Chapter 1

Visually Lossless Perceptual Image Coding Based on Natural-Scene Masking Models

Yi Zhang, Md Mushfiqul Alam and
Damon M. Chandler

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/65362

Abstract

Perceptual coding is a subdiscipline of image and video coding that uses models of human visual perception to achieve improved compression efficiency. Nearly, all image and video coders have included some perceptual coding strategies, most notably visual masking. Today, modern coders capitalize on various basic forms of masking such as the fact that distortion is harder to see in very dark and very bright regions, in regions with higher frequency content, and in temporal regions with abrupt changes. However, beyond these obvious forms of masking, there are many other masking phenomena that occur (and co-occur) when viewing natural imagery. In this chapter, we present our latest research in perceptual image coding using natural-scene masking models. We specifically discuss: (1) how to predict local distortion visibility using improved natural-scene masking models and (2) how to apply the models to high efficiency video coding (HEVC). As we will demonstrate, these techniques can offer 10–20% fewer bits than baseline HEVC in the ultra-high-quality regime.

Keywords: HEVC, visual masking, contrast gain control, adaptive quantization

1. Introduction

Recent advancements in digital signal processing technologies have made available a wide variety of digital media for end use by consumers and practitioners. It is estimated that more than 100 billion digital photos and videos are recorded, transmitted, and viewed annually just in the United States. Today, the tremendous popularity of ubiquitously connected digital imaging devices has made the Internet the standard means by which to share imagery. Of

course, digital images/videos have many uses beyond entertainment, including online education, video conferencing, remote medical diagnoses, and many others. Such widespread use of digital images and videos places a great demand on compression algorithms which are absolutely crucial for reducing the bandwidth requirements of storing and transmitting these images and videos.

To this end, state-of-the-art image/video compression algorithms exploit the fact that the human visual system (HVS) is an imperfect sensor. When a digital image/video is to be viewed by a human, an exact bit-for-bit reconstruction is unnecessary; rather, the data can be coded in a non-invertible or *lossy* fashion. Lossy compression is useful for applications where lower information fidelity can be tolerated, such as in consumer photography, computer vision, and machine vision applications. If the compression distortions are invisible, the compression is said to be *visually lossless*. Visually lossless compression techniques generally take advantage of a low-level psychophysical phenomenon such as visual masking. If, on the other hand, the compression distortions are visible, the compression is called *visually lossy*. Visually lossy compression techniques aim to generate the best-looking reconstructed version under the given bit-rate constraints. Both of these paradigms fall under the more general category of the so-called *perceptual coding*, owing to the need to model the human visual system (HVS), and in particular, how the HVS detects and perceives compression-induced distortions.

With the release of each new coding standard, the emphasis in perceptual coding research has largely shifted from the mid-quality regime toward the ultra-high-quality regime, with the aim of producing compressed images and videos which are visually equivalent to the originals. Thus, research in visually lossless compression has seen a recent resurgence in importance. In this chapter, we focus exclusively on visually lossless image compression. The key challenge in visually lossless compression is to automatically determine, on a per image basis, the maximum amount of compression that can be applied before the resulting image appears distorted. However, to tackle this challenge requires the ability to accurately and efficiently predict the visibility of local distortions in an image, a task which still remains elusive in the current research.

Perceptual coding strategies have long relied on well-known properties of the HVS largely derived from the visual psychophysics literature (e.g., see [1, 2]). Perhaps, the most well-known and widely used property is the *contrast sensitivity function* (CSF), which specifies the visibility of a narrowband spatial pattern (the *target* of detection) as a function of the pattern's spatial or temporal frequency. Previous psychophysical studies have shown that the minimum contrast needed to detect a visual target (e.g., distortions) varies with both the spatial frequency and the temporal frequency of the target. This minimum contrast is called the *contrast threshold*, and the inverse of this threshold is called *contrast sensitivity*. For targets consisting of spatial sine waves, the CSF is band-pass, indicating that we are least sensitive to very low-frequency and very high-frequency targets. The *temporal CSF* is an extension of the spatial CSF which takes into account sensitivity to time-varying targets, typically demonstrating a peak in sensitivity around 4–8 Hz.

The CSF can be thought of as a baseline visual sensitivity measure because the CSF is traditionally measured for targets shown against a blank background. However, for targets

http://dx.doi.org/10.5772/65362

consisting of compression distortions, this blank-background scenario occurs only when the distortions happen to appear in very smooth regions such as in the sky. In other image regions, such as in structures, textures, and hybrids regions, the distortions are generally more difficult to detect (i.e., they exhibit higher contrast detection thresholds), and therefore, visual sensitivity to the distortions is said to be reduced in these regions. This concept of *visual masking* has served as the cornerstone of modern perceptual coding.

