

## Research Article

# A Patch-Based Structural Masking Model with an Application to Compression

Damon M. Chandler,<sup>1</sup> Matthew D. Gaubatz,<sup>2</sup> and Sheila S. Hemami<sup>3</sup>

<sup>1</sup> School of Electrical and Computer Engineering, Oklahoma State University, Stillwater, OK 74078, USA

<sup>2</sup> Print Production Automation Lab, HP Labs, Hewlett-Packard, Palo Alto, CA 94304, USA

<sup>3</sup> School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA

Correspondence should be addressed to Matthew D. Gaubatz, matthew.gaubatz@hp.com

Received 26 May 2008; Accepted 25 December 2008

Recommended by Simon Lucey

The ability of an image region to hide or *mask* a given target signal continues to play a key role in the design of numerous image processing and vision systems. However, current state-of-the-art models of visual masking have been optimized for artificial targets placed upon unnatural backgrounds. In this paper, we (1) measure the ability of natural-image patches in masking distortion; (2) analyze the performance of a widely accepted standard masking model in predicting these data; and (3) report optimal model parameters for different patch types (textures, structures, and edges). Our results reveal that the standard model of masking does not generalize across image type; rather, a proper model should be coupled with a classification scheme which can adapt the model parameters based on the type of content contained in local image patches. The utility of this adaptive approach is demonstrated via a spatially adaptive compression algorithm which employs patch-based classification. Despite the addition of extra side information and the high degree of spatial adaptivity, this approach yields an efficient wavelet compression strategy that can be combined with very accurate rate-control procedures.

Copyright © 2009 Damon M. Chandler et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Visual masking is a general term that refers to the perceptual phenomenon in which the presence of a masking signal (the *mask*) reduces a subject's ability to visually detect another signal (the *target* of detection). Masking is perhaps the single most commonly used property of the human visual system for image processing. It has found extensive use in image and video compression [1–4], digital watermarking [5–9], unequal error protection [10], quality assessment [11–14], image synthesis [15, 16], in the design of printers and variable-resolution displays [17, 18], and in several other areas (e.g., in image denoising [19] and camera projection [20]). For most of these applications, the original image serves as the mask, and the distortions induced via the processing (e.g., compression artifacts, rendering artifacts, or a watermark) or specific objects of interest (e.g., in object tracking) serve as the target of detection. Predicting an image's ability to mask a visual target is thus of great interest to system designers.

The amount of masking imposed by a particular image is determined by measuring a human subject's ability to detect a target in the presence of the mask. A psychophysical experiment of this type would commonly employ a forced-choice procedure in which two images are presented to a subject. One image contains just the mask (e.g., an original image patch), and the other image contains the mask+target (e.g., a distorted image patch). The subject is then asked to select which of the two images contains the target. If the subject chooses the correct image, the contrast of the target is reduced; otherwise, the contrast of the target is increased. This process is repeated until the contrast of the target is at the subject's threshold of detectability.

The above forced-choice paradigm is noteworthy because computational models of visual masking operate in similar fashion [3, 4, 21–24]. A computational model of masking would first compute modeled neural responses to the mask, then compute modeled neural responses to the mask+target, and then deem the target detectable if the two sets of neural responses sufficiently differ. The neural responses are

commonly modeled via three stages: (1) a frequency-based decomposition which models the initially linear responses of an array of visual neurons, (2) application of a pointwise nonlinearity to the transform coefficients and inhibition based on the values of other coefficients (*gain control* [24–27]), and (3) summation of these adjusted coefficients across space, spatial frequency, and orientation so as to arrive at a single scalar response value for each image. The first two stages, in effect, represent the two images (mask and mask+target) as points in a feature space, and the target is deemed visually detectable if the two points are sufficiently distant (as measured, in part, in the third stage).

Standard approaches to the frequency-based decomposition include a steerable pyramid [22], a Gaussian pyramid [11], an overcomplete wavelet decomposition [28], radial filters [29], and cortex filters [21, 24, 30]. The standard approach to the summation stage employs a  $p$ -norm, typically with  $p \in [1.5, 4]$ . The area of greatest debate lies in the implementation of the second stage which models the pointwise nonlinearity and the gain control mechanism provided by the inhibitory pool [24, 27, 31]. Let  $x(u_0, f_0, \theta_0)$  correspond to the transform coefficient at location  $u_0$ , center frequency  $f_0$ , and orientation  $\theta_0$ . In a standard gain-control-based masking model, the (nonlinear) response of a neuron tuned to these parameters,  $r(u_0, f_0, \theta_0)$ , is given by

$$r(u_0, f_0, \theta_0) = g \cdot \frac{(w(f_0, \theta_0)x(u_0, f_0, \theta_0))^p}{b^q + \sum_{(u, f, \theta) \in S} (w(f, \theta)x(u, f, \theta))^q}, \quad (1)$$

where  $g$  is a gain factor,  $w(f, \theta)$  represents a weight designed to take into account the human contrast sensitivity function,  $b$  represents a saturation constant,  $p$  provides the pointwise nonlinearity to the neuron,  $q$  provides the pointwise nonlinearity to the neurons in the inhibitory pool, and the set  $S$  indicates which other neurons are included in the inhibitory pool.

Although numerous studies have shown that the response of a neuron can be attenuated based on the responses of neighboring neurons (see, e.g., [26, 32]), the actual contributors to the inhibitory pool remain largely unknown. Accordingly, the specific parameters used in gain-control-based masking models are generally fit to experimental masking data. For example, model parameters have been optimized for detection thresholds measured using simple sinusoidal gratings [4], to filtered white noise [3], and to threshold-versus-contrast (TvC) curves of target Gabor patterns with sinusoidal maskers [22, 24]. Typically,  $p$  and  $q$  are in the range  $2 \leq q \leq p \leq 4$ , and the inhibitory pool consists of neural responses in the same spatial frequency band ( $f_0$ ), at orientations within  $\pm 45^\circ$  of  $\theta_0$ , and within a local spatial neighborhood (e.g., 8-connected neighbors). Indeed, this approach has proved quite successful at predicting detection thresholds for targets placed against relatively simplistic masks such as sinusoidal gratings, Gabor patches, or white noise.

Image-processing applications however, are concerned with the detectability of specific targets presented against naturalistic, structured backgrounds rather than white noise or other artificial masks. It remains an open question of whether the model parameters need to be adjusted for masks consisting of natural images, and if so, what are the proper adjustments? Because very few studies have measured masking data for natural-image stimuli [33–35], the optimal model parameters for natural-image masks have yet to be determined. Consequently, the majority of current algorithms simply use the aforementioned standard parameter values (optimized for simplistic masks). Although we have previously shown that the use of these standard model parameters can provide reasonable predictions of masking imposed by textures [6], most natural images contain a mix of *textures*, *structures*, and *edges*. We have observed that application of these standard model parameters to natural images often leads to overestimates of the ability of edges and object boundaries to mask distortions. This shortcoming is illustrated in Figure 1, which depicts an original image of a horse (a), that same image to which wavelet subband quantization distortions oriented at  $90^\circ$  have been added (b), and the top ten  $32 \times 32$  patches which contain the most visible distortions (c) as estimated by a standard model of masking ((1) with  $p = 2.4$ ,  $q = 2.3$ ,  $b = 0.03$ , and  $g = 0.025$ ; see Section 3 for further details of the model implementation). Notice from the middle image of Figure 1 that the distortions are most visible in the flat regions of sky around the horse's ears. Yet, the masking model overestimates the ability of these structured regions to mask the distortion.

To address this issue, the goals of this paper are threefold. (1) We present the results of a psychophysical experiment which provides masking data using natural-image patches; our results confirm the fact that edges and other structured regions provide less masking than textures. (2) Based on these masking data, we present model parameters which are optimized for natural image content (textures, structures, and edges) and are thus better suited for applications which process natural images. (3) We demonstrate the utility of this model for image processing via a specific application to image compression; a classification-based compression strategy is presented in which quantization step sizes are selected on a patch-by-patch basis as a function of the patch classification into a texture, structure, or edge, and then based upon our masking data. Despite the requirement of additional side information, the use of our image-type-specific masking data results in an average rate savings of 8%, and produces images that are preferred by 2/3 of tested viewers over a standard gain-control-based compression scheme.

This paper is organized as follows. Section 2 details the visual masking experiment and the results. In Section 3, we apply a standard gain-control model of masking to the experiment stimuli and describe how this model must be adjusted based on local image content. An application of image-content-adaptive masking to compression is presented in Section 4, and general conclusions are provided in Section 5.

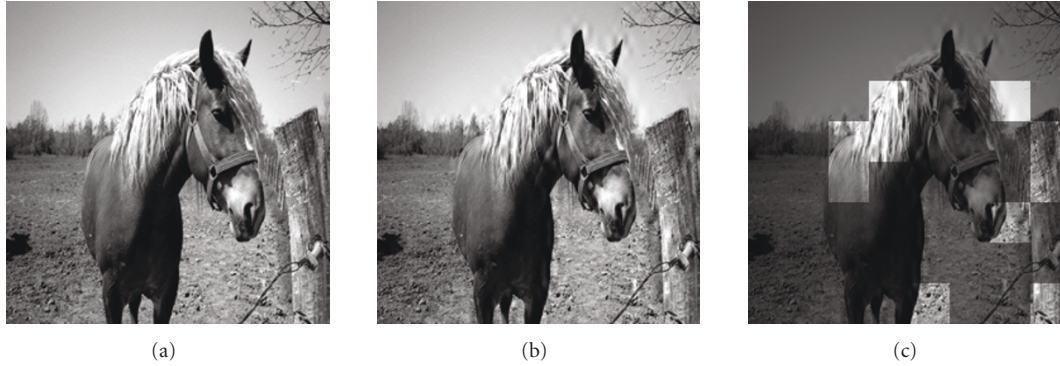


FIGURE 1: (a) Original  $256 \times 256$  image *horse*. (b) Distorted version of *horse* containing wavelet subband quantization distortions created via the following: (1) performing a three-level discrete wavelet transform of the original image using the 9/7 biorthogonal filters [36]; (2) quantizing the HL subband at the third decomposition level with a step size of 400; (3) performing an inverse discrete wavelet transform. (c) Highlighted regions correspond to the top ten  $32 \times 32$  patches containing the most visible distortions as deemed by the standard masking model; these blocks elicit the greatest difference in simulated neural responses between the original and distorted images (see Section 3 for details of the model implementation). Notice from (b) that the distortions are most visible in the regions above the horse's ears, whereas the masking model overestimates the ability of these regions in masking the distortion.

## 2. Visual Masking Experiment

In this work, a *texture* is defined as image content for which threshold elevation is reasonably well predicted by current masking models, and roughly matches our intuitive idea of what the term “texture” represents. An *edge* is one or more boundaries between homogeneous regions. A *structure* is neither an edge nor a texture, but which contains some recognizable organization.

To investigate the effects of patch contrast and type (texture, structure, edge) on the visibility of wavelet subband quantization distortions, a psychophysical detection experiment was performed. Various patches cropped from a variety of natural images served as backgrounds (masks) in this study. The patches were selected to contain either a texture, a structure, or an edge. We then measured the minimum contrast required for subjects to detect vertically oriented wavelet subband quantization distortions (targets) as a function of both the RMS contrast of each patch and the patch type.

We acknowledge that this division into three classes is somewhat restrictive and manufactured. Our motivation for using three categories stems primarily from our experience in applying masking models to image processing applications (compression [37–39], watermarking [6], and image and video quality assessment [13, 40]). We have consistently observed that the standard model of masking performs well on textures, but this same model always overestimates the masking ability of edges and other object boundaries. Thus, a logical first step toward extending the standard model of masking is to further investigate these two particular image types both psychophysically and computationally.

In addition, we have employed a third class, structures, to encompass regions which would not normally be considered an edge nor a texture. From a visual complexity standpoint, these are regions which are not as simple as an edge, but which are also less random (more structurally organized)

than a texture. We acknowledge that the structure class is broader than the other two classes, and that the term “structure” might not be the ideal label for all nonedge and nontexture patches. However, our motivation for using this additional class stems again from masking. For visual detection, we would expect structures to provide more masking than edges, but less masking than textures; thus, the structure class is a reasonable third choice to investigate. As discussed in this section, our psychophysical results confirm this rank ordering of masking ability. Furthermore, as we demonstrate in Section 4, improvements in visual quality can be achieved by modifying the standard model to take into account these three patch classes.

### 2.1. Methods

**2.1.1. Apparatus and Contrast Metric.** Stimuli were displayed on a high-resolution, Dell UltraScan P991 19-inch monitor at a display resolution of 28 pixels/cm. The display yielded minimum and maximum luminances of, respectively, 1.2 and 99.2  $\text{cd}/\text{m}^2$ , and an overall gamma of 2.3. Luminance measurements were made by using a Minolta LS-100 photometer (Minolta Corporation, Tokyo, Japan). The pixel-value-to-luminance response of this monitor was approximated via

$$L(X) = (0.7 + 0.026X)^{2.3}, \quad (2)$$

where  $L$  denotes the luminance in  $\text{cd}/\text{m}^2$ , and  $X$  denotes the 8 bit digital pixel value in the range 0–255. Stimuli were viewed binocularly through natural pupils in a darkened room at a distance of approximately 82 cm, resulting in a display visual resolution of 36.8 pixels/degree of visual angle [41].

