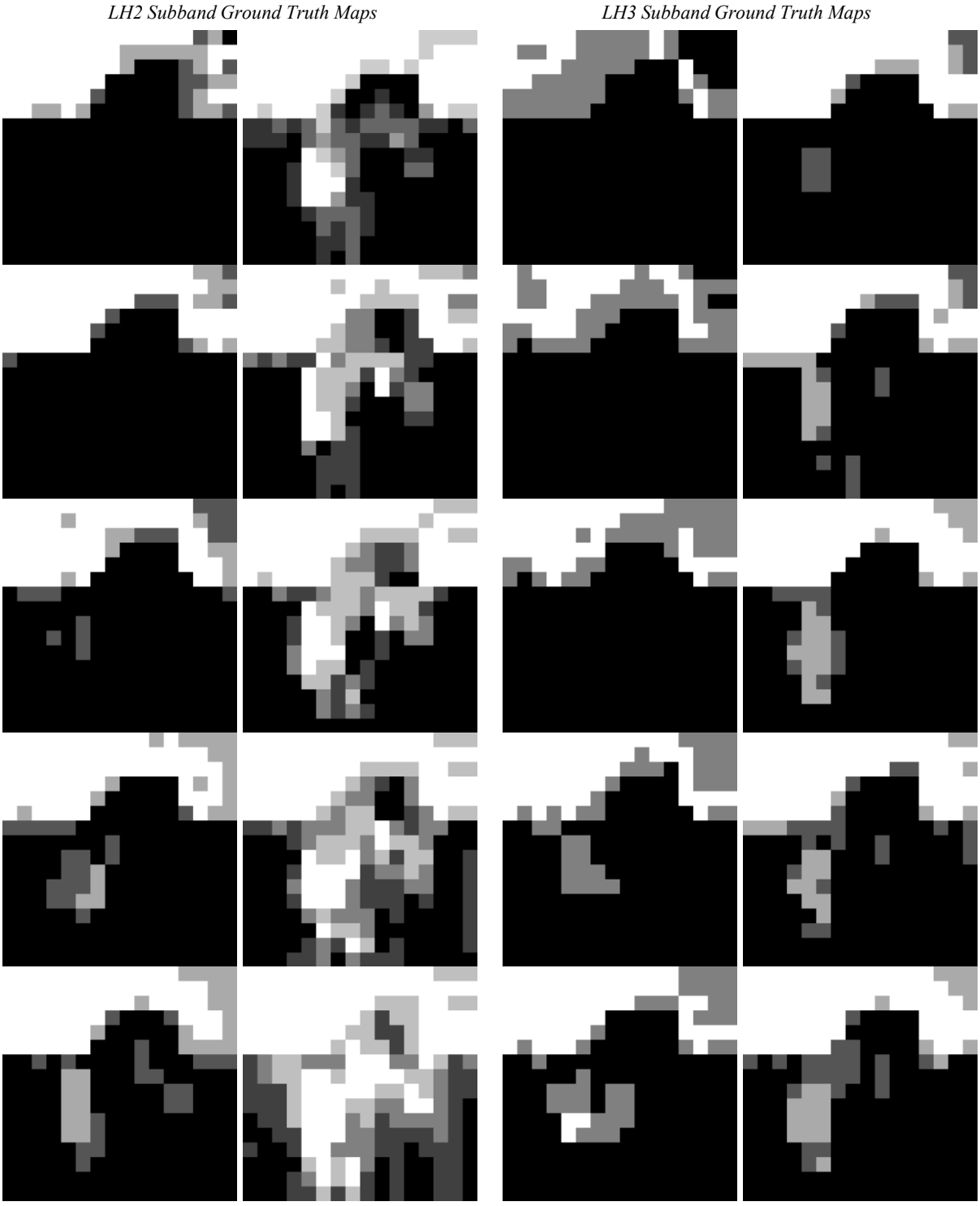

Figure 3.8: *Ground truth maps for horse generated by averaging subjects response to increasing α weights of distortions accross each image*

Figure 3.9: *Ground truth maps for lena generated by averaging subjects response to increasing α weights of distortions accross each image*

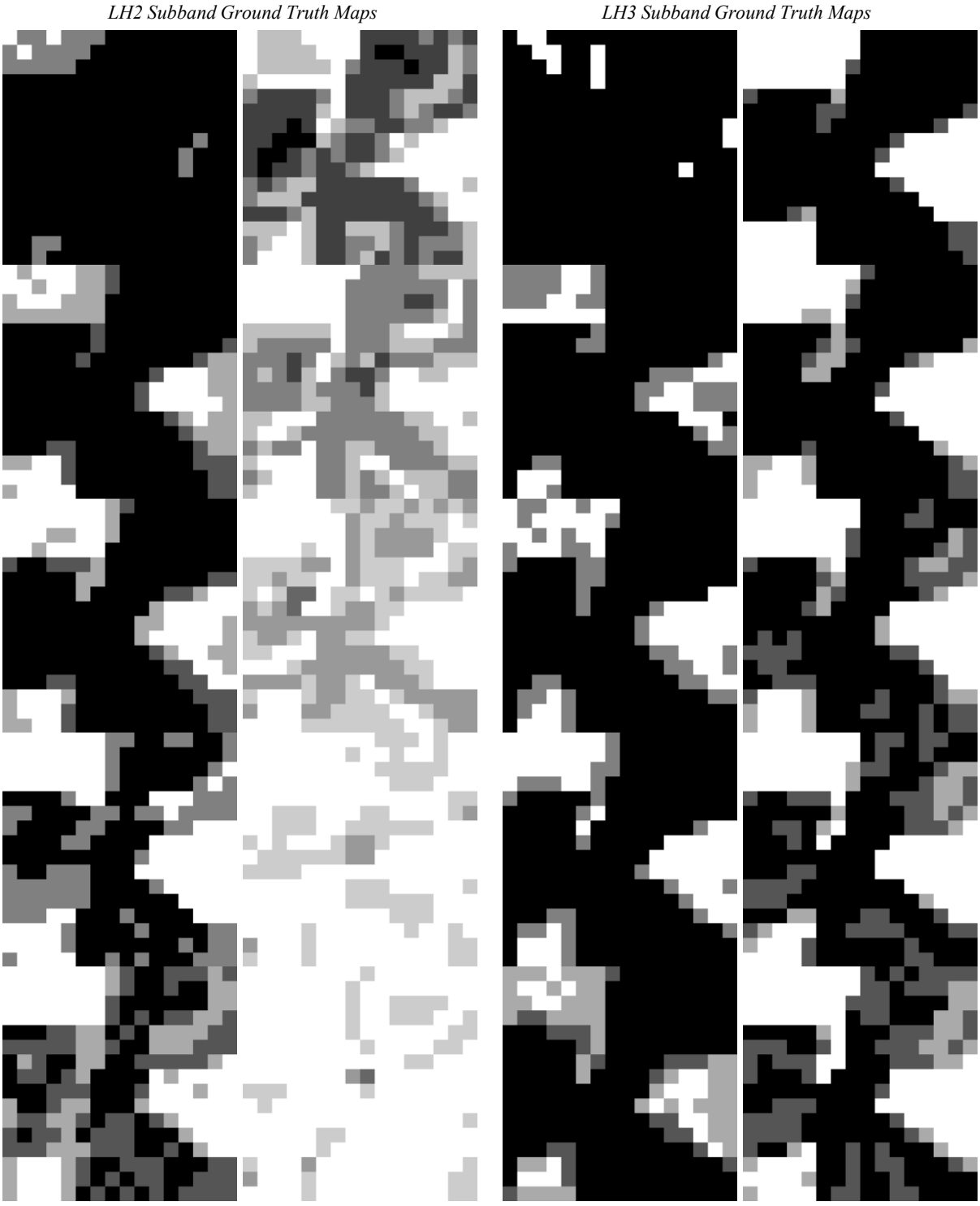

Figure 3.10: *Ground truth maps for lily generated by averaging subjects response to increasing α weights of distortions accross each image*