At the most general level, visual masking refers to a reduction or elimination in the visibility of one signal (called the "target") caused by the presence of another signal (called the "mask"). For image compression, the image serves as the mask, and the compression distortions serve as the targets of detection. There are various forms of visual masking which can occur and co-occur in images and video. For example, it is well-known that humans have a harder time seeing distortions in very bright regions of an image, an HVS property called *luminance masking*. To capitalize on this fact, modern coding schemes more coarsely quantize the coefficients corresponding to (devote fewer bits to) locations of higher luminance. A similar strategy can be used for very busy regions of an image (*contrast masking*) or during scene changes in video (*temporal masking*).

These low-level aspects of the HVS are so commonly used in image/video coding for two simple reasons: (1) they are easy to incorporate and (2) such low-level aspects have been well-documented in the visual psychology literature with accompanying computational models. However, most existing models of masking (and thus, existing perceptual coding techniques) are largely based on findings using artificial stimuli rather than on a true database of natural scenes. The advantage of these artificial masks is that they have well-defined features and parameters, which allows one to investigate the effects of specific mask properties on the detection thresholds. However, in image compression, the mask is necessarily an image, and thus, it remains unclear whether the results obtained using artificial masks can be used to predict the results obtained using natural scene masks. There are some studies using natural scenes as masks, but these studies either employed only a limited number of tested images, or the thresholds were limited to select spatial locations within images (e.g., [3–5]).

In this chapter, we present our latest research in visually lossless image compression which operates based on the concept of *masking maps* predicted from a natural-scene masking model built upon a large local masking database [6]. Specifically, we recently published the results of a large-scale psychophysical study designed to obtain local contrast detection thresholds (masking maps) for a database of natural images [6]. This database can serve as crucial ground-truth data for investigating on how local image content affects the visual masking thresholds. Using this database, we present an high efficiency video coding (HEVC)-based quantization scheme which uses the contrast gain control (CGC) with structure facilitation model trained on the database of local masking thresholds to predict a masking map for the to-be-compressed image. The masking map is then used to guide a spatially adaptive quantization scheme, which more coarsely quantizes the blocks that can induce greater masking, and vice-versa. Using this approach, our technique can generate compressed images in which the contrasts of the local compression artifacts are much closer to their masked visibility thresholds than when using standard HEVC.

This chapter is organized as follows: Section 2 provides a brief review of current visually lossless perceptual image compression algorithms. In Section 3, we describe the computational models used to predict the masking map for any given input image. In Section 4, we describe how to incorporate the masking map to perform spatially adaptive compression using HEVC. In Section 5, we analyze and discuss the performance of the proposed visually lossless compression method. General conclusions are presented in Section 6.

2. Previous work on perceptual image compression

As we mentioned, the goal of visually lossless image compression is to generate images containing distortions at or just below the visual detection threshold. To this end, previous work in this area has exploited properties of the HVS (most notably the CSF and visual masking) and has taken a variety of approaches toward incorporating these visual properties into the transform, quantization, and/or encoding stages. In this section, we briefly review previous work on perceptual (HVS-based) image compression.

Perceptual image compression techniques can be dated back as early as 1990s when Safranek et al. [7] published one of earliest attempts at incorporating HVS properties into compression through a system called perceptually tuned subband image coder (PIC). Three properties of low-level vision were modeled in PIC: (1) contrast sensitivity, (2) luminance masking, and (3) contrast masking. These properties were used to guide the selection of per-subband quantization step sizes designed to yield visually lossless results. Although PIC was initially designed for visually lossless compression, Pappas et al. [8] reported that this system can also be used for visually lossy compression, and high performance can be achieved when the perceptual thresholds are properly scaled. Also, Hontsch et al. [9] extended PIC by exploiting visual masking; they proposed a locally adaptive perceptual coder, which discriminates between image components based on their perceptual relevance.