Results are reported here in terms of RMS contrast [42], which is defined as the standard deviation of a pattern's luminances normalized by the mean luminance of the background upon which the pattern is displayed. RMS

contrast has been applied to a variety of stimuli, including noise [42], wavelets [43], and natural images [35, 44]. In this paper, results are reported in terms of the RMS contrast of the distortions (target) computed with respect to the mean luminance of the background-image patch (mask). Let  $\mathbf{I}$  and  $\hat{\mathbf{I}}$  denote an original and distorted image patch, respectively, and, let  $\mathbf{E} = \hat{\mathbf{I}} - \mathbf{I} + (1/N)\sum_{i=1}^N I_i$  denote the mean-offset distortions. The RMS contrast of the distortions,  $C$ , is given by

$$C = \frac{1}{\mu_{L(\mathbf{I})}} \left( \frac{1}{N} \sum_{i=1}^N [L(E_i) - \mu_{L(\mathbf{E})}]^2 \right)^{1/2}, \quad (3)$$

where  $\mu_{L(\mathbf{I})} = (1/N)\sum_{i=1}^N L(I_i)$  denotes the average luminance of  $\mathbf{I}$ , and  $N$  denotes the number of pixels in  $\mathbf{I}$ , and where  $\mu_{L(\mathbf{E})} = (1/N)\sum_{i=1}^N L(E_i)$  denotes the average luminance of the mean-offset distortions  $\mathbf{E}$ . The quantities  $L(I_i)$  and  $L(E_i)$  correspond to the luminance of the  $i$ th pixel of the image patch and the mean-offset distortions, respectively.

**2.1.2. Stimuli.** Stimuli used in this study consisted of image patches containing wavelet subband quantization distortions. Each stimulus was composed of a mask upon which a target was placed. The masks were  $64 \times 64$ -pixel image patches. The targets were wavelet subband quantization distortions.

**Masks.** The masks used in this study consisted of  $64 \times 64$ -pixel patches cropped from 8 bit grayscale images chosen from a database of high-resolution natural images. Fourteen  $64 \times 64$  masks were used, four of which were visually categorized as containing primarily texture, five of which were visually categorized as containing primarily structure, and five of which were visually categorized as containing primarily edges. Figure 2 depicts each mask along with its assigned common image name.

To investigate the effect of mask contrast on target detectability, the RMS contrast of each mask was adjusted via

$$\tilde{\mathbf{I}} = \alpha(\mathbf{I} - \mu_{\mathbf{I}}) + \mu_{\mathbf{I}}, \quad (4)$$

where  $\mathbf{I}$  and  $\tilde{\mathbf{I}}$  denote the original and contrast-adjusted images, respectively, where  $\mu_{\mathbf{I}} = (1/N)\sum_{i=1}^N I_i$  denotes the mean pixel value of  $\mathbf{I}$ , and where the scaling factor  $\alpha$  was chosen via bisection such that  $\tilde{\mathbf{I}}$  was at the desired RMS contrast. (The RMS contrast of each mask was computed by using (3) with  $L(I_i)$  and  $\mu_{L(\mathbf{I})}$  in place of, resp.,  $L(E_i)$  and  $\mu_{L(\mathbf{E})}$ .) RMS contrasts of 0.08, 0.16, 0.032, and 0.64 were used for all masks. To test the low-mask-contrast regime, two masks from each category were further adjusted to RMS contrasts of 0.01, 0.02, and 0.04 (images *fur* and *wood* from the texture category, images *baby* and *pumpkin* from the structure category, and images *butterfly* and *sail* from the edges category). Figures 3, 4, and 5 depict the adjusted-contrast textures, structures, and edges, respectively.

**Targets.** The visual targets consisted of distortions generated via quantization of a single wavelet subband. The subbands

were obtained by applying a discrete wavelet transform (DWT) to each  $64 \times 64$  patch using three decomposition levels and the 9/7 biorthogonal DWT filters (also used by Watson et al. [41], and by Ramos and Hemami [45], see also [35]). The distortions were generated via uniform scalar quantization of the HL3 subband (the subband at the third level of decomposition corresponding to vertically oriented wavelets). The quantizer step size was selected such that the RMS contrast of the resulting distortions was as requested by the adaptive staircase procedure (described in the following section). At the display visual resolution of 36.8 pixels/degree, the distortions corresponded to a center spatial frequency of 4.6 cycles/degree of visual angle.

Figures 6, 7, and 8 depict the masks from each category (texture, structure, and edge, resp.) along with each mask+target (distorted image). All masks in these figures have an RMS contrast of 0.32. All targets (distortions) are at an RMS contrast of 0.1. For illustrative purposes, the bottom row of each figure depicts just the targets placed upon a solid gray background set to the mean pixel value of each corresponding mask (i.e., the image patch has been replaced with its mean pixel value to facilitate viewing of just the distortions).

**2.1.3. Procedures.** Contrast thresholds for detecting the target (distortions) in the presence of each mask (patch) were measured by using a spatial two-alternative forced-choice procedure. On each trial, observers concurrently viewed two adjacent images placed upon a solid gray  $25 \text{ cd/m}^2$  background. One of the images contained the mask alone (nondistorted patch) and the other image contained the mask+target (distorted image patch). The image to which the target was added was randomly selected at the beginning of each trial. Observers indicated via keyboard input which of the two images contained the target. If the choice was incorrect (target undetectable), the contrast of the target was increased; if the choice was correct (target detectable), the contrast of the target was decreased. This process was repeated for 48 trials, whereupon the final target contrast was recorded as the subject's threshold of detection.

Contrast threshold was defined as the 75% correct point on a Weibull function, which was fitted to the data following each series of trials. Target contrasts were controlled via a QUEST staircase procedure [46] using software derived from the Psychophysics Toolbox [47, 48]. During each trial, an auditory tone indicated stimulus onset, and auditory feedback was provided upon an incorrect response. Response time was limited to within 7 seconds of stimulus onset.

The experiment was divided into 14 sessions, one session for each mask. Each session began with 3 minutes each of dark adaptation and adaptation to a uniform  $25 \text{ cd/m}^2$  display, which was then followed by a brief practice session. Before each series of trials, subjects were shown a high contrast, spatially randomized version of the distortions to minimize subjects' uncertainty in the target. Each subject performed the entire experiment two times; the thresholds reported in this paper represent the average of the two experimental runs.

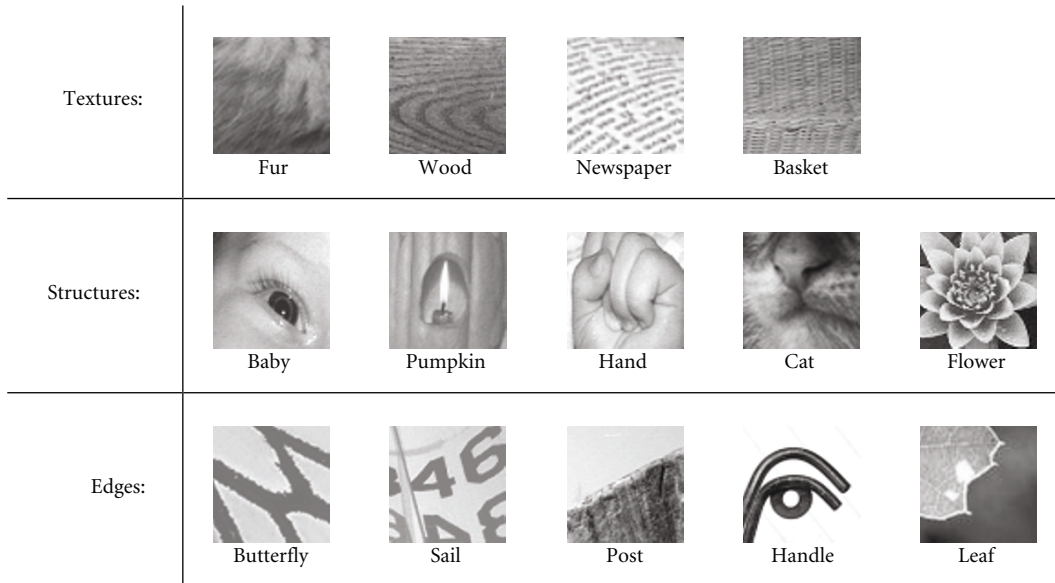


FIGURE 2: Image patches used as masks in the experiment. Textures: *fur, wood, newspaper, and basket*; structures: *baby, pumpkin, hand, cat, and flower*; edges: *butterfly, sail, post, handle, and leaf*.

Textures	$C = 0.64$	$C = 0.32$	$C = 0.16$	$C = 0.08$	$C = 0.04$	$C = 0.02$	$C = 0.01$
Fur							
Wood							
Newspaper					n/a	n/a	n/a
Basket					n/a	n/a	n/a

FIGURE 3: Contrast-adjusted versions of the textures used in the experiment. Note that only two images were tested in the very-low-contrast regime (RMS contrasts of 0.01, 0.02, and 0.04).

2.1.4. *Subjects.* Four adult subjects (including one of the authors) participated in the experiment. Three of the subjects were familiar with the purpose of the experiment; one of the subjects was naive to the purpose of the experiment. Subjects ranged in age from 20 to 30 years. All subjects had either normal or corrected-to-normal visual acuity.

2.2. *Masking Results and Analysis*

2.2.1. *Detection Thresholds as a Function of Mask Contrast.* The results of the experiment for two images of each type are shown in Figure 9 in the form of (TvC) curves in which masked detection thresholds are plotted as a function of

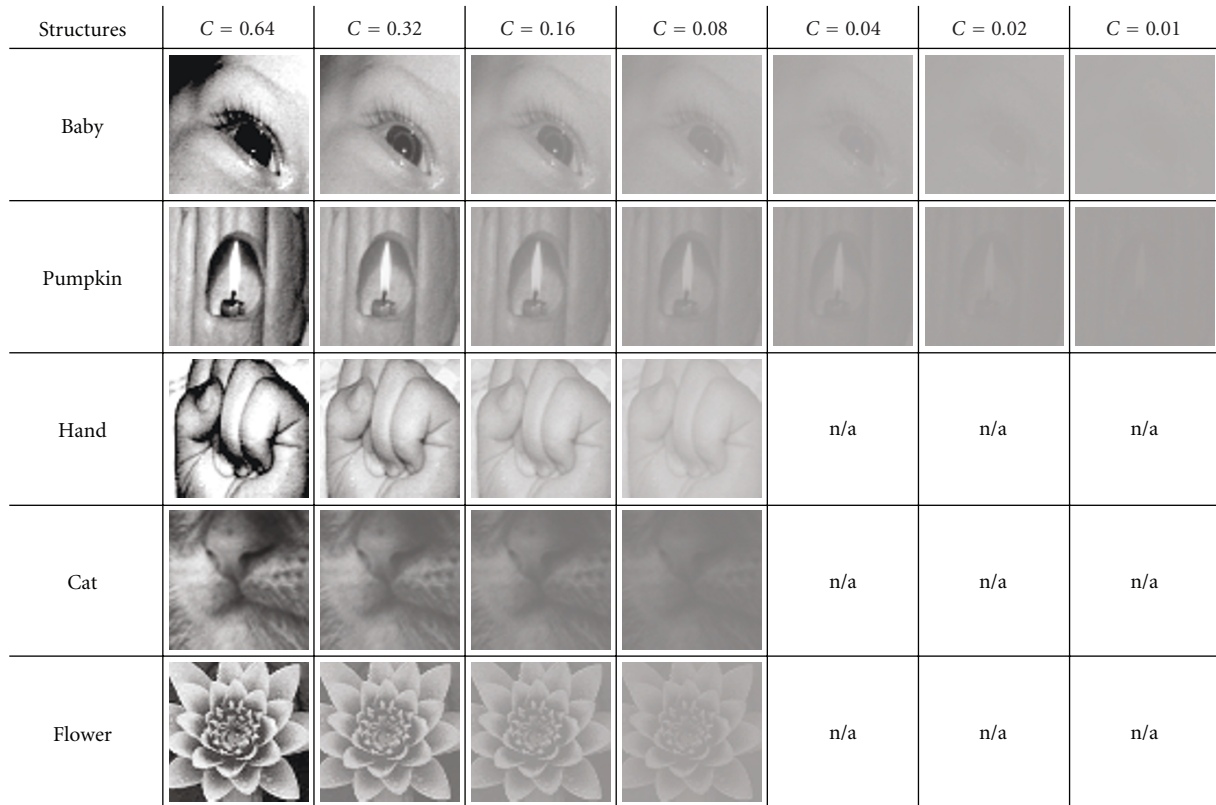


FIGURE 4: Contrast-adjusted versions of the structures used in the experiment. Note that only two images were tested in the very-low-contrast regime (RMS contrasts of 0.01, 0.02, and 0.04).

the contrast of the mask. Figure 9(a) depicts the average results for textures along with individual TvC curves for images *fur* and *wood*. Figure 9(b) depicts the average results for structures along with individual TvC curves for images *baby* and *pumpkin*. Figure 9(c) depicts the average results for edges along with individual TvC curves for images *sail* and *butterfly*. In each graph, the horizontal axis denotes the RMS contrast of the mask, and the vertical axis denotes the RMS contrast of the target. The dashed line in each graph corresponds to the average TvC curve computed for all masks of a given category (texture, structure, and edge). Data points in the individual TvC curves indicate contrast detection thresholds averaged over all subjects. Error bars denote standard deviations of the means over subjects (individual TvC curves) and over masks (average TvC curves).