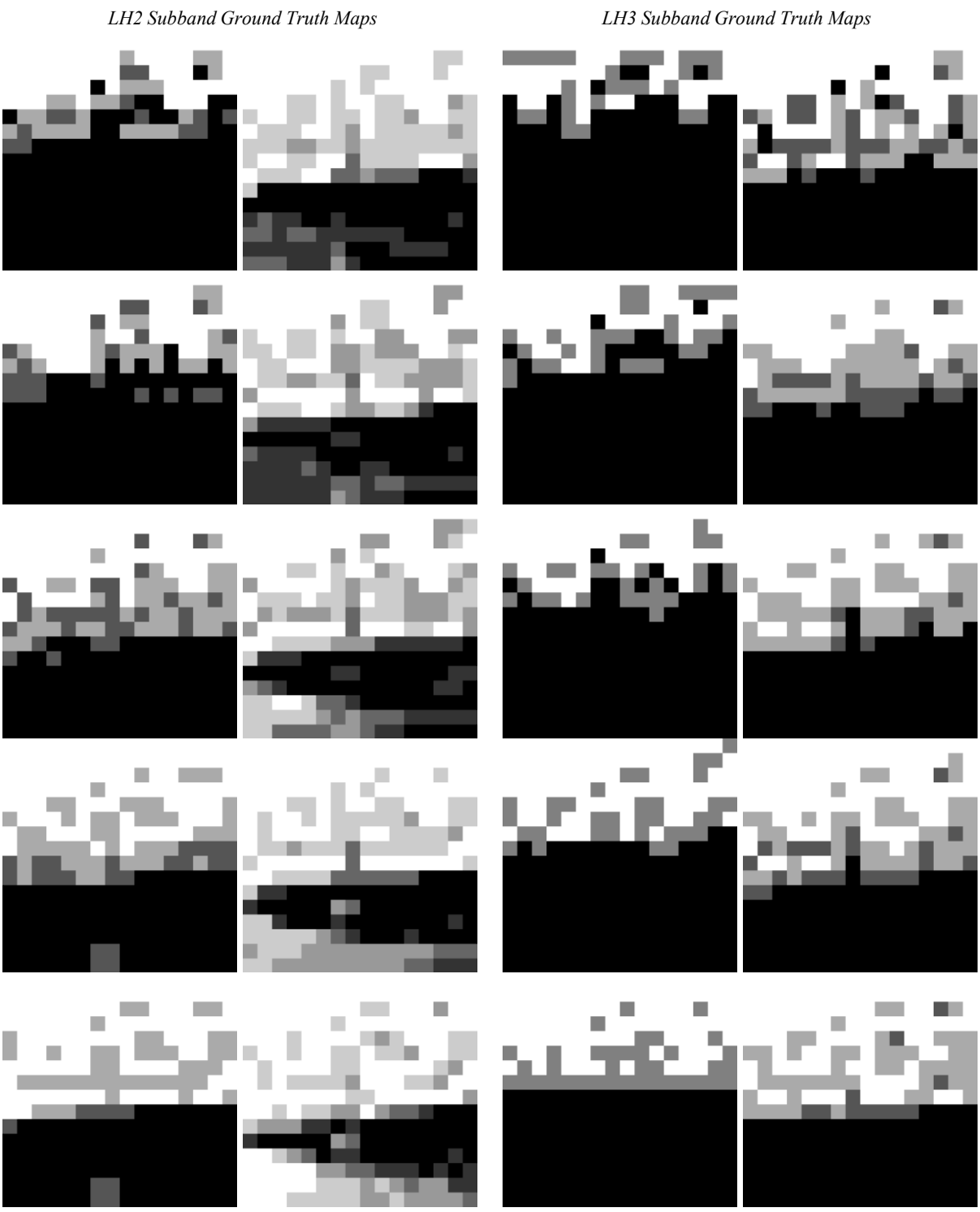

Figure 3.11: *Ground truth maps for seagulls generated by averaging subjects response to increasing α weights of distortions accross each image*

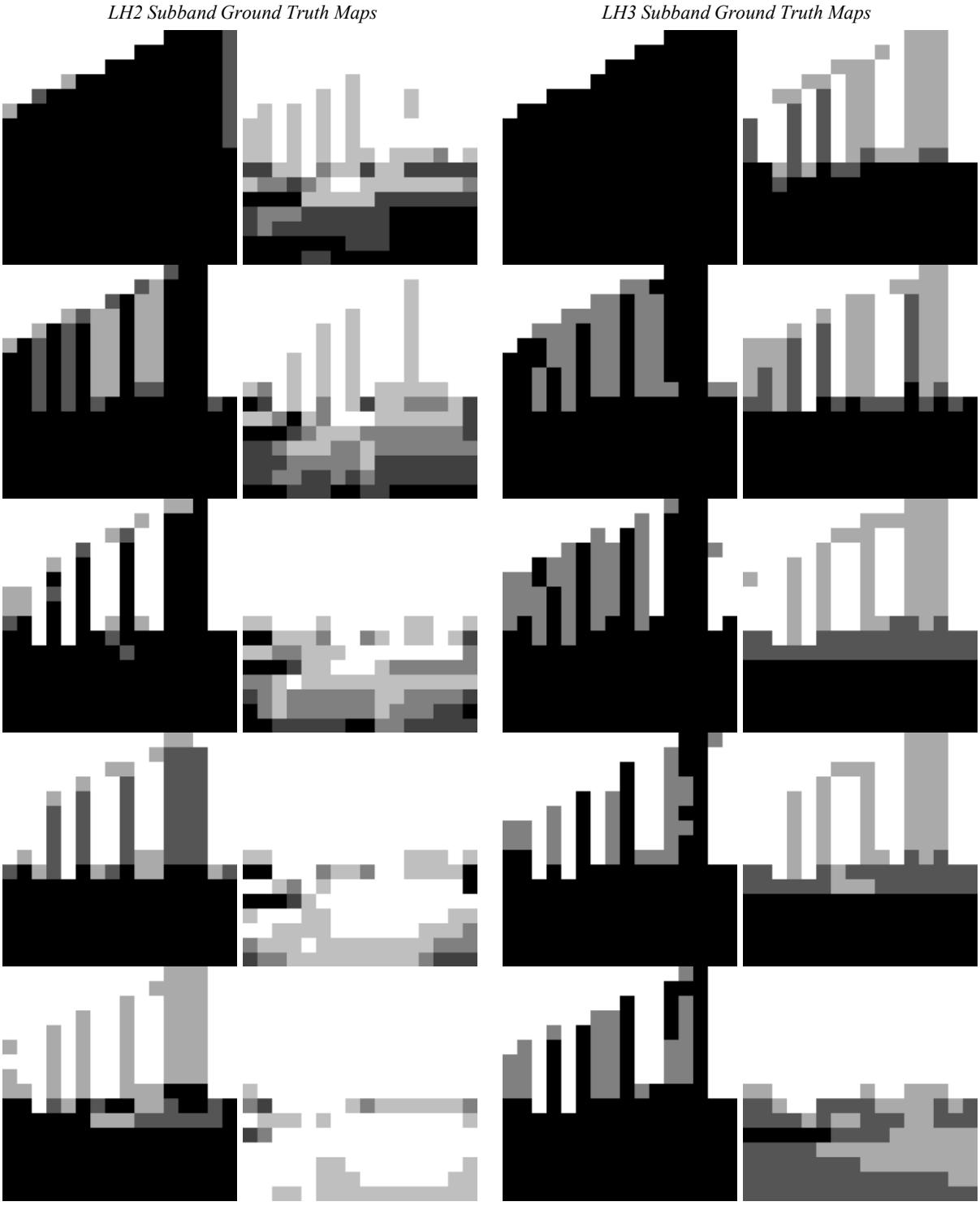

Figure 3.12: *Ground truth maps for wall generated by averaging subjects response to increasing α weights of distortions accross each image*

CHAPTER 4

ALGORITHM AND APPLICATIONS

4.1 Introduction

The VEPM, outlined in Figure 4.1 and Figure 4.2, uses spatial properties of a source image to determine regions in the image where wavelet domain distortions are visible. The feature maps of luminance, contrast and entropy are image dependent and are combined to provide an activity map. The activity scaling parameter (P_{act}) determines a threshold where the error activity must exceed the image activity in order for the error to be considered visible. Let $\mathbf{X}$ denote an MxN pixel source image, $\mathbf{M}$ denotes maps generated with subscripted values indicating respective properties or domain transforms. The algorithm is tailored to perform two tasks as follows:

1. Given two images, an original and distorted image, it can predict the regions of the distorted image where a human observer will detect the distortions.

2. Provide visibly enhanced compressed images at the same or increased ratios when combined with a compression algorithm in comparison with regular compression.