Later research on compression has exploited the properties of the HVS and employed the CSF to regulate the quantization step size in order to minimize the visibility of compression artifacts. For example, Nadenau et al. [10] incorporated HVS properties into a wavelet-based coding algorithm via a noise-shaping filtering stage which preceded quantization. Albanesi [11] proposed a method for incorporating HVS characteristics directly into the transform stage of a wavelet-based coder via the design of analysis and synthesis filters based on the CSF. Antonini et al. [12] introduced a wavelet coder which employed a CSF-weighted distortion criterion during bit allocation. O'Rourke et al. [13] proposed a wavelet-based image compression technique based on two properties of the HVS: orientation sensitivity and contrast sensitivity. Specifically, the diamond-shaped frequency passband of the HVS was exploited for the design of the compression scheme, and the logarithm of the contrast sensitivity was employed for bit allocation. Lai et al. [14] presented an image compression scheme in which contrast-sensitivity and visual masking adjustments were performed within a wavelet-based coder using a low-pass model of the CSF and a local measure of visual distortion. In two similar approaches, Beegan et al. [15] used a "CSF mask" to adjust transform coefficients prior to the

quantization, and Wei et al. [16] used a "visual compander." Also, in [17], Zhang et al. proposed luminance and chrominance CSF-based weighting in the discrete-wavelet-packet-transform domain to reduce perceptible information of the high-dynamic-range images.

There are also some researchers who conducted psychophysical experiment to measure visibility thresholds for compression artifacts in unnatural images and/or on natural scenes. For example, Watson et al. [18] measured visual detection thresholds for both individual wavelet basis functions and simulated wavelet subband quantization distortions presented against a gray background. The thresholds were modeled as a function of the spatial frequency of the distortions, and the model was then used to compute quantizer step sizes for each wavelet subband. In [19], Watson's approach was extended to lower rate coding via models of visual masking and summation. Nadenau et al. [5] measured the visibility thresholds of quantization noise in natural scenes and compared five visual masking models to predict the visibility thresholds. They concluded that a masking model considering local activity of the wavelet subbands performed better than point-wise contrast masking models.

In a recent study, Chandler et al. [3] proposed a new kind of masking called the *structural masking* by psychophysically measuring the visibility thresholds of wavelet distortions placed on small patches categorized in three groups: texture, structure, and edges. The authors have also proposed different set of values of parameters of contrast-gain control model [20] for three different categories and have shown that the category-specific masking model showed better compression results for wavelet-type compression schemes. Similarly, in [21], Chandler et al. proposed a visually lossless compression algorithm based on psychophysical detection experiments of wavelet distortion on radiograph images.

Several other studies have specifically focused on the visually lossless compression of JPEG and JPEG2000 compression schemes. For example, Oh et al. [22] developed a visually lossless compression model which allocates the code streams of the JPEG2000 encoder by measuring visibility thresholds via a wavelet statistics-based quantization distortion model and a visual masking model. In [23], Ponomarenko et al. pointed out that the visual quality of input (to-be-compressed) image has a large effect on the compression performance. Thus, they adaptively adjusted the scaling factor of the JPEG quantization matrix based on the estimated blur and noise content of the input image and showed that such a compression scheme gives larger compression ratio compared to super-high quality mode of consumer digital cameras. Leung et al. [24] proposed a JPEG2000-based visually lossless compression scheme for CT images in which the visibility thresholds varied according to the viewing window/display size of the CT image.

3. Computational models of local masking

This section describes the computational masking models that we developed to predict the masking map for the given input (to-be-compressed) image. First, we describe the ground-truth database used to train the models. Next, we describe a modified version of the model put forth by Watson and Solomon, which operates by simulating V1 neural responses with

contrast gain control (CGC). Here, we have modified the model and optimized its parameters to provide the best predictions for the aforementioned database. In addition, we describe an extension of the model to deal with structural facilitation which we earlier reported in [3]. Structural facilitation refers to the reduction in threshold (increased distortion visibility) in parts of the image containing highly recognizable structure.

3.1. Database of local masking in natural scenes

In [6], we performed a large-scale psychophysical experiment in which we measured thresholds for detecting simulated distortions placed within each 85 × 85 block of every image from the CSIQ database [25]. The simulated distortion was a narrowband log-Gabor noise target whose center frequency was chosen to be near the peak of visual sensitivity (3.6 cycles/degree of visual angle). The thresholds were obtained using a three-alternative forced-choice procedure [26]; we employed at least three subjects per image, with at least two trials per subject. The end result of the experiment was a masking map for each of the 30 CSIQ images; each entry in each map denotes the minimum contrast required for a human subject to detect distortions at that location in the image.