As shown in Figure 9, for mask contrasts below 0.04, the thresholds for all three image types (edges, textures, and structures) are roughly the same. Average thresholds when the mask contrast was at the minimum contrast tested of 0.01 were as follows. The error measurement reported for each threshold (represented by a  $\pm$  sign) denotes one standard deviation of the mean over the tested images.

- (i) Textures:  $0.0080 \pm 0.0002$ ,
- (ii) Structures:  $0.0082 \pm 0.0001$ ,
- (iii) Edges:  $0.0089 \pm 0.0028$ .

Notice that the average thresholds for textures and structures are within each other's standard deviation as well within as the standard deviation of the average threshold for edges. These data therefore suggest that at very low mask contrasts, in the regime in which the mask is nearly undetectable and certainly visually unrecognizable, masking is perhaps due primarily to either noise masking or low-level gain-control mechanisms (e.g., inhibition amongst V1 simple cells) [24], and not due to higher-level visual processing.

As previous masking studies have shown, when the contrast of the mask increases, so does the contrast threshold for detecting a target placed upon that mask. Our results support this finding; in general, the greater the mask contrast, the greater the detection threshold. However, as shown in Figure 9, the TvC curves for the three categories demonstrate a marked divergence as the contrasts of the masks increase. Average thresholds when the mask contrast was 0.64 (the maximum contrast tested) were as follows.

- (i) Textures:  $0.1233 \pm 0.0384$ ,
- (ii) Structures:  $0.07459 \pm 0.0218$ ,
- (iii) Edges:  $0.0288 \pm 0.0120$ .

The large variation in elevations suggests that the effectiveness of a particular image patch at hiding distortion depends both on the contrast of the patch and on the content within the patch.

Edges	$C = 0.64$	$C = 0.32$	$C = 0.16$	$C = 0.08$	$C = 0.04$	$C = 0.02$	$C = 0.01$
Butterfly							
Sail							
Post					n/a	n/a	n/a
Handle					n/a	n/a	n/a
Leaf					n/a	n/a	n/a

FIGURE 5: Contrast-adjusted versions of the edges used in the experiment. Note that only two images were tested in the very-low-contrast regime (RMS contrasts of 0.01, 0.02, and 0.04).

Textures	Fur	Wood	Newspaper	Basket
Mask				
Mask+target				
Target				

FIGURE 6: Targets used in the experiment consisted of wavelet distortions generated by quantizing the HL3 DWT subband of each mask. Shown here are: Texture masks (original images) at an RMS contrast of 0.32. Masks+targets (distorted images) in which the distortions are at an RMS contrast of 0.1. Targets (distortions) shown against a gray background set to the mean pixel value of the corresponding mask (shown for illustrative purposes only). In the experiment, subjects were shown both the mask and mask+target and were asked to choose which of the two images contained the target (distortions); the contrast of the target was decreased (or increased) if the subject's choice was correct (or incorrect, resp.). This process was repeated for 48 trials, whereupon the final target contrast was recorded as the subject's threshold of detection.

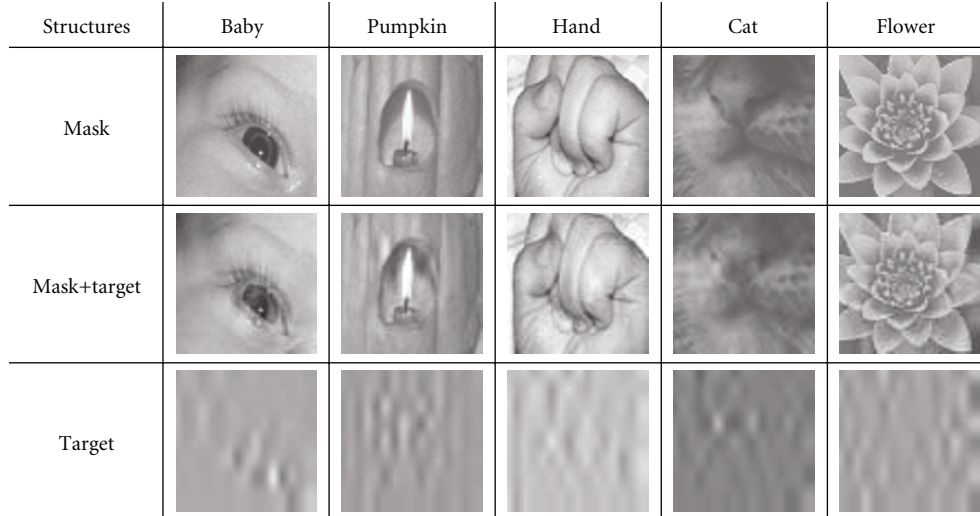


FIGURE 7: Structure masks (original images) at an RMS contrast of 0.32. Masks+targets (distorted images) in which the distortions are at an RMS contrast of 0.1. Targets (distortions) shown against a gray background set to the mean pixel value of the corresponding mask (shown for illustrative purposes only).

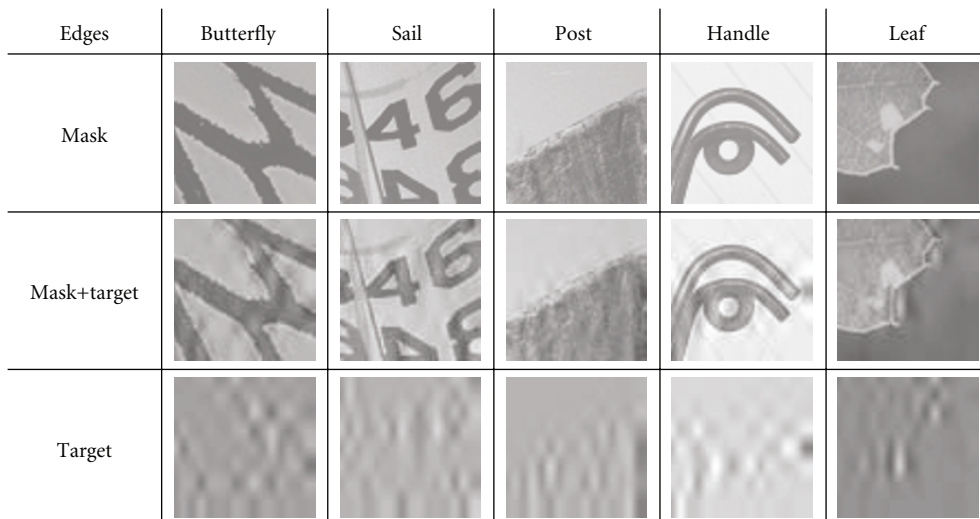


FIGURE 8: Edge masks (original images) from each category at an RMS contrast of 0.32. Masks+targets (distorted images) in which the distortions are at an RMS contrast of 0.1. Targets (distortions) shown against a gray background set to the mean pixel value of the corresponding mask (shown for illustrative purposes only).

### 2.2.2. Detection Thresholds as a Function of Mask Category.

The influence of patch content (mask category) on detection thresholds is further illustrated in Figures 10 and 11, which depict relative threshold elevations defined as

$$\text{relative threshold elevation} = \frac{CT}{\overline{CT}_{\text{edge}}}, \quad (5)$$

where  $CT$  denotes the contrast detection threshold for a given mask contrast, and  $\overline{CT}_{\text{edge}}$  denotes the contrast detection threshold averaged over all edges of the same contrast. Thus, the relative threshold elevation provides a measure of the extent to which a given mask increases thresholds (for elevations  $>1.0$ ) or decreases thresholds (elevations  $<1.0$ ) relative to an edge of the same contrast. The

relative threshold elevation was computed separately for each subject and each mask.

Figure 10 depicts relative threshold elevations, averaged over all subjects and all images of each category, plotted as a function of mask contrast. Observe that at low mask contrasts (0.01–0.04), relative threshold elevations are largely independent of category, that is, on average, low-contrast edges are equally as effective as low-contrast textures and structures at masking distortions. However for higher mask contrasts (0.16–0.64), relative threshold elevations are indeed category-specific. In general, as the contrast of the mask increases, textures exhibit progressively greater masking than structures, and structures exhibit progressively greater masking than edges.



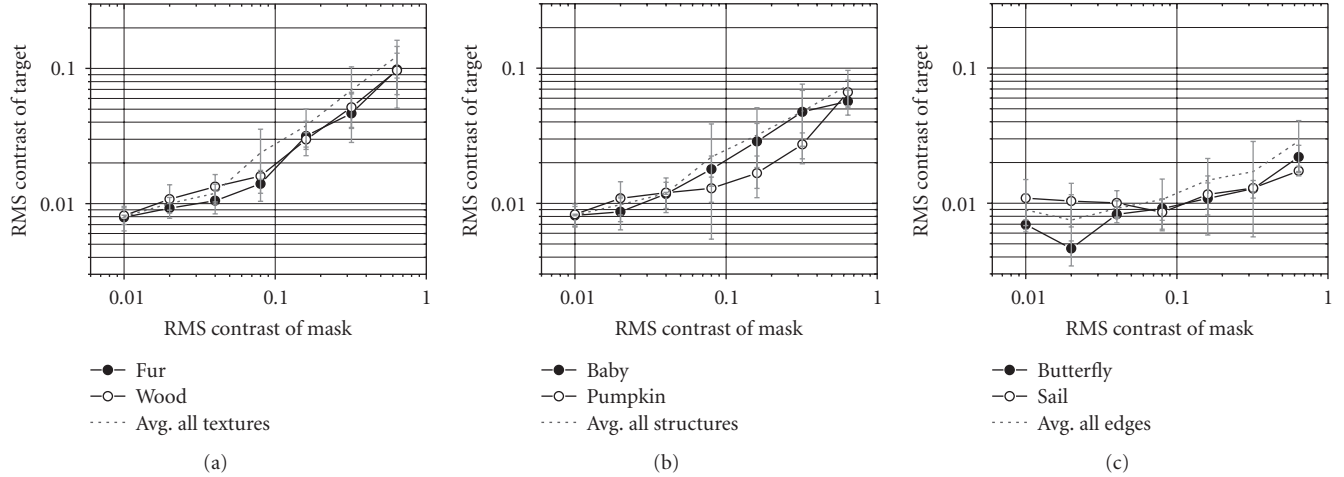


FIGURE 9: Threshold-versus-contrast (TvC) curves obtained from the masking experiment. (a) Average TvC curve for textures (dashed line) and individual TvC curves for *fur* and *wood*. (b) Average TvC curve for structures and individual TvC curves for *baby* and *pumpkin*. (c) Average TvC curves for edges and individual TvC curves for *butterfly* and *sail*. In each graph, the horizontal and vertical axes correspond to the RMS contrast of the mask (image) and the RMS contrast of the target (distortion), respectively. Data points in the individual TvC curves indicate contrast detection thresholds averaged over all subjects. Error bars denote standard deviations of the means.

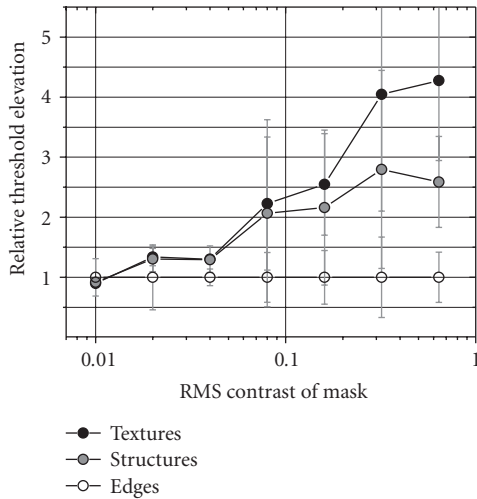


FIGURE 10: Average relative threshold elevations for each mask category plotted against mask contrast. For increasingly greater mask contrasts, textures and structures demonstrate increasingly greater threshold elevations over edges at the same contrast.

Figure 11 depicts relative threshold elevations for the 0.32 and 0.64 contrast conditions plotted for each of the 14 images. The data points denote relative threshold elevations averaged over all subjects; error bars denote standard deviations of the means over subjects. The dashed lines denote average relative threshold elevations for each of the three image types. The images depicted on the horizontal axis have been ordered by eye to represent a general transition from simplistic edge to complex texture (from left to right). Indeed, notice that the data generally demonstrate a corresponding left-to-right increase in relative threshold elevation.

Thus, on average, high-contrast (0.32–0.64) textures elevate detection thresholds approximately 4.3 times greater than high-contrast edges, and high-contrast structures elevate thresholds approximately 2.6 times greater than high-contrast edges. We call this effect *structural masking* which attributes elevations in threshold to the structural content (texture, structure, and edge) of the mask. (We are currently investigating the relationship between structural masking and *entropy masking* [49]. Entropy masking attributes elevations in thresholds to a subject’s unfamiliarity with a mask. A computational model of entropy masking has yet to be developed.) These findings demonstrate that a proper measure of masking should account both for mask contrast and for mask *type*. In the following section, we use these masking data to compute optimal mask-type-specific parameters for use in a gain-control-based masking model.