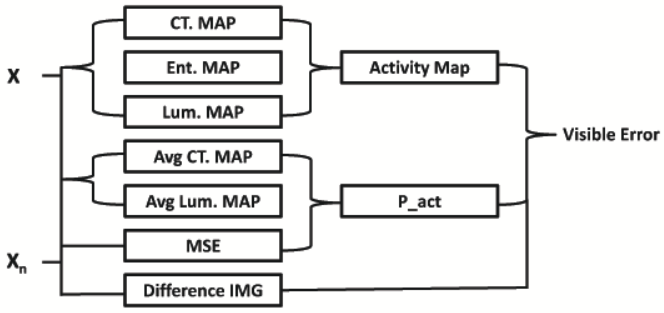

Figure 4.1: *VEPM (linear)* Process Block Diagram

4.2 Feature Maps

4.2.1 Preprocessing

The images were first converted into luminance values $\mathbf{X_{lum}} = (b + k\mathbf{X})^{\gamma}\ cd/m^2$, where $b = 0$, $k = 0.02874$ and $\gamma = 2.2$ according to sRGB standard. $\mathbf{X_{lum}}$ was sent through a CSF function, Equation (4.1)

$$\mathbf{A}(f) = 2.6 * (0.0192 + 0.114 * f) * e^{-(0.114*f)^{1.1}} \tag{4.1}$$

As described in 2.2.2 to model the HVS response to spatial frequencies. Statistical computations were performed on a blocks of size $m \times m$ with 50% overlap of neighboring adjacent blocks.

4.2.2 Luminance Map

The average luminance of each block M_{lum} was used to adjust the contrast and entropy maps to discard regions of extreme luminosity or extreme darkness as described in Section 4.1.

4.2.3 Contrast Map

The local spatial contrast of an image provides information of the range of intensities contained in that image. First the contrast map was computed using $\mathbf{M_{ct}} =$

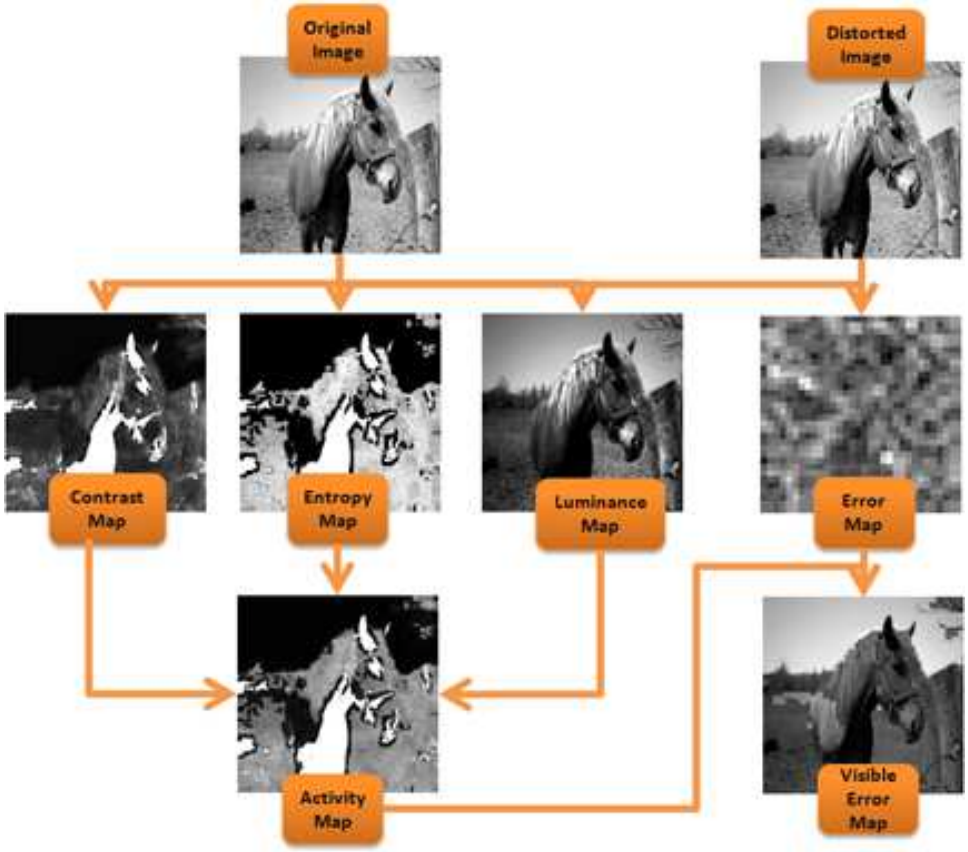

Figure 4.2: VEPM Process Block Diagram

$\sigma_{\mathbf{x_{lum}}}/\mu_{\mathbf{x_{lum}}}$ as shown in Figure 4.3 (c). The contrast map was then adjusted logarithmically to stretch the intensity values of dark regions using $\mathbf{M_{ct_{adj}}} = log(\mathbf{M_{ct}} + 1)$.

4.2.4 Entropy Map

To compute the entropy map $\mathbf{M_{ent}}$ a symmetric gray level co-occurrence matrix (GLCM) to extract activity information using $\mathbf{M_{ent}} = \sum_{i,j=0}^{N-1} \mathbf{P}_{i,j}(-ln\mathbf{P}_{i,j})$. To reduce edge effects which tend to inflate entropy values, blocks $\mathbf{x}$ of the image were subdivided into 4 equal sized blocks with each having its entropy calculated. The entire block was then assigned the minimum entropy value of the subdivided blocks. The entropy map was also adjusted via scaling and thresholding of low entropy regions which were mapped to zero.

4.2.5 Activity Map

An activity map was generated by combining the various individual feature maps via a β-norm

$$\mathbf{M_{Act}} = ((0.5 * \mathbf{M_{ct_{adj}}})^{\beta} + (0.5 * \mathbf{M_{ent_{adj}}})^{\beta})^{1/\beta} \tag{4.2}$$

Note that we apply a threshold to $\mathbf{M_{act}}$ to handle the case in which the image activity is low; specifically, values of $\mathbf{M_{act}} < \mathbf{0.005}$ are set to zero.

The various computed maps of image *horse* are shown in Figure 4.3. Regions of higher intensity denote high activity and therefore would mask distortion better than regions of low intensity. AS stated earlier very low intensity regions according to the luminance map have been given amplified intensities in the activity map to illustrate their masking abilities.

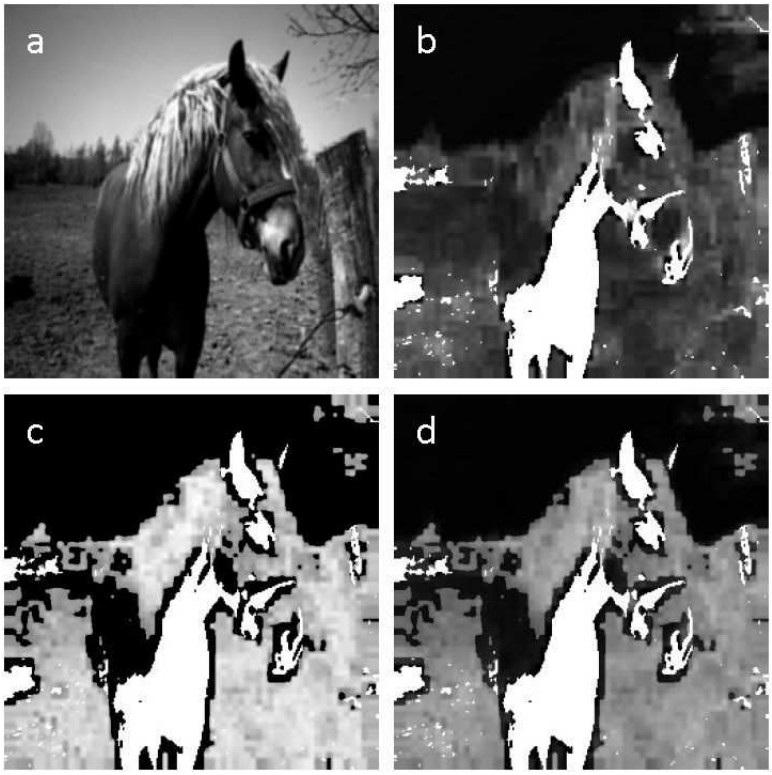

Figure 4.3: *from left to right:* a) luminance map b) contrast map c) entropy map d) activity map

4.3 Distortion Visibility Evaluation

4.3.1 Activity Scaling Parameter (P_{act})

To evaluate the masking capability of each region, a block based comparison of average error intensity and average scaled image activity was made. If the error activity was above the scaled image activity, the error was considered visible. Specifically, a map indicating regions of visible distortion $\mathbf{Vis_{err}}$ is obtained as follows:

$$\mathbf{Vis_{err}} = \begin{cases} 1 & if\ \mathbf{M_{err}} > P_{act} \times \mathbf{M_{act}}, \\ 0 & otherwise. \end{cases} \tag{4.3}$$

Here (P_{act}) is an activity scaling parameter which is adaptively computed on a per image basis. (P_{act}) represents the value by which the ratio of the error activity $\mathbf{M_{err}}$ to the image activity $\mathbf{M_{act}}$ must exceed in order for the distortion to be deemed visible at the corresponding location in the image.