### 3. Fitting a Gain-Control Model to Natural Images

The standard gain-control model, which has served as a cornerstone in current understanding of the nonlinear response properties of early visual neurons, has proved quite successful at predicting thresholds for detecting targets placed against relatively simplistic masks. However, gain-control models do not explicitly account for image content; rather, they employ a relatively oblivious inhibitory pool which imposes largely the same inhibition regardless of whether the mask is a texture, structure, or edge. Such a strategy is feasible for low-contrast masks, but, as demonstrated by our experimental results, high-contrast textures, structures, and edges impose significantly different elevations in thresholds (i.e., structural masking is observed).

In this section, we apply a computational model of gain control to the masking data from the previous section and

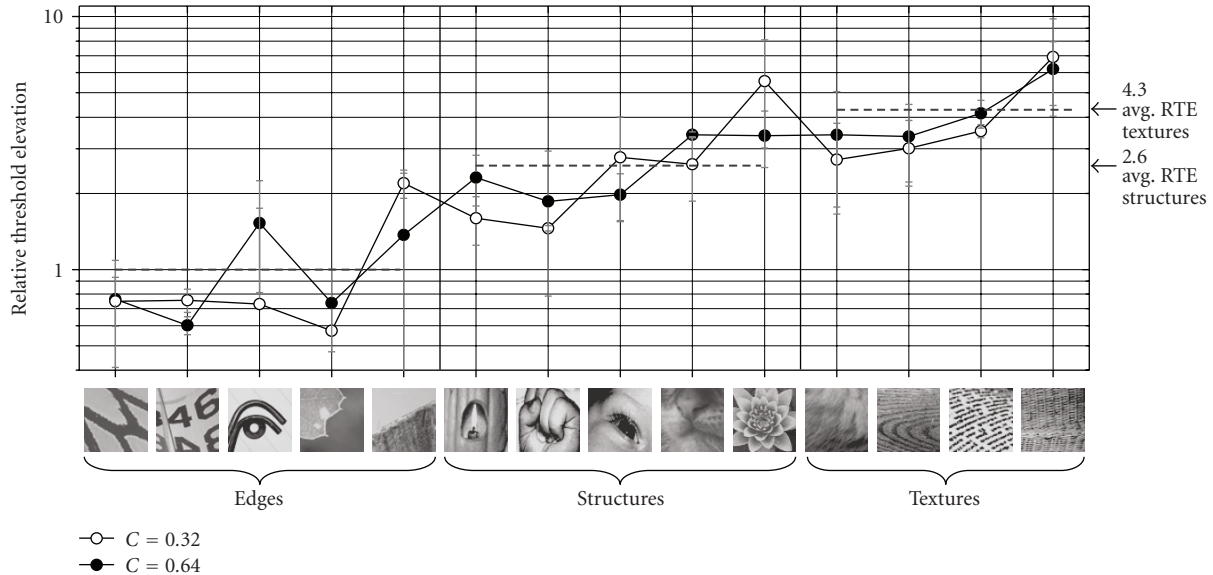


FIGURE 11: Relative threshold elevations averaged over all subjects for each of the 14 masks at contrasts of 0.32 and 0.64.

report the optimal model parameters. We demonstrate that when the model is implemented with standard parameter values, the model can perform well in predicting thresholds for textures. However, these same parameters lead to over-estimates of the amount of masking provided by edges and structures. Here, we report optimal model parameters for different patch types (textures, structures, and edges) which provide a better fit to the masking data than that achieved by using standard parameter values.

**3.1. A Discussion of Gain Control Models.** The standard model of gain control described in Section 1 contains many parameters that must be set. However, we emphasize that this model is used extensively in the visual neuroscience community to mimic an underlying physical model consisting of an array of visual neurons. This neurophysiological underpinning limits the choice of model parameters to those which are biologically plausible. Here, before discussing the details of our model implementation, we provide general details regarding the selection of these parameters. For convenience, we repeat (1) as follows:

$$r(u_0, f_0, \theta_0) = g \cdot \frac{(w(f_0, \theta_0)x(u_0, f_0, \theta_0))^p}{b^q + \sum_{(u, f, \theta) \in S} (w(f, \theta)x(u, f, \theta))^q}. \quad (6)$$

As mentioned previously, this gain-control equation models a nonlinear neural response, which is implemented via a ratio of responses designed to mimic neural inhibition observed in V1 (so-called *divisive normalization*). The numerator models the excitatory response of a single neuron, and the denominator models the inhibitory responses of the neurons which impose the normalization.

**3.1.1. The Input Gain  $w(f, \theta)$  and Output Gain  $g$ .** The parameters  $w(f, \theta)$  and  $g$  model the input and output gain of each neuron, respectively. The input gain  $w(f, \theta)$  is

designed to account for the variation in neural sensitivity to different spatial frequencies and orientations. These gains are typically chosen to match the human contrast sensitivity function derived based on detection thresholds measured for unmasked sine-wave gratings (e.g., [21]). Others have selected the gains to match unmasked detection thresholds for Gabor or wavelet targets, which are believed to better probe the sensitivities of visual neurons (e.g., [24]). We have followed this latter approach.

The output gain  $g$  can be viewed as the sensitivity of the neuron following divisive normalization. This parameter is typically left as a free parameter that is adjusted to match TvC curves. We too have left  $g$  as a free parameter.

**3.1.2. The Excitatory and Inhibitory Exponents  $p$  and  $q$ , and the Semisaturation Constant  $b$ .** The parameters  $p$  and  $q$ , when used with (1), are designed to account for the fact that visual neurons exhibit a nonlinear response to contrast. A neuron's response increases with increasing stimulus contrast, but this response begins to level off at higher contrasts. In [22],  $p$  and  $q$  were fixed at the same value of  $p = q = 2$ . (Indeed, in terms of biological plausibility, using the same value for  $p$  and  $q$  is logical.) However, as noted by Watson and Solomon [24], this setting of  $p = q$  leads to premature response saturation in the model. In both [24, 26], this side effect is avoided by selecting separate values for  $p$  and  $q$ , with the condition that  $p > q$  to prevent an eventual decrease in response at high contrast. Typically,  $p$  and  $q$  are in the range  $2 \leq q \leq p \leq 4$ . Most often, either  $p$  or  $q$  is fixed, and other is left as a free parameter. We have followed this latter approach ( $p$  fixed,  $q$  free).

The parameter  $b$  is used to prevent division by zero (and thus an infinite response) in the absence of masking. In [24],  $b$  was allowed to vary, which resulted in optimal values between 0.02 and 0.08. We report the results of using both  $b$

fixed and  $b$  free, each of which leads to an optimal value near 0.03 which is well within the range reported in [24].

**3.2. Model Implementation.** As mentioned in Section 1, computational models of gain control typically employ three stages: (1) a frequency-based decomposition which models the initially linear responses of an array of visual neurons, (2) computation of nonlinear neural responses and inhibition, and (3) summation of modeled neural responses. The individual neural responses were modeled by using (1) with specific details of the model as described in what follows.

The initially linear responses of the neurons ( $x(f, u, \theta)$  in (1)) were modeled by using a steerable pyramid decomposition with four orientations,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$ , and three levels of decomposition performed on the original and distorted images. This decomposition was applied to the luminance values of each image computed via (2). The CSF weights  $w(f, \theta)$  were set to values of 0.068, 0.266, and 0.631 for bands from the first through third decomposition levels, respectively, and the same weights were used for all four orientations. These CSF weights were selected based on our previous study utilizing wavelet subband quantization distortions presented against a gray background (i.e., in the absence of a mask) [35].

The inhibitory pool consisted of those neural responses at orientations within  $\pm 45^\circ$  of the orientation of the responding neuron and within the same frequency band as the responding neuron. Following from [13] (see also [24]), a Gaussian-weighted summation over space, implemented via convolution, was used for the inhibitory pool. Specifically, the spatial extent of the inhibitory pooling was selected to be a  $3 \times 3$  window (8 connected neighbors) in which the contribution of each neighbor was determined by the taps of a  $3 \times 3$  Gaussian filter created via the outer product of one-dimensional filters with impulse response  $[1/6, 2/3, 1/6]$ .

To determine if a target is at the threshold of detection, the modeled neural responses to the mask are compared with the modeled neural responses to the mask+target. Let  $\{r_{\text{mask}}(u, f, \theta)\}$  and  $\{r_{\text{mask+target}}(u, f, \theta)\}$  denote the sets of modeled neural responses computed via (1) for the mask and mask+target, respectively. The model deems the target detectable if  $\{r_{\text{mask}}(u, f, \theta)\}$  and  $\{r_{\text{mask+target}}(u, f, \theta)\}$  sufficiently differ as measured via

$$d = \left( \sum_u \left( \sum_\theta \left( \sum_f |r_{\text{mask}}(u, f, \theta) - r_{\text{mask+target}}(u, f, \theta)|^{\beta_f} \right)^{\beta_\theta/\beta_f} \right)^{\beta_u/\beta_\theta} \right)^{1/\beta_u}, \quad (7)$$

with  $\beta_\theta = \beta_f = 1.5$  and  $\beta_u = 2$ ; the value of  $\beta_\theta = \beta_f = 1.5$  was selected based on our previous study on summation of responses to wavelet subband quantization distortions [35]. The model predicts the target to be at the threshold of detection when  $d$  reaches some chosen critical value, typically  $d = 1$  [6, 24] which is also used here.

To use the model to predict contrast detection thresholds, a search procedure is used in which the contrast of the target

is iteratively adjusted until  $d = 1$ . Here, for targets consisting of wavelet subband quantization distortions, we have used the following bisection search.

- (1) Compute the responses to the original image:  $\{r_{\text{mask}}(u, f, \theta)\}$ .
- (2) Generate baseline distortions  $\mathbf{e} = \hat{\mathbf{I}} - \mathbf{I}$ , where  $\mathbf{I}$  denotes the original image, and  $\hat{\mathbf{I}}$  denotes the distorted image created by quantizing the HL3 DWT subband with a step size of 100.
- (3) Let  $v_{\text{lo}} = 0$  and  $v_{\text{hi}} = 50$ .
- (4) Compute  $v = (1/2)(v_{\text{lo}} + v_{\text{hi}})$ .
- (5) Adjust the contrast of the distortions contained in the distorted image via  $\hat{\mathbf{I}} = \mathbf{I} + v\mathbf{e}$ .
- (6) Compute the responses to the distorted image from step (5):  $\{r_{\text{mask+target}}(u, f, \theta)\}$ .
- (7) Compute  $d$  via (7).
- (8) If  $d \approx 1$ , then exit.
- (9) If  $d > 1$  (i.e.,  $v$  is too large), then let  $v_{\text{hi}} = v$ , and then go to step (4).
- (10) If  $d < 1$  (i.e.,  $v$  is too small), then let  $v_{\text{lo}} = v$ , and then go to step (4).

This procedure typically converges in less than 10 iterations, whereupon the contrast of the distortions which elicited  $d \approx 1$  is taken to be the contrast detection threshold. The contrast of the distortions is measured via (3).

**3.3. Optimal Parameters and Model Predictions.** The parameters which are typically adjusted in a gain-control model are  $p$ ,  $q$ ,  $b$ , and  $g$ . Others have reported that the specific values of  $p$  and  $q$  have less effect on model performance than the difference between these parameters; one of these parameters is therefore commonly fixed. Here, we have used a fixed value of  $p = 2.4$ . Similarly, the value of  $b$  is often fixed based on the input range of the image data; we have used a fixed value of  $b = 0.035$ .

The free parameters,  $q$  and  $g$ , were chosen via an optimization procedure to provide the best fit to the TvC curves for each of the separate image types (Figure 9). Specifically,  $q$  and  $g$  were selected via a Nelder-Mead search [50] to minimize the standard-deviation-weighted cost function

$$J(\mathbf{C}_t, \hat{\mathbf{C}}_t) = \sum_i \frac{(\log C_{t,i} - \log \hat{C}_{t,i})^2}{\sigma_{t,i}}. \quad (8)$$

Here,  $\mathbf{C}_t$  denotes the vector of contrast detection thresholds measured for images of type  $t$ , and  $C_{t,i}$  denotes its  $i$ th element (the threshold measured for the  $i$ th contrast-adjusted mask of type  $t$ ).  $\hat{\mathbf{C}}_t$  denotes the vector of contrast thresholds predicted for those images by the model, and  $\hat{C}_{t,i}$  denotes its  $i$ th element. The value  $\sigma_{t,i}$  denotes the standard deviation of  $C_{t,i}$  across subjects. The optimization was performed separately for textures, structures, and edges.

TABLE 1: Model parameters, correlation coefficients, and RMS errors. Parameter values in italics were fixed, while others were optimized. Correlations coefficients and RMS errors shown in boldface denote maximum correlations and minimum errors for each column.

		Textures	Structures	Edges
Parameters	$p$	<i>2.4</i>	<i>2.4</i>	<i>2.4</i>
	$q$	2.32	1.94	1.72
	$b$	<i>0.035</i>	<i>0.035</i>	<i>0.035</i>
	$g$	0.020	0.040	0.091
Corr. coeff.	w/texture params	<b>0.943</b>	0.892	0.556
	w/structure params	0.891	<b>0.912</b>	<b>0.586</b>
	w/edge params	0.674	0.889	0.577
RMS error	w/texture params	<b>0.147</b>	0.213	0.485
	w/structure params	0.296	<b>0.176</b>	0.304
	w/edge params	0.542	0.444	<b>0.215</b>

Table 1 lists the resulting model parameters and the correlation coefficients and RMS errors between the experimental and predicted (log) thresholds using each set of parameters. Notice from these data that the optimal model parameters vary substantially for the three image types. The value of  $p = 2.4$  (fixed) and  $q = 2.3$  (optimized for textures) are well within the range of what is considered standard for a gain-control model. However, as evidenced by the decrease in correlation and increase in RMS error when using the incorrect parameters, the parameter values do not generalize across image type. For example, when the texture-optimal parameters are applied to edges, or when the edge-optimal parameters are applied to structures, the resulting RMS error is more than twice the RMS error in prediction performance compared to that achieved by using the correct optimal parameters. Similarly, applying the edge-optimal parameters to textures results in nearly four times the RMS error in prediction performance compared to that achieved by using the texture-optimal parameters.

This latter assertion is demonstrated in Figures 12, 13, and 14, which depict model predictions for textures, structures, and edges, respectively, using each set of optimal model parameters. Notice that, as expected, the standard (texture) model performs quite well for textures; in general, the model predictions are within the error bars. Using these same parameters yields reasonably accurate results for structures *baby*, *cat*, and *flower*; however, the predicted thresholds for *pumpkin* and *hand* are clearly overestimated in the high-contrast regime. For edges, the standard (texture) model severely overestimates thresholds in the high-contrast regime.

We have experimented with various values of  $p$  and  $b$ , with generally little effect on model performance if these parameters are set to reasonable (biologically plausible) values. If  $p$  is adjusted, the optimal value of  $q$  changes so as to maintain the quantity  $p-q$ . Similarly, an optimization in which  $b$  is allowed to vary along with  $q$  and  $g$  results in  $b = 0.035$ ,  $0.034$ , and  $0.032$  for textures, structures, and edges, respectively, and an insignificant change in  $q$  and  $g$  from the values listed in Table 1. Using these optimal values of  $b$  as opposed to a fixed value of  $b = 0.035$  results in a negligible effect on model performance for structures and edges. We

have also experimented with a joint optimization of  $p$ ,  $q$ ,  $b$ , and  $g$ , which results in values similar to those listed in Table 1 (while absolute values of  $p$  and  $q$  will vary depending on the initial values of these parameters, the quantity  $p-q$  remains consistent).

The model presented here represents an improved version of a model which we previously proposed in [51]. The primary improvement of our current model over the previous model is that the current model better represents a standard masking model. In particular, our model in [51] took as input pixel-value images as opposed to luminance-value images. The choice of using luminance values versus pixel values is still debatable, since pixel values more closely correlate with perceived brightness (“lightness” or  $L^*$ ) [52]. However, the use of luminance is more common in neural modeling, and thus luminance-value images are used in the current model. The model in [51] also used a broad inhibitory pool which consisted of neural responses at all orientations, whereas standard models typically limit the inhibitory pooling over orientation to within  $\pm 45^\circ$  of the orientation of the responding neuron [3, 4, 24], as is used in the current model. In terms of parameter values, our previous model used  $p = 2.4$ , as is used in the current model; however, in [51] two different values of  $q$  were used:  $q = 2.75$  for textures and  $q = 2.45$  for structures and edges; these latter values are nonstandard, since most models, including our current model, employ  $p > q$ . In [51],  $b$  and  $g$  were fixed at  $b = 0.03$  and  $g = 1$ , and a new parameter  $g_m$  (which represented an inhibition modulation term) was employed. In our current model, better fits are achieved without using  $g_m$ , and instead allowing both  $g$  and  $q$  to vary for each image type.

Although the use of model parameters optimized for separate image types does not achieve a perfect fit to the data, the thresholds are rarely overestimated for structures and edges as they are when using standard model parameters. These results indicate that when applying a masking model in image-processing applications, one must take into account the *type* of content contained in local image regions and select the model parameters accordingly (e.g., according to Table 1). Without this adjustment, visible distortions will emerge in regions containing edges and/or structures. In the

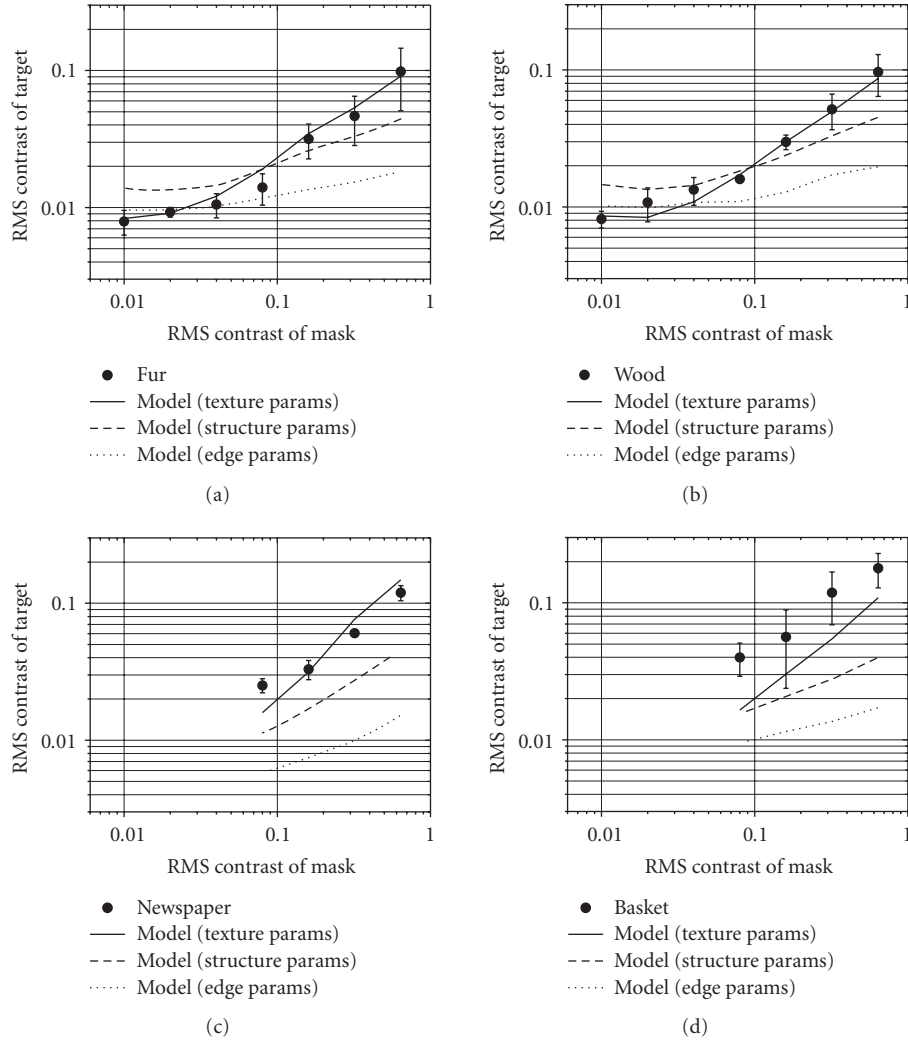


FIGURE 12: Predictions of the gain-control model for textures using parameters optimized for textures (solid line), for structures (dashed line), or for edges (dotted line). The solid circles denote the masking results obtained for each of the textures in the psychophysical experiment.

following section, we demonstrate how adapting a masking model to local image patches can lead to improved results for one common application of masking which is image compression.

#### 4. Application of the Structural Masking Model to Compression

A wavelet-based image coder that leverages the results of the previous masking study is described in this section. To fully take advantage of the study outlined in Section 2, it is necessary to classify each image patch into one of the three mask types (textures, structures, or edges). Quantization step sizes associated with each patch are derived based on its type and contrast, which jointly make up side information needed by the coder. Explicit as well as implicit compression techniques are applied in order to control the size of a compressed file containing this side information. A block

diagram of the proposed coder is illustrated in Figure 15. Rate-control techniques used to achieve arbitrary target coded rates are also discussed.

*4.1. Segmentation and Classification.* The first phase in the compression algorithm is to perform analysis on an input image in the spatial domain. The image must first be segmented into patches. Then, the contrast (see (3)) and type of each patch must be determined. These data are used to compute detection thresholds for each patch of image data, which in turn are used to determine quantization step sizes. Therefore, this information must be available at a decoder in order to make the compression process invertible. Any segmentation and classification schemes can be used for this stage. Though more sophisticated algorithms will yield more accurate step sizes and the best performance with respect to image quality, there is a tradeoff between the granularity of the segmentation scheme and the amount of

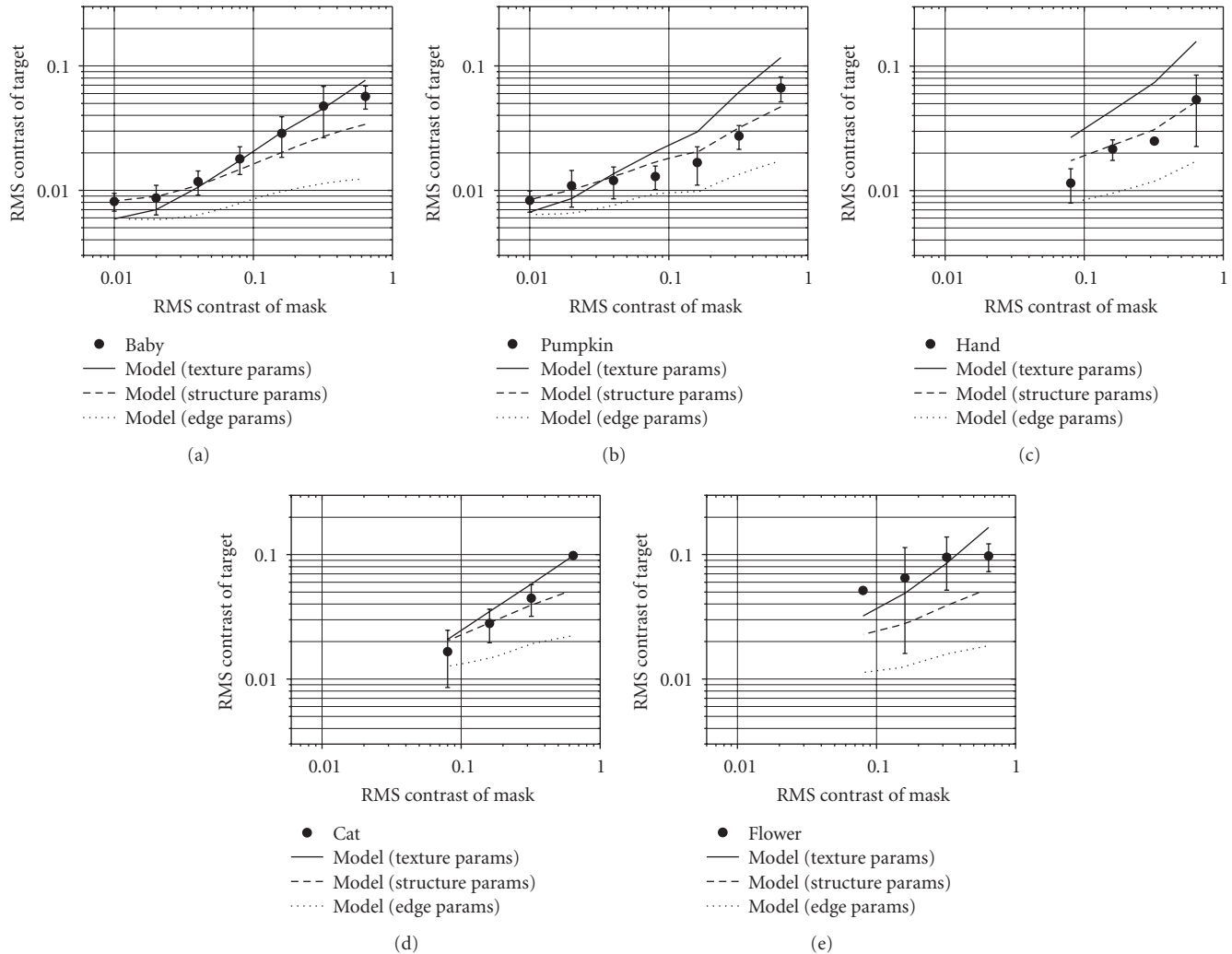


FIGURE 13: Predictions of the gain-control model for structures using parameters optimized for textures (solid line), for structures (dashed line), or for edges (dotted line). The solid circles denote the masking results obtained for each of the structures in the psychophysical experiment.

side information that must be coded in order to invert the compression process. Thus, a relatively simple rectangular-patch-based approach is used herein that associates a contrast and a type to each  $M \times M$  image patch. This choice is made for convenience and efficiency; mask types and mask contrast values are determined by local, disjoint computations. In order to compensate for the lack of granularity in the segmentation procedure, the contrast of each patch is set to the minimum contrast of all subregions of the patch.