4.3.2 Linear Function Based Model

In order to predict (P_{act}) on a per-image basis, four test images *horse, balloon, baby, and frog* were used as a training set for a general linear model. Specifically, A linear regression model given in Equation (4.2) was used to model the P_{act} as a function of the average luminance of $\mathbf{X}$,the average contrast of $\mathbf{X}$ and the mean square error between $\mathbf{X}$ and its distorted version:

$$P_{act} = 5.722 * \mathbf{X}_{avgct} + 0.149 * \mathbf{X}_{avglum} - 0.071 * \mathbf{X}_{mse} \tag{4.4}$$

The psychophysical experiment provided ground truth data maps which were averaged as shown in the bottom row of Figure 3.2 for scale 2 subband with SS distortion magnitude of $\alpha = 12.5$. Using MATLAB's inbuilt fminsearch function and averaging over starting points of $\mathbf{P_{act}} = \begin{bmatrix} 5 & 20 & 35 & 65 & 75 \end{bmatrix}$, the average P_{act} for four test images *horse, balloon, baby, and frog* were obtained.

4.3.3 Image Quality (IQ) Based Model

Training on a set of data and testing on a separate set is useful in many applications today, however, in the area of image processing, the diversity of images in both content and quality cannot be fitted by a simple training and testing set. In [38] Larson and Chandler suggest that the human visual system uses several strategies to determine the quality of an image. Local spatial properties, as described above of contrast and luminance as well as the suggested entropy can be related to the *Most Apparent Distortion MAD* measure provided in [38]. *MAD* employs a dual strategy namely; a high-quality regime whereby the HVS seeks distortions in the image's presence and a low-quality regime in which image content is sought in the presence of distortions. Therefore if the distortions have large magnitude, more visible blocks are required to define the whole image in the presence of the distortions as opposed to when the distortions are of low magnitude.

As described earlier, (P_{act}) the image error $\mathbf{M_{err}}$ to the image activity $\mathbf{M_{act}}$ ratio, in this case can be regarded as the quality ratio whereby distortions in an image are considered acceptable to the observer. In Table 3.1, the quality ratings of the distorted images were measured. Here we define a function that relates the perceptual quality of the distorted image's to (P_{act}). The conducted experiment, gave the range of (P_{act}) = 40 indicating no visibility of distortions and (P_{act}) = 0 for distortions that were completely visibile over the entire image. The quality measure (Q_{MAD}) range for *MAD* was taken to be (Q_{MAD}) $\approx$ 0, for a watermarked image $\mathbf{X_m}$ to be indistinguishable from the original $\mathbf{X}$ and (Q_{MAD}) $\geq$ 150 for a watermarked image $\mathbf{X_m}$ to be perceptibly unrecognizable from the original $\mathbf{X}$. Figure 4.4 shows graphs the relationship between (P_{act}) and Q_{MAD} defined by Equation (4.5).

$$P_{act} = \frac{-40}{(1 + e^{-\frac{1}{Q_{MAD}}})} + 40 \tag{4.5}$$

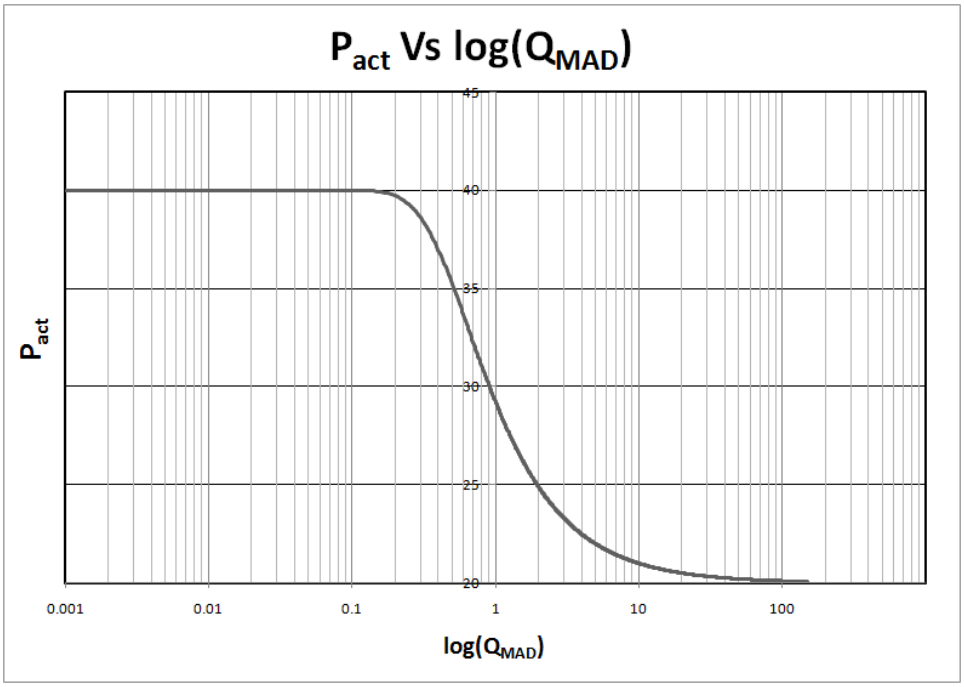

Figure 4.4: Graph of P_{act} vs. $\log Q_{MAD}$

4.4 Results

To evaluate the performance of the model, three test images *lena*, *flowers* and *seagulls* Figure 4.5 were distorted at the LH3 with a magnitude of $\alpha = 9.5$. The test involved assessing how well the VEPM algorithm performed as compared to ground truth and also as compared to a Watson-based visual model. MSE was used to quantitatively evaluate the performance of both models.

The values of P_{act} calculated using Equation (4.4) above for the three images are given in Table 4.1. The linear regression model provided good estimates of P_{act} for the test images, higher P_{act} values indicate high detection of wavelet-based distortions in the image.

The visible error maps computed by the VEPM closely matches that of ground truth data with an MSE of 0.1206 for *lena*, 0.083 for *flowers* and 0.0799 for *seagulls*. Although Watson's model MSE values could indicate it being closely matched with the VEPM it indicated several regions where ground truth considered distortions