A multistage classifier used to designate each patch as texture, structure, or edge data is described in more detail. A greedy approach to classification is presented. Prior to employing the algorithm, all images patches are labeled as unclassified. During each stage, the classifier selects patches that are associated with only one specific type of mask (texture, structure, or edge content). Only unclassified patches are processed in subsequent stages. Let  $m$  index each patch in the image, let  $\sigma_m^2$  denote its variance, and let  $\sigma_{m,j}^2$  denote the variance of the  $j$ th subpatch of patch  $m$ .

(1) The first stage classifies certain patches as edges or textures based on the patch variance. Specifically,

- (i) if  $\sigma_m^2 > K_{\text{edge}} \times (\text{no. of subpatches such that } \sigma_{m,j}^2 > T_M)$ , patch  $m$  is labeled an edge, or
- (ii) if  $\sigma_m^2 > K_{\text{texture}} \times (\text{no. of subpatches such that } \sigma_{m,j}^2 > T_M)$ , patch  $m$  is labeled a texture,

where  $K_{\text{edge}}$ ,  $K_{\text{texture}}$  and  $T_M$  are constants.

(2) The second stage is composed of three substages, each of which classifies a patch as an edge if a per-patch metric exceeds a threshold similar in form to those used in the first stage. The metrics, in the order in which they are applied, are

- (i) average Canny edge-detector output,
- (ii) average Laplacian of Gaussian detector output,
- (iii) variance of  $\sigma_{m,j}^2 / \text{mean of } \sigma_{m,j}^2$ .

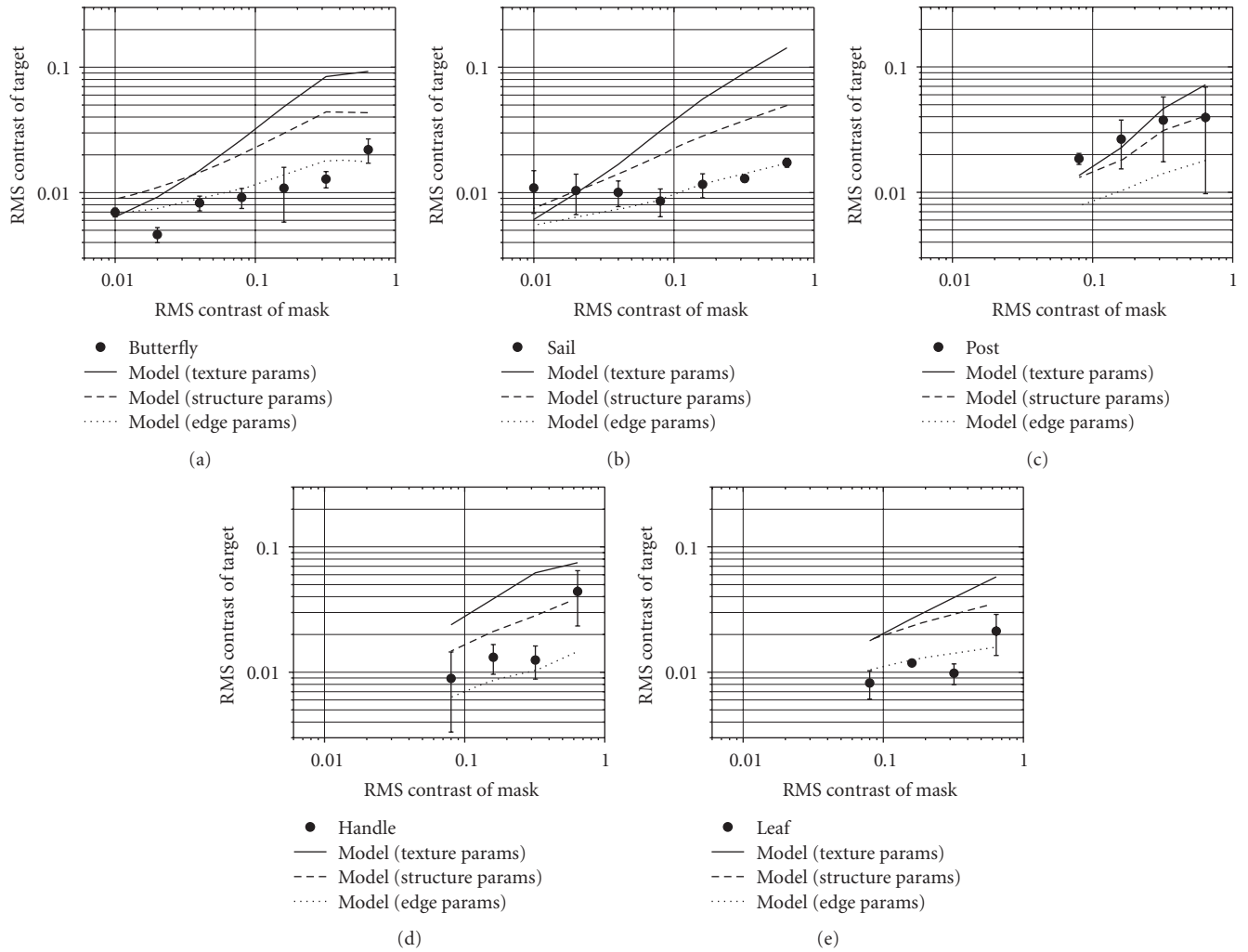


FIGURE 14: Predictions of the gain-control model for edges using parameters optimized for textures (solid line), for structures (dashed line), or for edges (dotted line). The solid circles denote the masking results obtained for each of the edges in the psychophysical experiment.

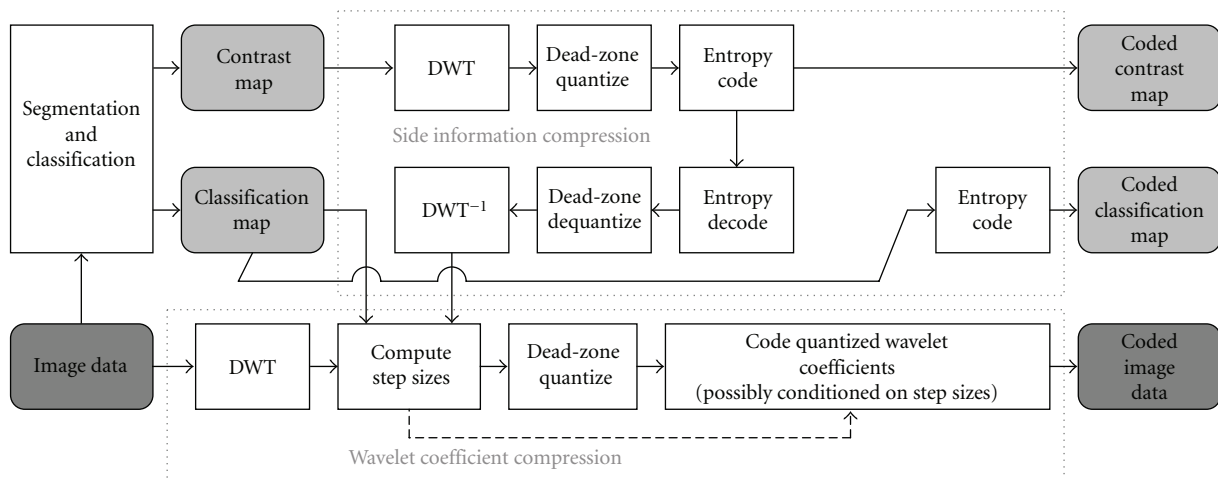


FIGURE 15: Block diagram representing the proposed coder. The major components perform segmentation/classification, side information coding and wavelet coefficient coding. If coding quantized wavelet coefficients conditioned on step size values, more compression gain is achieved through a nonembedded coding strategy (see Table 3).

TABLE 2: Confusion matrix for the example patch type classifier.

		Predicted patch type		
		Texture	Structure	Edge
Actual patch type	Texture	3450	2137	766
	Structure	184	1043	1038
	Edge	9	380	1233

- (3) The third stage labels all unclassified patches as structures.

This classifier was tuned without using any images in the test set. Performance of this classifier is evaluated in the following way. Ten of the test images were hand-annotated by three expert viewers, classifying each  $16 \times 16$  block of pixels as a patch of texture, structure, or edge information. This data constitutes a ground truth that can be compared with the classifier. The confusion matrix for the example classifier described above is given in Table 2. In general, the proposed algorithm predicts the correct class in over half the trials. The performance of the classifier can certainly be improved. The main goal of the example coder described herein, however, is to illustrate some of the gains achievable via type-based classification, and the classifier provides enough functionality to accomplish this goal.

After the segmentation and classification phase is complete, each patch in the image is associated with a contrast value computed via (3), and a mask type. The set of all patch contrasts, denoted  $C_{\text{rms}}(m)$ , is referred to as a *contrast map*, and the set of all patch types, denoted  $\text{type}(m)$ , is referred to as a *classification map*. Examples of these quantities are illustrated in Figure 16. Samples in the contrast map are from a continuum of values, and elements in the classification map are from a discrete set. Clearly, there is a tradeoff between the patch size and the amount of auxiliary information produced by this stage. Methods for representing these data efficiently are therefore key to the success of utilizing this information.

**4.2. Explicit Side Information Coding.** In order to generate the step sizes at a decoder, the contrast and classification maps must be conveyed as side information. To improve the efficiency of this maneuver, both maps are compressed. The contrast map is compressed in a lossy fashion, and the classification map is losslessly compressed. It is important to perform this compression prior to determining quantization step sizes. Otherwise, the decoder will incorrectly estimate these values during dequantization. The classification map can be represented efficiently with an arithmetic entropy coder. Because the contrast map contains relatively low-frequency information (e.g., see Figure 16), a standard wavelet transform coding framework can be used to represent this information economically. A five-level two-dimensional DWT is applied to the contrast map, followed by dead-zone quantization. In doing so, it is important to maintain a high-quality representation of the contrast map, otherwise the perceptual advantage of saving the contrast map is diminished. The compressed contrast map is given by  $\hat{C}_{\text{rms}}(m)$ . The size of this map must usually be kept

between 2–4 times smaller than the size of uncoded map to achieve this goal. While the explicit coding procedure reduces the overhead associated with the side information, the compressed contrast *and* classification maps can still take up a noticeable percentage of the size of the bitstream. An implicit method of rate reduction for the side information is used that is tied closely with the mechanism used to code the quantized wavelet coefficients, and is discussed in Section 4.4.

**4.3. Quantization Step Size Computation.** Once the (compressed and decompressed) contrast and classification maps have been established, quantization step sizes for each wavelet coefficient can be derived. With these step sizes, wavelet coefficients in each location are quantized to the threshold of perceived distortion. In other words, the following quantization scheme is designed to create *visually lossless* images. Lossy compression to an arbitrary coded bit rate can also be achieved based on this approach, and is discussed as well.

A quantization step size is derived for each individual wavelet coefficient in each subband as follows. In order to compute this step size, each wavelet coefficient must be associated with a mask type and a contrast value. Because these quantities are computed on a per-patch basis, this assignment proceeds differently depending on the number of coefficients in the subband relative to the patch size. Suppose the image to be compressed is  $N \times N$  pixels in size, and that the patch size is  $M \times M$  pixels. For wavelet subbands with fewer than  $(N/M) \times (N/M)$  coefficients, there are multiple entries in the contrast map that correspond to each wavelet coefficient. In this case, the average contrast computed over all patches corresponding to each coefficient is associated with that coefficient. This operation is conceptually equivalent to resampling the contrast map such that the resampled version is the same size as the wavelet subband. A pictorial representation of this process is illustrated in Figure 17. An analogous method is used to determine contrast values for subbands with more than  $(N/M) \times (N/M)$  coefficients. Assignment of mask types to each wavelet coefficient proceeds in the same way.

After each wavelet coefficient is associated with a contrast and a mask type, these values are mapped to a set of detection thresholds.  $\text{CT}_{\text{type}}(u_0, f_0, \theta_0)$  denotes the per-type contrast threshold associated with wavelet coefficient  $x(u_0, f_0, \theta_0)$ . The underlying assumption is that if patches of wavelet coefficients are quantized to induce the threshold contrast, the quantization error will be barely visible in the presence of the original image data. The masking model presented earlier can be used for this purpose, but is very computationally intensive. Thus, the following alternative is used for practical reasons. Previous texture masking experiments have demonstrated that TvC curves for images patches with distortions associated with a range of spatial frequencies  $f$  [53] can be described with a *frequency-adaptive* model of the form

$$\text{CT}(m, f_0) = \frac{(H(f_0) + C_{\text{rms}}(m)^{\alpha(f_0)})^{1/\beta(f_0)}}{N(f_0)}, \quad (9)$$



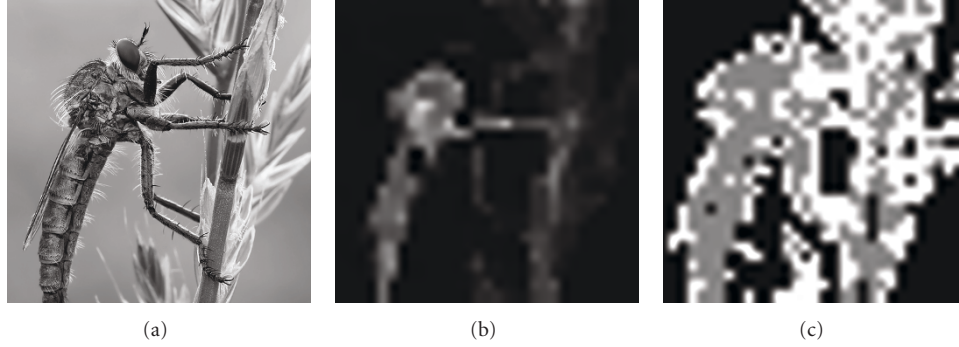


FIGURE 16: (a) An example image *bug*, (b) with a contrast map, and (c) classification map. In the contrast map, dark regions denote low values and light regions denote higher values. In the classification map, black regions denote texture, grey regions denote structure, and white regions represent edges. This information must be transmitted to the decoder.

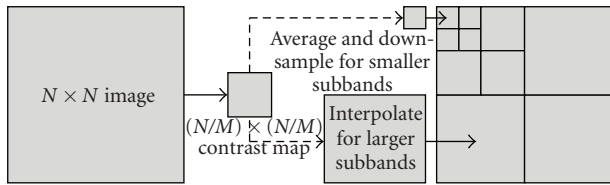


FIGURE 17: Diagram of relationship between contrast maps and wavelet subband coefficients. In order to associate each coefficient in a subband with a contrast value,  $\hat{C}_{rms}(u_0, f_0, \theta_0)$  (which is needed to predict the step size for the coefficient), the contrast map  $\hat{C}_{rms}(m)$  is resampled in order to match the number of map entries with subband coefficients. A similar approach is used to associate each wavelet coefficient with a patch type.

where  $C_{rms}(m)$  denotes the mask contrast for patch  $m$ . This same model is adopted for each patch of coefficients, on a per-mask-type basis. Mask-type-specific TvC data presented herein can be used to fit this model for wavelet subbands with a center spatial frequency of 4.6 cycles/degree; the relative differences between average contrast thresholds for the different frequencies in [53] can be used to derive relationships mapping  $\hat{C}_{rms}(u_0, f_0, \theta_0)$  to  $CT_{type}(u_0, f_0, \theta_0)$  for different frequencies (and types).

When this value has been computed for each wavelet coefficient, local step sizes may be determined as the step sizes that induce  $CT_{type}(u_0, f_0, \theta_0)$  in the distortion image. These values, given by  $Q^{threshold}(u_0, f_0, \theta_0)$ , may be found via iterative bisection search. One difficulty with determining  $Q^{threshold}(u_0, f_0, \theta_0)$  is that contrast thresholds are measured in the luminance domain. As a result, an iterative approach can be cumbersome. Nevertheless, faster alternatives exist. The contrast threshold can be mapped to an MSE distortion value (using the method described in [53]), and a variety of models for wavelet coefficient data can be used to map this MSE distortion to a step size.

Though the presented experiment and proposed coder yield performance that is optimized for at-threshold (visually lossless) compression, it can be extended to produced coded images at any target rate. The emergence of a number of new rate-control procedures [54, 55], which essentially can

be applied on a per-patch basis, yield the ability to do so quickly and accurately. To create coded images at any target rate, the at-threshold step sizes must be modified. One way to do so is the following. Prior to any quantization or coding, each wavelet coefficient is scaled by a spatially-selective weight that is inversely proportional to the step sizes which quantize each coefficient to the threshold of detection. In other words, each wavelet coefficient  $x(u_0, f_0, \theta_0)$  is first normalized by  $Q^{threshold}(u_0, f_0, \theta_0)$ , yielding modified coefficients  $\hat{x}(u_0, f_0, \theta_0) = x(u_0, f_0, \theta_0)/Q^{threshold}(u_0, f_0, \theta_0)$ . A traditional (patch-adaptive) MSE-based rate-distortion optimization procedure is carried out to determine step size  $Q_{MSE}(u_0, f_0, \theta_0)$  that will optimally compress  $\hat{x}(u_0, f_0, \theta_0)$  at a target rate. Assuming dead-zone quantization where the size of the dead-zone is twice the step size, the final quantization indices are given by

$$\begin{aligned} i(u_0, f_0, \theta_0) &= \left\lfloor \frac{\hat{x}(u_0, f_0, \theta_0)}{Q_{MSE}(u_0, f_0, \theta_0)} \right\rfloor \\ &= \left\lfloor \frac{x(u_0, f_0, \theta_0)}{Q^{threshold}(u_0, f_0, \theta_0) \cdot Q_{MSE}(u_0, f_0, \theta_0)} \right\rfloor. \end{aligned} \quad (10)$$

Note that this operation is equivalent to quantizing the original wavelet coefficients  $x(u_0, f_0, \theta_0)$  by step sizes  $\hat{Q}(u_0, f_0, \theta_0) = Q^{threshold}(u_0, f_0, \theta_0) \cdot Q_{MSE}(u_0, f_0, \theta_0)$ . A similar approach has been previously used combining a spatially-adaptive quantization scheme with a JPEG-2000 coder [1].

**4.4. Implicit Side Information and Quantized Wavelet Coefficient Coding.** A number of methods are available to code the actual image data. The image may be compressed in a *nonembedded* fashion, where wavelet coefficient quantization indices are separated into a *significance map*, *refinement bits*, and *sign bits*. One advantage of this approach is that it can use the previously coded side information to further reduce the overall bit-rate as follows. The locations of nonzero entries in each subband significance map are coded first (which is why the bitstream is not refinable below the subband level). Then, the values of the entries are coded *conditioned on the step sizes*



FIGURE 18: Test images used to evaluate the proposed coder. These images were chosen to span a range of mask types and contrast content.

used to create each entry. Refinement bits are represented with a simple adaptive arithmetic coder, and sign bits are inserted into the stream as needed uncoded. Though this approach is essentially nonembedded, all fully received subbands in a partially received image can be decoded.

Another alternative that provides a higher level of robustness to bitstream truncation errors is to generate the step sizes, quantize the wavelet data, then simply compress the resulting quantization indices using a more traditional embedded wavelet coder. Several examples of

coders of this nature are capable of representing quantized wavelet data in fine (per-bit-plane) layers [56, 57]. If the coded data is truncated at any point during transmission, all the received information can be reconstructed into a partially received image. There is thus a tradeoff between the achievable efficiency and robustness of the coded image due to the requirement of explicit side information, which is analyzed later. A diagrammatic representation of this part of the coder is illustrated in bottom portion of Figure 15.

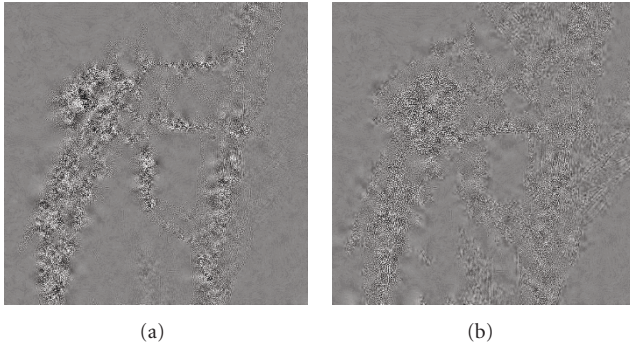


FIGURE 19: (a) Residual images resulting from compressing image *bug* using the texture-masking-based approach in [58], and (b) the proposed structural-masking-based approach, which have been contrast-enhanced to emphasize differences.

**4.5. Compression Results.** To test the proposed method, images were compressed at visually lossless rates using the proposed method and compared with the implementation of the coder in [58]. The main differences between these coders are that (1) the proposed coder requires additional rate to specify the results of the classification stage, such that the decoder can invertibly derive the quantization step sizes from the contrast map, (2) the proposed coder derives step sizes from TvC curves tailored to the mask type of each region (texture, structure, or edge) instead of TvC curves for just for textures, and (3) the proposed coder segments images into  $16 \times 16$  blocks instead of  $8 \times 8$  blocks, since content classification, instead of a highly granular segmentation, can be used to ensure appropriate selection of step sizes. Test images consist of twenty-four  $512 \times 512$  8 bit grayscale images, collected from standard databases as well as the Internet, illustrating a range of mask types. These images are illustrated in Figure 18.

**4.5.1. Compression Performance.** For the tested images, the proposed coder demonstrates a reduction in overall coded rate by an average factor of about 8%. Rates for individual test images are listed in Table 3, as well as the reductions achieved in bits-per-pixel (bpp). Reductions occur for two main reasons: first, the side rate associated with a  $16 \times 16$  pixel block segmentation is smaller than that associated with an  $8 \times 8$  pixel block segmentation, and second, the mask-type-specific segmentation results in slightly higher contrast values and thus slightly more aggressive step sizes for certain regions. At the same time, however, the proposed coder spends more bits representing certain regions classified as structures or edges (reflected in Figure 19). The PSNR values in the table corroborate this notion; sometimes the PSNR increases due to finer quantization, whereas sometimes it decreases due to coarser quantization. There are a few examples in which the rate increases. In some cases, this event occurs due to patch classification error; if the model for a patch is not chosen appropriately, coding efficiency will be reduced.

**4.5.2. Subjective Verification.** Results of a perceptual test comparing the two coders involving six of the test images is illustrated in Table 4. Ten nonexpert viewers were presented with both images and asked which was of better quality. On average, 2/3 of the viewers preferred the images created with the proposed method. Figure 19 illustrates a comparison between residuals, which show that the proposed coder places smaller errors in the regions corresponding to the structures on *bug* and larger errors in some edge regions that can mask more distortion. The fact that image quality is maintained, even when PSNR decreases, suggests that the proposed coder is making more efficient use of masking properties. We invite interested readers to view the results for all 24 test images at <http://foulard.ece.cornell.edu/gaubatz/piv08>.

**4.5.3. Limitations and Extensions.** The proposed nonembedded coding scheme has demonstrated the ability to efficiently represent an image despite the side information to specify highly spatially adaptive perceptual quantization parameters. Still, it does not result in a fully scalable image representation. It is more difficult to take advantage of the side information for implicit rate gain when representing the data in a finely layered fashion. The far right column in Table 3 illustrates this effect for the tested images. Creating a fully-scalable representation results in an increase in coded image size of roughly 0.01 to 0.04 bpp.

## 5. Conclusions

In this paper, we have advocated for the use of a patch-based structural masking model in which neural modeling is coupled with a classification scheme that selects appropriate model parameters based on the type of structural content (texture, structure, and edge) contained in local image patches.

The results of a psychophysical masking experiment using masks consisting of natural-image patches revealed that the ability of an image patch to mask wavelet subband quantization distortions depends not only on mask contrast, but also on whether the mask is a texture, structure, or edge. As previous masking studies have found, our results revealed that detection thresholds increase as the contrast of the mask increases. For very low mask contrasts of 0.01, 0.02, and 0.04, detection thresholds were similar for textures, structures, and edges. However, the thresholds for the three types demonstrated a marked divergence as the contrasts of the masks increased. High-contrast textures elevated detection thresholds approximately 4.3 times over those for high-contrast edges, and high-contrast structures elevated thresholds approximately 2.6 times over those for high-contrast edges. These results demonstrate that a proper measure of masking in images should account both for local contrast and for local image content.

By fitting a gain-control-based masking model to these data, we reported model optimal parameters for textures, structures, and edges. We found the optimal model parameters to vary substantially for the three image types. The optimal parameter values for textures were similar to standard

TABLE 3: Comparison between the coded rate/side rate/PSNR difference achieved with an implementation of the texture-based coder in [58] and the proposed classification-based coder.

Image	Overall rate (bpp)		Change in overall rate (%)	Side rate (bpp)		Change in side rate (bpp)	Change in PSNR (dB)	Extra rate for layers (bpp)
	Texture-based	Texture/structure/edge-based		Texture-based	Texture/structure/edge-based			
Baby	0.921	0.817	-11.229	0.015	0.017	0.002	-0.94	0.005
Beans	1.224	1.178	-3.721	0.062	0.026	-0.036	+0.58	0.029
Bug	1.085	0.913	-15.848	0.043	0.024	-0.019	-0.45	0.027
Casino	0.974	1.017	+4.415	0.048	0.025	-0.023	+1.94	0.029
Duck	1.333	1.287	-3.419	0.047	0.022	-0.025	+0.19	0.026
Fruits	1.098	1.030	-6.204	0.023	0.021	-0.002	-0.86	0.013
Gazebo	1.231	1.207	-1.930	0.051	0.026	-0.025	-0.03	0.027
Geckos	1.186	1.297	+9.350	0.054	0.024	-0.030	+0.95	0.037
Harbour	1.394	1.164	-16.542	0.042	0.024	-0.018	-1.10	0.030
Horse	1.396	1.254	-10.185	0.041	0.022	-0.019	-0.22	0.023
House	1.432	1.168	-18.471	0.038	0.022	-0.016	-1.37	0.025
Jackal	1.865	1.599	-14.263	0.064	0.028	-0.036	-0.12	0.025
Lander	1.394	1.304	-6.482	0.047	0.025	-0.022	-0.31	0.021
Lena	0.639	0.650	+1.841	0.033	0.022	-0.011	+0.26	0.007
Mill	1.672	1.658	-0.813	0.060	0.029	-0.031	+1.11	0.037
Pelicans	0.615	0.508	-17.368	0.030	0.018	-0.011	-1.09	0.017
Rainriver	1.089	0.979	-10.088	0.051	0.023	-0.028	+0.81	0.022
Rhino	1.667	1.729	+3.694	0.069	0.028	-0.041	+1.34	0.044
Roommates	0.806	0.713	-11.491	0.030	0.021	-0.009	-1.14	0.018
Seagulls	1.172	0.965	-17.708	0.046	0.024	-0.022	-0.37	0.045
Stock	0.962	0.932	-3.109	0.045	0.027	-0.018	-1.12	0.031
Sun	0.444	0.353	-20.466	0.018	0.017	-0.001	-2.46	0.009
Temple	0.884	0.776	-12.256	0.038	0.022	-0.016	-0.95	0.016
Wall	1.134	0.877	-22.721	0.028	0.020	-0.009	-1.76	0.003
Average	1.151	1.057	-8.119	0.043	0.023	-0.019	-0.30	0.024

TABLE 4: Comparison of the subjective verification results achieved with the texture-based coder in [58] and the proposed classification-based coder.