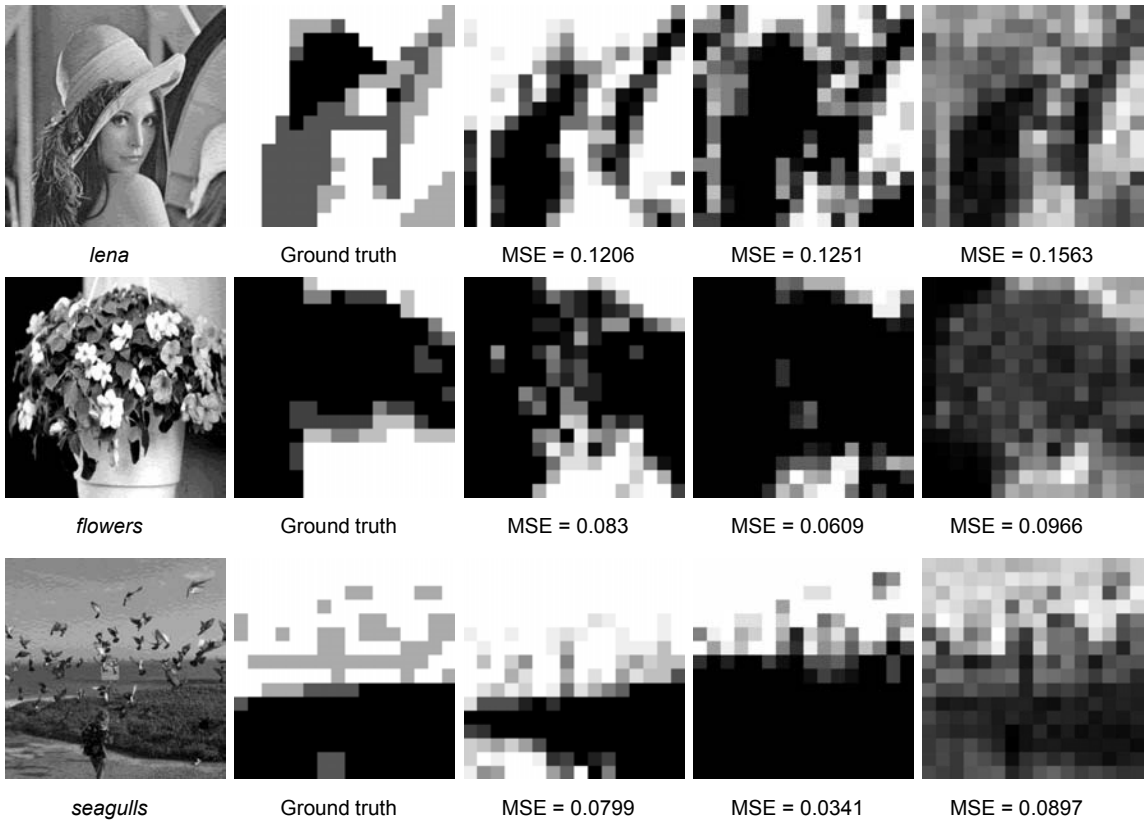

Figure 4.5: *Ground truth Maps compared to VEPM(linear) map VEPM(IQ) map and Watson map*

highly visible as not visible or completely masked by the Image. In the image of *flowers*, ground truth maps visible error in regions of low activity almost completely ignoring the presence of distortion in the leaves and petals region. VEPM detected this response and although it predicted visibility in some floral regions, its overall response indicated its sensitivity to regions of activity and low contrast. These results are extended in wavelet domain distortions at other scales and subbands.

4.5 VEPM Application

In this section, we examine the results of applying the *activity map* generated by the *VEPM* algorithm to a compression scheme. The block diagram of the process is depicted in the following Figure 4.6.

Table 4.1: Computed P_{act}, VEPM and Watson MSE values

Image	**VEPM(IQ)**	**VEPM(linear)**	**Watson**
Lena	0.1251	0.17	0.1563
Flowers	0.0609	0.083	0.0966
Seagulls	0.0341	0.0799	0.0897
Average MSE	0.0734	0.1328	1.1547

An image is analyzed by the *VEPM* and subsequently, the *activity map* generated is multiplied with the original image to reduce the coefficients used in compressing regions of elevated masking effects. The propsed scheme is illustrated in the block diagram, Figure 4.6. The image is then passed to a JPEG2000 compression algorithm with the resulting compressed image scaled back with the *activity map*. The result is an image whereby bits reserved for compressing all areas of the image are pushed to areas where distortions are more visible generating an image with a higher visual quality.

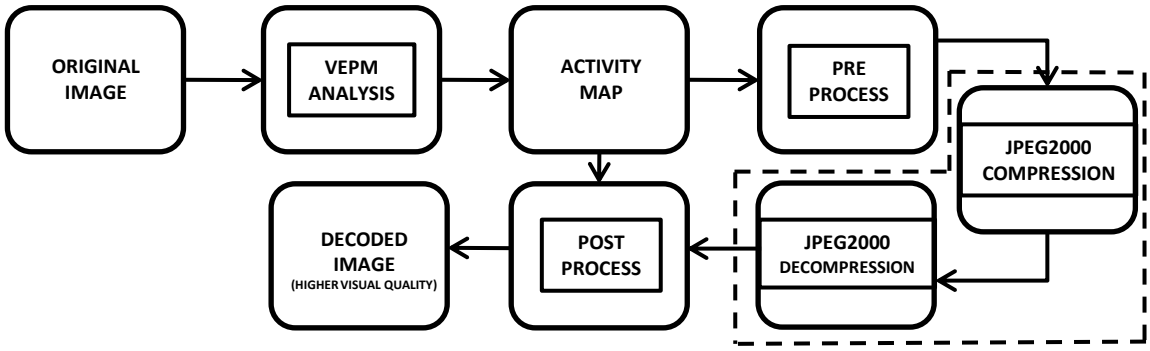

Figure 4.6: Block diagram of VEPM compression scheme.

At higher bit rates 0.75 *bpp*, the compressed image of *horse* with regular JPEG2000 is found to have a poorer image visual quality than when compressed with the *VEPM* scheme with a *MAD* rating of 142.126 as compared to 139.237. In addition, when

both images are compressed at 0.50 *bpp* and 0.25 *bpp* , *VEPM* once again prevails as providing a compressed image with higher visual quality ratings of 57.225 to 78.450 and 24.639 to 53.623 respectively. *VEPM* however, fails in the image of *balloon* due to the activity map overcompensating for image distortions in regions of low contrast. The quality ratings of other images are provided in Table 4.2. Overall, as the average measure indicates, *VEPM* out-performs regular JPEG2000 with a *MAD* rating of 48.331 to 54.481. The other quality assessment measures lie within statistical signicance of each other and thus did not provide a meaningful comparison. This can be attributed to *MAD's* dual strategy, which has led it to beat other image quality assessment algorithms [38]. The next Chapter, shall try to investigate further other influences that could potentially cause the detection of distortions to vary based on contextual influence.

Figure 4.7: Left: *horse* Top: Regular JPEG2000 Bottom: VEPM-JPEG2000 *(0.75bit/pix)*

Figure 4.8: Left: *horse* Top: Regular JPEG2000 Bottom: VEPM-JPEG2000 *(0.50bit/pix)*

Figure 4.9: Left: *horse* Top: Regular JPEG2000 Bottom: VEPM-JPEG2000 *(0.25bit/pix)*

Table 4.2: Quality ratings comparing JPEG2000 with the proposed scheme for 10 different images at bit rates (Bits per Pixel) of 0.25 *bpp*, 0.50 *bpp*, 0.75 *bpp* using *PSNR*, *SSIM*, *VIF* and *MAD*

		PSNR		SSIM		VIF		MAD	
IMAGE	*bpp*	*VEPM*	*JP2*	*VEPM*	*JP2*	*VEPM*	*JP2*	*VEPM*	*JP2*
baby	0.25	35.206	35.380	0.886	0.889	0.467	0.463	28.260	28.669
	0.50	37.838	38.228	0.924	0.929	0.659	0.665	5.228	5.965
	0.75	39.426	39.794	0.942	0.946	0.764	0.786	0.255	0.207
balloon	0.25	27.868	28.207	0.838	0.851	0.350	0.361	147.092	141.214
	0.50	31.551	31.834	0.902	0.915	0.518	0.542	79.594	68.608
	0.75	34.257	34.955	0.932	0.943	0.644	0.658	48.957	35.160
flowers	0.25	29.160	29.372	0.829	0.854	0.380	0.383	127.448	124.378
	0.50	33.113	33.646	0.889	0.917	0.583	0.586	47.001	41.866
	0.75	35.743	36.229	0.917	0.942	0.703	0.729	21.178	16.225
frog	0.25	37.495	37.631	0.921	0.922	0.526	0.531	40.035	39.334
	0.50	40.335	40.386	0.947	0.947	0.709	0.705	12.307	13.589
	0.75	41.960	42.092	0.961	0.961	0.803	0.806	6.593	7.075
hawaii	0.25	27.281	27.639	0.839	0.835	0.311	0.299	72.462	104.205
	0.50	30.756	31.200	0.918	0.912	0.498	0.477	24.238	43.604
	0.75	33.942	34.432	0.951	0.948	0.602	0.617	13.026	20.710
horse	0.25	25.722	26.074	0.733	0.745	0.271	0.271	139.237	142.126
	0.50	28.400	28.756	0.834	0.837	0.406	0.417	57.225	78.450
	0.75	30.186	30.912	0.882	0.882	0.528	0.514	24.639	53.623
lena	0.25	33.992	34.142	0.888	0.888	0.448	0.450	47.646	49.941
	0.50	37.029	37.275	0.925	0.925	0.638	0.640	14.957	17.496
	0.75	38.773	38.981	0.943	0.942	0.749	0.746	5.771	8.256
lily	0.25	32.461	32.935	0.860	0.868	0.404	0.407	49.069	52.808
	0.50	35.674	36.172	0.917	0.925	0.594	0.609	16.161	14.594
	0.75	37.891	38.558	0.944	0.951	0.719	0.723	6.664	10.058
seagulls	0.25	27.222	27.266	0.770	0.772	0.287	0.282	159.779	187.951
	0.50	29.734	29.926	0.854	0.864	0.445	0.452	85.140	108.856
	0.75	32.008	32.327	0.906	0.907	0.562	0.538	44.220	70.448
wall	0.25	30.705	30.739	0.801	0.803	0.361	0.351	93.285	101.670
	0.50	33.200	33.549	0.877	0.882	0.549	0.539	32.431	33.506
	0.75	35.404	35.477	0.923	0.919	0.645	0.665	15.048	13.842
AVERAGE		33.478	33.804	0.888	0.894	0.537	0.540	48.831	54.481

CHAPTER 5

PSYCHOPHYSICAL EXPERIMENT II

5.1 Introduction

In Chapter 3 we examined how the regions of an image masked the distortions as detected by a human observer. The basis of that experiment was to illustrate the effects of luminance, contrast and entropy masking. However, these three properties alone fail to completely characterize the human visual system's varying response to a distorted texture in an image when this texture is found in different areas of the image. Therefore, there seems to be some factors influencing the visual systems' ability to project the same thresholds for similar textures in different locations distorted with the same α magnitude as in Equation (3.1).

Most computational models of masking have focused heavily on a bottom-up approach in which modeled neurons are influenced by a local neighborhood in radial frequency, orientation, and position (see, e.g., Daly [39], Teo & Heeger [40], Legge et al. [14] and Watson & Solomon [41]). In particular, most models of masking are operated in a highly spatially localized manner in which local detection thresholds are determined by considering only a relatively small spatial neighborhood around each location; a larger spatial context is used only implicitly at coarser spatial scales. However, masking in natural images may also be influenced by top-down factors (e.g., recognition or category) which are handled by higher-level visual mechanisms [42], [43]. Thus, it remains unclear whether a model that considers only low-level image attributes can fully account for masking in natural images.

One higher-level attribute, which is the focus of this Chapter, is the effect of

natural vs. unnatural surrounds (contexts) on the detection of distortions. Previous studies on the effect of context in object detection have shown that contextual influence is exerted on the detection of objects in a scene. Biederman et al. asked observers to discriminate a target in normal versus jumbled scenes and found observers were more accurate on normal scene context [44]. Easier detection of semantically consistent objects versus semantically inconsistent objects in a scene was reported by Boyce et al. [45]. Thus, previous research has provided evidence that there is a need to further explore the contextual influence of scenes on detection of objects. Hollingworth and Henderson however, raise the doubt that the methodology of the experiments to determine contextual dependence on signal detection did not eliminate response bias from sensitivity measures [46].

When an image patch is distorted, the patch itself provides masking, but the context in which this patch is displayed might have different effects: (1) it could provide additional masking, or (2) it might provide facilitation, or (3) it might have no effect. Here, we investigate whether a natural vs. unnatural vs. no context (solid gray) can influence detection of distortions in natural-image patches. When a patch is placed within an unnatural context, the ability to recognize the patch's subject matter is greatly reduced; thus, a logical assumption is that an unnatural context should provide greater masking when factors such as luminance and contrast are held constant. To test this, we measured thresholds for detecting a gabor target at 6.2 cy/deg in natural-image patches which were placed within various contexts, including a solid gray background and the patch's original context. A log-sigmoid blend was used to eliminate edge effects. The contexts were adjusted using histogram specification to match the histogram of the original context in order to control for differences in brightness, contrast, and other first-order statistical properties of the luminance distribution. RMS contrast thresholds for visually detecting the distortions were then measured using a spatial two-alternative forced-choice procedure. This

Chapter discusses the results of the experiment and investigates if detection thresholds for distortion in natural images are context-dependent.

5.2 Methods

5.2.1 Apparatus and Subjects

Stimuli used in the experiment were displayed on a LaCie 324, 24-inch LCD monitor at a resolution of 1900×1200 pixels. The display monitor yielded a minimum luminance of 0.17 cd/m^2 and a maximum luminance of 266 cd/m^2, at a frame rate of 60Hz. The monitor was calibrated to determine its overall gamma of 2.6 by using a Minolta LS-110 photometer to map its digital-pixel-value to luminance response. Subjects viewed the stimuli binocularly through natural pupils in a darkened room at a distance of approximately 93 cm.

A total of six subjects participated in this experiment CV, DC, KV, LV, PV and TP. Subjects were engineering graduate students of varying degree of experience in evaluation and detection of wavelet coefficient distortions in images. Their objective was to mark regions where they could see an error in the distorted image. The authors of the experiment did not participate in the study so as not to influence thresholds due to familiarity and adaptation. Subjects ranged in age from 22 to 35 years. All subjects had either normal or corrected visual acuity.

5.2.2 Stimuli

The images used in the experiment were gray scale 8-bit images. The stimuli consisted of a patch that was extracted from five original context images Figure 5.1 centered and alpha blended to seven different background textures. There were a total of 35 stimuli, where each stimulus consisted of a patch blended to the seven different textures. A through explanation of the patches and textures used is given below:

Figure 5.1: Five natural images *church, eagle, lily, train* and *zebra* in which patches were extracted

Patches: A patch size of 128×128 was extracted from the five original context images Figure 5.2 of size 512×512 such that each patch consisted part of the region of interest from the original context images. Once extracted, a log-sigmoid feathering was applied to the four sides of the patch in order to avoid edge effects when blending with the background and reduce the patch size to 96×96. This reduction in size was necessary because the patch sizes were meant to subtend about 2.13 degrees of viewing angle (size of the fovea).

Figure 5.2: Five natural images patches *church, eagle, lily, train and zebra*

Textures: Seven images with different textures Figure 5.3 of size 384×384 were used as a background image for which the patch was centered. A Log-sigmoid feathering was also applied to the center area of the textures where the patch was placed in order to avoid edge effects.

5.2.3 Procedure

Two images were presented on the screen against on a gray background, side-by-side for the observer. One image contained the distorted patch plus the context texture;

Figure 5.3: Seven textures *weave, rocks, tree, marble, straw, gray, original-context*

the other image contained the original patch plus the context texture. The distorted patch was created by scaling the magnitude of the gabor target. By using a fixed size gabor across different targets, we ensure that contrast masking remains fixed along a single patch. A log-sigmoid feathering was also applied to the sides of the entire image in order to avoid edge effects at the boundary of the image. Histogram specification was used in order to match the histograms of the textures to those of the original contexts from which the patches were extracted. Via keyboard input, the observer indicated which image contained the distorted patch.

A staircase algorithm was employed to converge on the subject's threshold by using a three down one up method. The step-size varied according to the $\mathbf{SC_{step}} = \begin{bmatrix} 30 & 5 & 2.5 & 1 \end{bmatrix}$ vector. The distortion levels were adjusted based on the subject's response to the previous stimuli. The staircase algorithm selected the next step from the step-size vector if a reversal occured. A reversal is defined as the subject getting a correct response followed by an incorrect response. If the subject detected the distortions for three consecutive trials, the contrast on the following stimuli was lowered making it more difficult for the subject to notice, however, a single incorrect response causes the target's contrast to be elevated. The maximum number of trials

was 60. Each stimulus was presented a second apart in order to limit burn in and accommodation of the stimulus in view. Each subject was given seconds to detect the distortion and key in a response. Audio feedback for correct and incorrect detection responses were given to the subject.

5.3 Results and Discussion

To quantify each subject's threshold of detection for the distortion, the RMS contrast of the distortion within the distorted patch was computed. Each subject took the experiment four times with the first set was discarded as the familiarity with the experiment changed thresholds drastically from the first to second attempt. Subsequent trials of the experiment yielded more consistent results within each subject. The resulting contrast thresholds for images church and zebra are listed in Table 1. The contrast thresholds for each subject are also graphed in Figure 5 along with the mean thresholds and standard deviation error bars.

The results reveal that when all other histogram-based factors such as contrast and luminance are equal to that of the patch's original context, there is an influence of context on the detection of the distortions within the patch. Generally, detection of distortions in natural images is subject to contextual influences and varies from image to image. The location of fixation of the subject to detect distortion was found to be a function of the contrast of the masking image. This demonstrated thresholds were affected by peripheral influence.