Image	Preference (no. of viewers)	
	Texture-based	Texture/structure/edge-based
Beans	3	7
Bug	2	8
Gazebo	3	7
House	4	6
Rhino	4	6
Temple	4	6

parameter values derived previously for simplistic masks. Accordingly, the standard model performs well in predicting detection thresholds for masks consisting of textures; the model predictions fall largely within the error bars of the measured thresholds. However, application of these standard parameter values to masks consisting of structures and edges revealed that the model does not generalize across image

type. When the texture-optimal parameters were applied to edges and when the edge-optimal parameters were applied to structures, more than twice the RMS error in prediction performance was observed as compared to that achieved by using the correct optimal parameters. Application of the edge-optimal parameters to textures resulted in nearly four times the RMS error in prediction performance compared to that achieved by using the texture-optimal parameters. These findings advocate that when gain-control-based models of masking are used in image-processing applications, one must take into account the *type* of content contained in local image patches and select the model parameters accordingly.

The utility of the patch-based structural masking approach was demonstrated via an application to image compression. The proposed compression scheme classifies each patch as a texture, structure, or edge, and then selects quantization step sizes based on the corresponding type-specific masking results obtained in the psychophysical experiment. Although additional explicit side information is needed to convey the spatially adaptive step sizes to the decoder, the proposed approach yields an efficient

wavelet compression strategy that can be combined with very accurate rate-control procedures. Testing revealed that applying the patch-type-specific adaptations saves on average 8% of the total bit-rate.

## References

- [1] W. Zeng, S. Daly, and S. Lei, "Point-wise extended visual masking for JPEG-2000 image compression," in *Proceedings of the International Conference on Image Processing (ICIP '00)*, vol. 1, pp. 657–660, Vancouver, Canada, September 2000.
- [2] I. Höntsch and L. J. Karam, "Locally adaptive perceptual image coding," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1472–1483, 2000.
- [3] C. J. van den Branden Lambrecht, "A working spatio-temporal model of the human visual system for image restoration and quality assessment applications," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '96)*, vol. 4, pp. 2291–2294, Atlanta, Ga, USA, May 1996.
- [4] S. Winkler, "Visual quality assessment using a contrast gain control model," in *Proceedings of the 3rd IEEE Workshop on Multimedia Signal Processing (MMSP '99)*, pp. 527–532, Copenhagen, Denmark, September 1999.
- [5] J. Huang and Y. Q. Shi, "Adaptive image watermarking scheme based on visual masking," *Electronics Letters*, vol. 34, no. 8, pp. 748–750, 1998.
- [6] M. A. Masry, D. M. Chandler, and S. S. Hemami, "Digital watermarking using local contrast-based texture masking," in *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers (ACSSC '03)*, vol. 2, pp. 1590–1594, Pacific Grove, Calif, USA, November 2003.
- [7] I. G. Karybali and K. Berberidis, "Efficient spatial image watermarking via new perceptual masking and blind detection schemes," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 256–274, 2006.
- [8] W. Liu, L. Dong, and W. Zeng, "Optimum detection for spread-spectrum watermarking that employs self-masking," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 4, pp. 645–654, 2007.
- [9] A. Koz and A. A. Alatan, "Oblivious spatio-temporal watermarking of digital video by exploiting the human visual system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 326–337, 2008.
- [10] T.-T. Lam, L. J. Karam, and G. P. Abovsleman, "Robust image coding using perceptually-tuned channel-optimized trellis-coded quantization," in *Proceedings of the 42nd IEEE Midwest Symposium on Circuits and Systems (MWSCAS '99)*, vol. 2, pp. 1131–1134, Las Cruces, Minn, USA, August 1999.
- [11] "Jndmetrix technology," Sarnoff Corporation, <http://www.sarnoff.com>.
- [12] A. B. Watson, Q. J. Hu, J. F. McGowan III, and J. B. Mulligan, "Design and performance of a digital video quality metric," in *Human Vision and Electronic Imaging IV*, vol. 3644 of *Proceedings of SPIE*, pp. 168–174, San Jose, Calif, USA, January 1999.
- [13] M. A. Masry and S. S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 133–146, 2004.
- [14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [15] J. A. Ferwerda, "Fundamentals of spatial vision," in *Applications of Visual Perception in Computer Graphics*, V. Interrante, Ed., pp. 1–27, ACM SIGGRAPH, Orlando, Fla, USA, 1998.
- [16] B. Walter, S. N. Pattanaik, and D. P. Greenberg, "Using perceptual texture masking for efficient image synthesis," *Computer Graphics Forum*, vol. 21, no. 3, pp. 393–399, 2002.
- [17] D. I. Neuhoff and T. N. Pappas, "Perceptual coding of images for halftone display," *IEEE Transactions on Image Processing*, vol. 3, no. 4, pp. 341–354, 1994.
- [18] Q. Yu and K. J. Parker, "Quality issues in blue noise halftoning," in *Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts III*, G. Beretta and R. Eschbach, Eds., vol. 3300 of *Proceedings of SPIE*, pp. 376–385, San Jose, Calif, USA, January 1998.
- [19] S.-M. Lu, H.-C. Pu, and C.-T. Lin, "A HVS-directed neural-network-based approach for impulse-noise removal from highly corrupted images," in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics (ICSMC '03)*, vol. 1, pp. 72–77, Washington, DC, USA, October 2003.
- [20] D. Wang, I. Sato, T. Okabe, and Y. Sato, "Radiometric compensation in a projector-camera system based properties of human vision system," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, p. 100, San Diego, Calif, USA, June 2005.
- [21] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing, and Digital Display III*, A. B. Watson, Ed., vol. 1666 of *Proceedings of SPIE*, pp. 179–206, San Jose, Calif, USA, February 1993.
- [22] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Human Vision, Visual Processing, and Digital Display V*, vol. 2179 of *Proceedings of SPIE*, pp. 127–141, San Jose, Calif, USA, February 1994.
- [23] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*, E. Peli, Ed., pp. 245–283, World Scientific, River Edge, NJ, USA, 1995.
- [24] A. B. Watson and J. A. Solomon, "A model of visual contrast gain control and pattern masking," *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2379–2391, 1997.
- [25] W. S. Geisler and D. G. Albrecht, "Cortical neurons: isolation of contrast gain control," *Vision Research*, vol. 32, pp. 1409–1410, 1992.
- [26] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, no. 2, pp. 181–197, 1992.
- [27] J. M. Foley, "Human luminance pattern mechanisms: masking experiments require a new model," *Journal of the Optical Society of America A*, vol. 11, no. 6, pp. 1710–1719, 1994.
- [28] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Transactions on Image Processing*, vol. 8, no. 5, pp. 717–730, 1999.
- [29] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Transactions on Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.
- [30] P. Le Callet and D. Barba, "A robust quality metric for color image quality assessment," in *Proceedings of the International Conference on Image Processing (ICIP '03)*, vol. 1, pp. 437–440, Barcelona, Spain, September 2003.
- [31] T. S. Meese and D. J. Holmes, "Adaptation and gain pool summation: alternative models and masking data," *Vision Research*, vol. 42, no. 9, pp. 1113–1125, 2002.

- [32] D. G. Albrecht, W. S. Geisler, and A. M. Crane, "Nonlinear properties of visual cortex neurons: temporal dynamics, stimulus selectivity, neural performance," in *The Visual Neurosciences*, L. Chalupa and J. Werner, Eds., pp. 825–837, MIT Press, Boston, Mass, USA, 2003.
- [33] T. Caelli and G. Moraglia, "On the detection of signals embedded in natural scenes," *Perception and Psychophysics*, vol. 39, no. 2, pp. 87–95, 1986.
- [34] M. J. Nadenau and J. Reichel, "Image-compression-related contrast-masking measurements," in *Human Vision and Electronic Imaging V*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3959 of *Proceedings of SPIE*, pp. 188–199, San Jose, Calif, USA, January 2000.
- [35] D. M. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *Journal of the Optical Society of America A*, vol. 20, no. 7, pp. 1164–1180, 2003.
- [36] J. D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1053–1060, 1995.
- [37] M. D. Gaubatz, D. M. Chandler, and S. S. Hemami, "Spatial quantization via local texture masking," in *Human Vision and Electronic Imaging X*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 5666 of *Proceedings of SPIE*, pp. 95–106, San Jose, Calif, USA, January 2005.
- [38] D. M. Chandler and S. S. Hemami, "Suprathreshold image compression based on contrast allocation and global precedence," in *Proc. Human Vision and Electronic Imaging 2003*, vol. 5007, pp. 73–86, Santa Clara, Calif, USA, January 2003.
- [39] D. M. Chandler and S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 397–410, 2005.
- [40] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [41] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, no. 8, pp. 1164–1175, 1997.
- [42] B. Moulden, F. A. A. Kingdom, and L. F. Gatley, "The standard deviation of luminance as a metric for contrast in random-dot images," *Perception*, vol. 19, no. 1, pp. 79–101, 1990.
- [43] F. A. A. Kingdom, A. Hayes, and D. J. Field, "Sensitivity to contrast histogram differences in synthetic wavelet-textures," *Vision Research*, vol. 41, no. 5, pp. 585–598, 2001.
- [44] P. J. Bex and W. Makous, "Spatial frequency, phase, and the contrast of natural images," *Journal of the Optical Society of America A*, vol. 19, no. 6, pp. 1096–1106, 2002.
- [45] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *Journal of the Optical Society of America A*, vol. 18, no. 10, pp. 2385–2397, 2001.
- [46] A. B. Watson and D. G. Pelli, "QUEST: a Bayesian adaptive psychometric method," *Perception and Psychophysics*, vol. 33, no. 2, pp. 113–120, 1983.
- [47] D. H. Brainard, "The psychophysics toolbox," *Spatial Vision*, vol. 10, no. 4, pp. 433–436, 1997.
- [48] D. G. Pelli, "The Video Toolbox software for visual psychophysics: transforming numbers into movies," *Spatial Vision*, vol. 10, no. 4, pp. 437–442, 1997.
- [49] A. B. Watson, R. Borthwick, and M. Taylor, "Image quality and entropy masking," in *Human Vision and Electronic Imaging II*, vol. 3016 of *Proceedings of SPIE*, pp. 2–12, San Jose, Calif, USA, February 1997.
- [50] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.
- [51] S. S. Hemami, D. M. Chandler, B. G. Chern, and J. A. Moses, "Suprathreshold visual psychophysics and structure-based visual masking," in *Visual Communications and Image Processing 2006*, J. G. Apostolopoulos and A. Said, Eds., vol. 6077 of *Proceedings of SPIE*, pp. 1–15, San Jose, Calif, USA, January 2006.
- [52] C. Poynton, "The rehabilitation of gamma," in *Human Vision and Electronic Imaging III*, B. E. Rogowitz and T. N. Pappas, Eds., vol. 3299 of *Proceedings of SPIE*, pp. 232–249, San Jose, Calif, USA, 1998.
- [53] D. M. Chandler, N. L. Dykes, and S. S. Hemami, "Visually lossless compression of digitized radiographs based on contrast sensitivity and visual masking," in *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, Y. Jiang and M. Eckstein, Eds., vol. 5749 of *Proceedings of SPIE*, pp. 359–372, San Diego, Calif, USA, February 2005.
- [54] M. D. Gaubatz and S. S. Hemami, "Robust rate-control for wavelet-based image coding via conditional probability models," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 649–663, 2007.
- [55] M. D. Gaubatz and S. S. Hemami, "Efficient entropy estimation based on doubly stochastic models for quantized wavelet image data," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 967–981, 2007.
- [56] 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, no. 3, pp. 243–250, 1996.
- [57] P. Simard, D. Steinkraus, and H. Malvar, "On-line adaptation in image coding with a 2-D tarp filter," in *Proceedings of the Data Compression Conference (DCC'02)*, pp. 23–32, Snowbird, Utah, USA, April 2002.
- [58] M. D. Gaubatz, D. M. Chandler, and S. S. Hemami, "Spatially-selective quantization and coding for wavelet-based image compression," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '05)*, vol. 2, pp. 209–212, Philadelphia, Pa, USA, March 2005.