In the image *church*, the distortions presented to the subjects were harder to detect when the contextual scene was the *marble*. When the contextual scene was that of the original image, the contrast thresholds were also elevated as the recognition of the full scene reduced the subject's certainty in the location of the distortions (though, this was purely a perceptual phenomenon; the location of the distortions did not actually change). Generally, detection of distortions presented against the blank gray

context provided the lowest contrast thresholds. These results also hold in the case of *zebra* in which the gray context provided the lowest contrast thresholds. However, the distortions were also easy to detect in the zebra's original context, perhaps due to the repeated nature of the patterns on the *zebra* making detection similar to detecting distortions in a sinusoidal grating. The high contrast of the *zebra* patch facilitated the detection of the gabor target in *marble* as opposed to *marble* having the most masking effect for the detection of the gabor in the *church* patch.

It was also noted that some texture context (*marble*) had similar patterns as that of the patch (*church*) itself and provided larger error bars for detection of distortion. Subjects also mentioned that the cues which they observed as the context changed also differed. This is an indication of the effect of context in masking the appearance of distortions. We believe that our results suggest the influence of context on masking, but the fact that the thresholds are quite close means that it's probably not recognition at play. Instead, some other low-level factor beyond that captured by the histogram may explain the results. This suggests that current neural models of masking may only need to better take into account additional low-level properties rather than incorporating higher-level properties such as recognition.

Table 5.1: Subject's contrast thresholds to detection of context-masked gabor target

PATCH	**SUBJECT**	*original*	*weave*	*rocks*	*tree*	*marble*	*grass*	*gray*
church	KV trial 2	0.0120	0.0158	0.0116	0.0146	0.0233	0.0074	0.0103
	KV trial 3	0.0077	0.0095	0.0086	0.0133	0.0180	0.0074	0.0095
	TP trial 2	0.0219	0.0233	0.0193	0.0193	0.0310	0.0236	0.0151
	TP trial 3	0.0116	0.0154	0.0041	0.0112	0.0233	0.0146	0.0065
	DC trial 2	0.0133	0.0120	0.0120	0.0185	0.0241	0.0074	0.0099
	DC trial 3	0.0146	0.0103	0.0124	0.0185	0.0236	0.0112	0.0095
	PV trial 2	0.0219	0.0340	0.0202	0.0307	0.0323	0.0223	0.0146
	PV trial 3	0.0077	0.0146	0.0142	0.0198	0.0228	0.0172	0.0091
zebra	KV trial 2	0.0328	0.0201	0.0185	0.0175	0.0185	0.0282	0.0165
	KV trial 3	0.0175	0.0124	0.0099	0.0145	0.0165	0.0216	0.0129
	TP trial 2	0.0653	0.0395	0.0405	0.0271	0.0287	0.0539	0.0308
	TP trial 3	0.0359	0.0261	0.0271	0.0201	0.0313	0.0364	0.0271
	DC trial 2	0.0170	0.0277	0.0242	0.0226	0.0446	0.0277	0.0297
	DC trial 3	0.0129	0.0231	0.0140	0.0251	0.0328	0.0483	0.0226
lily	CV trial 2	0.0139	0.0191	0.0127	0.0163	0.0204	0.0290	0.0180
	CV trial 3	0.0196	0.0143	0.0147	0.0163	0.0171	0.0196	0.0155
	LV trial 2	0.0191	0.0139	0.0187	0.0200	0.0180	0.0237	0.0139
	LV trial 3	0.0159	0.0108	0.0108	0.0151	0.0099	0.0103	0.0111
train	CV trial 2	0.0094	0.0118	0.0118	0.0118	0.0064	0.0130	0.0069
	CV trial 3	0.0058	0.0094	0.0106	0.0153	0.0100	0.0142	0.0020
	LV trial 2	0.0053	0.0233	0.0148	0.0142	0.0106	0.0277	0.0069
	LV trial 3	0.0058	0.0106	0.0161	0.0106	0.0142	0.0239	0.0058

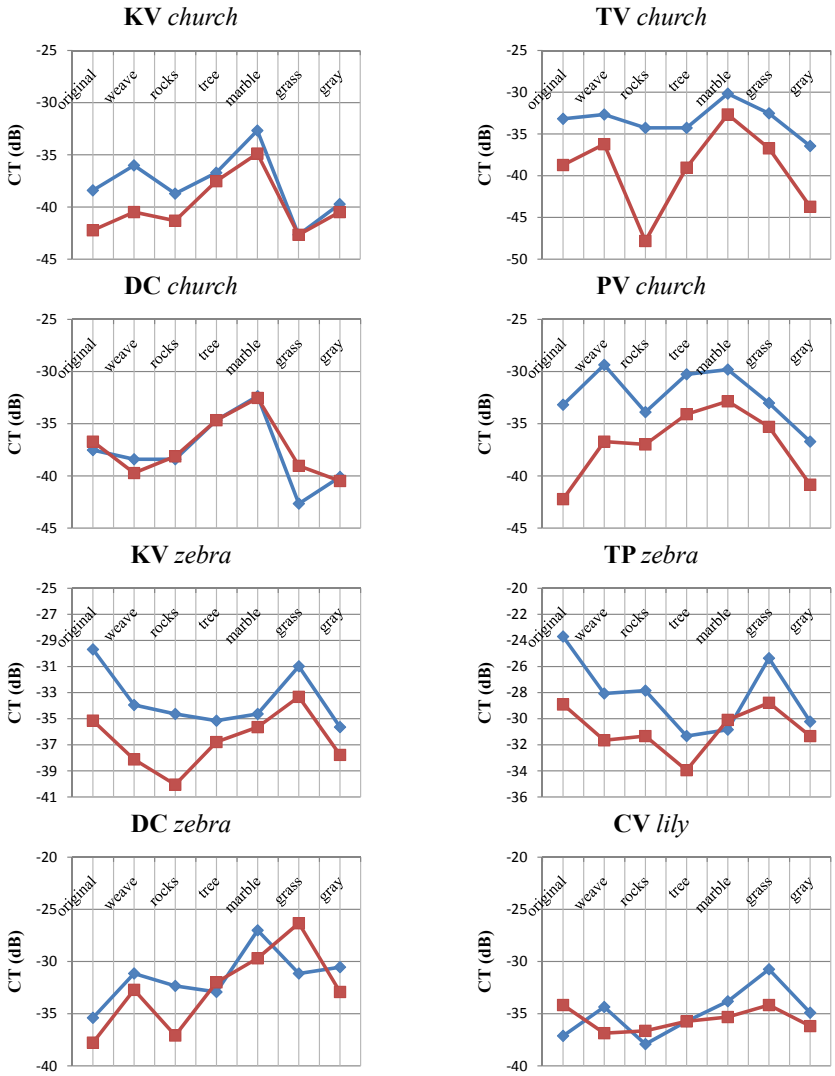

Figure 5.4: log-contrast thresholds for *church*, *zebra*, and *lily*

CHAPTER 6

CONCLUSIONS

Here, we proposed an algorithm to predict the regions in an image where distortions are detected. A psychophysical experiment was performed to determine ground truth maps, which were then used as part of training and testing of the proposed model. Two methods were used to evaluate the algorithms human visible error detection threshold, namely; a linear method and an image quality based method. Our preliminary results demonstrate that the algorithm performs well and correlates well with human generated maps. Furthermore, employing the algorithm in a compression scheme yielded compressed images with higer quality ratings than that of regular JPEG2000 compressed images at the same bit rate. Secondly, another psychophysical experiment has been performed to determine the influence of context on the detectability of distortions in an image. It was found that context may facilitate or mask the detection of a target distortion. We encourage more study in this area and an implementation of context masking in image applications. The results of this thesis indicates there is an advantage of implementing a human visual model in image applications. We are currently investigating quality perception as a measure of tolerance of image distortions.

BIBLIOGRAPHY

[1] S. OTTO H. SCHADE, "Optical and photoelectric analog of the eye," *J. Opt. Soc. Am.*, vol. 46, no. 9, pp. 721–738, 1956.

[2] E. Schonfeld, "Who has the most photos of them all? hint: It is not facebook," April 2009. http://techcrunch.com/2009/04/07/who-has-the-most-photos-of-them-all-hint-it-is-not-facebook/.

[3] AFP, "24 hours of video uploaded to youtube every minute," May 2010. http://www.google.com/hostednews/afp/article/ALeqM5jbG1BC7l8u9Njm7Yo9QV7GLN4uMA/.

[4] Seagate, "Seagate breaks capacity ceiling with world's first 3 terabyte external desktop drive," June 2010. http://www.seagate.com/ww/v/index.jsp?locale=en-US&name=freeagent-goflex-3tb-external-pr&vgnextoid=0d9352a0ca089210VgnVCM1000001a48090aRCRD.

[5] J. T. Fulton, *Biological Vision: A 21st Century Tutorial*. Trafford Publishing, 1st ed., 2004.

[6] P. Thompson, "Margaret thatcher: a new illusion," *Perception*, vol. 9, no. 4, pp. 483–484.

[7] S. E. Palmer, *Vision science: from photons to phenomenology*. MIT Press, 1999.

[8] R. L. DeValois and K. K. DeValois, "Spatial vision," 1990.

[9] D. J. Graham, D. M. Chandler, and D. J. Field, "Decorrelation and response equalization with center-surround receptive fields," *Journal of Vision*, vol. 4, no. 8, p. 276, 2004.

[10] J. A. Ferwerda, "Fundamentals of spatial vision," in *In Applications of visual perception in computer graphics*, 1998.

[11] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the animal, -cats visual cortex," *Journal of Physiology (London*, no. 148, pp. 574–591, 1962.

[12] E. PELI, "Contrast in complex images," *J. Opt. Soc. Am. A*, no. 7, pp. 2032–2040, 1991.

[13] D. G. Pelli, *Effects of visual noise*. PhD thesis, Cambridge University, Cambridge, England, 1980. doctoral dissertation.

[14] G. E. Legge and J. M. Foley, "Contrast masking in human vision," *J. Opt. Soc. Am.*, vol. 70, no. 12, pp. 1458–1471, 1980.

[15] A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," 1997.

[16] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. Morgan Kauffman Publishers, 2002.

[17] A. S. Lewis and G. Knowles, "Image compression using the 2-d wavelet transform," *IEEE Transactions Image Processing*, no. 1, pp. 240–250, 1992.

[18] M. Barni, F. Bartolini, and A. Piva, "Improved wavelet-based watermarking through pixel-wise masking," *IEEE Transactions on Image Processing*, no. 10, pp. 783–791, 2001.

[19] C. Podilchuk and Z. Wenjun, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, no. 16, pp. 525–539, 1998.

[20] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," in *2001 IEEE SIGNAL PROCESSING MAGAZINE 45 A.G. Bors and I. Pitas, Image watermarking using DCT domain constraints, in IEEE Proc. Int. Conf. on Image Processing*, pp. 1164–1175, 1997.

[21] M. A. Epstein, M. S. Pasieka, W. P. Lord, S. T. Wong, and N. J. Mankovich, "Security for the digital information age of medicine: Issues, applications, and implementation," *Journal of Digital Imaging*, vol. 11, pp. 33–34, feb 1998.

[22] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. Morgan Kauffman Publishers, 2002.

[23] K. A. Navas and M. Sasikumar, "Survey of medical image watermarking algorithms," 2007.

[24] H. Yoshimura, Y. Inoue, H. Tanaka, N. Fujita, N. Hirabuki, Y. Narumi, and H. Nakamura, "Operating data and unsolved problems of the dicom modality worklist: An indespensable tool in an electronic archiving environment," *Radiation Medicine*, pp. 68–73, 2003.

[25] B. M. Planitz and A. J. Maeder, "A study of block-based medical image watermarking using a perceptual similarity metric," 2005.

[26] J. Nayak, S. P. Bhat, S. M. Kumar, and R. U. Acharya, "Reliable and robust transmission and storage of medical images with patient information," 2004.

[27] Z. Fan and Z. Hongbin, "Digital watermarking capacity and reliability," 2004.

[28] X. Luo, Q. Cheng, and J. Tan, "A lossless data embedding scheme for medical images in application of e-diagnosis," pp. 17–21, 2003.

[29] G. Wang and R. Ni-ni, "A fragile watermarking scheme for medical image," pp. 3406–3409, 2005.

[30] A. Wakatani, "Digital watermarking for roi medical images by using compressed signature image," 2002.

[31] A. Deepthi and U. Niranjan, "Watermarking medical images with patient information," vol. 20, 1998.

[32] J. M. Zain, L. P. Baldwin, and M. Clarke, "Reversible watermarking for authentication of dicom images," pp. 3337–3340, 2004.

[33] H.-K. Lee, H.-J. Kim, K.-R. Kwon, and J.-K. Lee, "Roi medical image watermarking using dwt and bit-plane," 2005.

[34] S. A. Giakoumaki, S. Pavlopoulos, and D. Koutsouris, "A multiple watermarking scheme applied to medical image management," NaN.

[35] E. F. Badran, M. A. Sharkas, and O. A. Attallah, "Multiple watermark embedding scheme in wavelet-spatial domains based on roi of medical images," 2009.

[36] W. Osberger, A. J. Maeder, and D. McLean, "A computational model of the human visual system for image quality assessment," 1997.

[37] D. M. Chandler, N. L. Dykes, and S. S. Hemami, "Visually lossles compression of digitized radiographs based on contrast sensitivity and visual masking," in *Proceedings of SPIE*, vol. 5749, pp. 359–372, 2005.

[38] E. C. Larson and D. M. Chandler, "Most apparent distortion: a dual strategy for full-reference image quality assessment," vol. 7242, p. 72420S, SPIE, 2009.

[39] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," pp. 179–206, 1993.

[40] P. Teo and D. J. Heeger, "Perceptual image distortion," in *in Proc. SPIE*, vol. 2179, pp. 982–986, 1994.

[41] A. B. Watson and J. A. Solomon, "Model of visual contrast gain control and pattern masking.," *J Opt Soc Am A Opt Image Sci Vis*, vol. 14, pp. 2379–2391, September 1997.

[42] H. GW, R. MJ, and P. CJ, "Top-down processes in object identification: evidence from experimental psychology, neuropsychology and functional anatomy," *Review]. Philos Trans R Soc Lond B Biol Sci*, pp. 352–1275, 1997.

[43] P. Cavanagh, "What's up in top-down processing," in *Representations of Vision*, pp. 295–304, Cambridge University Press, 1991.

[44] I. Biederman, A. L. Glass, and E. W. Stacy, "Searching for objects in real-world scenes," *J. Exp. Psychol*, vol. 97, pp. 22–27, 1973.

[45] S. J. Boyce, A. Pollatsek, and K. Rayner, "Effect of background information on object identification," *Journal of Experimental Psychology: Human Perception and Performance*, no. 15, pp. 556–566, 1989.

[46] J. M. Henderson and A. Hollingworth, "High-level scene perception. annual review of psychology," in *In*, pp. 243–271, University, 1999.

[47] E. Peli, "Contrast sensitivity function and image discrimination," *J. Opt. Soc. Am. A*, vol. 18, no. 2, pp. 283–293, 2001.

Name: Osayamen Osaretin Imade Date of Degree: July, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: HUMAN VISION MODELS OF PERCEPTUAL IMAGE DISTORTIONS

Pages in Study: 64 Candidate for the Degree of Master of Science

Major Field: Electrical Engineering

Digital media's prevalence in today's society is placing an increasing strain on the technology to provide, transmit and store these contents. The demand for higher quality content in digital media has led to drastic increase in storage requirements over the past three decades. To meet the challenge of storing and securing digital media, this thesis proposes an insight into how the human vision system (HVS) can be characterized to determine thresholds of visibility of visual distortions. The use of these results can increase the amount of watermarking information applied to an image as well as applications in quantization error detection.

In this thesis, we first propose a framework for predicting the regions of natural images that visually disguise distortions created as a result of modification of wavelet domain coefficients. The visual error perception algorithm adaptively predicts the visual perceptibility threshold of spread spectrum watermarking added in the wavelet subband. Spatial statistical feature maps combined with ground truth data from psychophysical experiments enabled the generation of an activity scaling parameter that evaluates the masking thresholds of image regions. We also demonstrate the correlation with an image quality assessment algorithm to the detectability of distortions in an image. Subsequently the algorithm is combined with a compression scheme to yield compressed images of higher visual quality.

Secondly, we presents the results of another psychophysical experiment designed to investigate the effect of a scene's context on the detection of distortions presented in natural-image patches. Via a two-alternative forced-choice experiment, we measured thresholds for detecting 6.2 c/deg gabor target in image patches which were placed in various image surrounds (contexts), including various textures, a solid-gray background, and the patch's original context. The contexts were adjusted using histogram specification to control for differences in brightness, contrast, and other first-order statistical properties of the luminance distribution. Our results revealed that the context in which a patch is placed does indeed affect the ability to detect distortions in that patch. The findings suggest that characterization and implementation of a human visual system's ability to detect errors has potential in providing perceivable greater quality in image applications.

ADVISOR'S APPROVAL: Dr. Damon M. Chandler