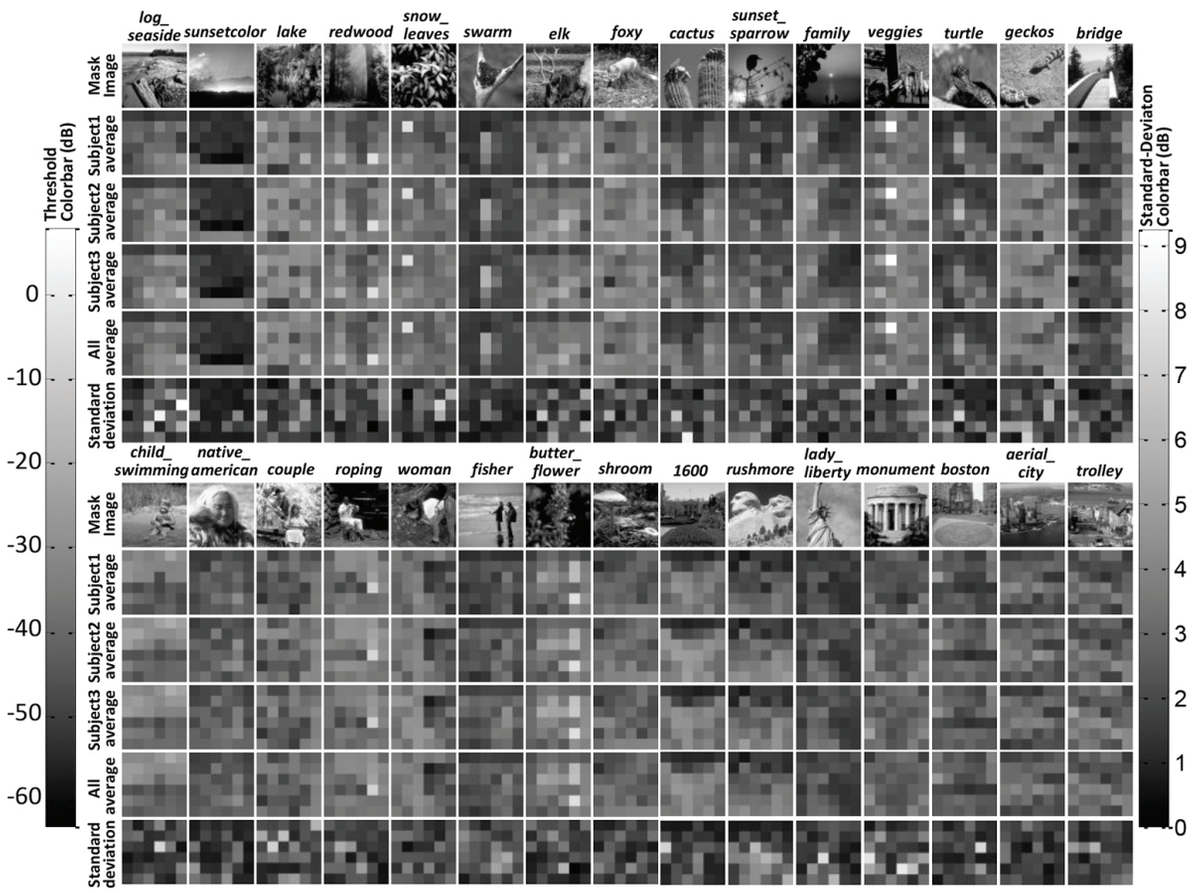

Figure 1. Masking maps and the corresponding standard deviation maps for all 30 images in the CSIQ database. See text for details.

Figure 1 shows the masking maps from the database. Each map consists of 36 values corresponding to the 36 blocks of the associated image. Brighter map values denote higher thresholds (i.e., more masking); darker maps values denote lower thresholds (less masking). The first and seventh rows of **Figure 1** show the 30 mask images. Below the mask images, the first, second, and third images show the average maps of the two trials of Subject 1, Subject 2, and Subject 3. The remaining rows show the average maps (taken across all six trials; 2 × 3 subjects), and the corresponding maps of the standard deviations of each average. Note that the averages and standard deviations are on different scales; please refer to the respective color bars shown in **Figure 1**. Overall, the subjects were in high agreement with each other and with themselves across separate trials.

In the following subsection, we describe the contrast gain control with structure facilitation model which operates by simulating V1 neural responses to predict these masking maps.

3.2. Contrast gain control with structure facilitation (CGC+SF) model

Contrast masking [27] has been widely used for predicting distortion visibility in images and videos [28, 42–44]. Among the many existing models of contrast masking, those which simulate the contrast gain-control response properties of V1 neurons are most widely used. Although several *contrast gain control* (CGC) models have been proposed in previous studies (e.g., Refs. [20, 27, 30, 31, 41]), in most cases, the model parameters are selected based on results obtained using either unnatural masks [20] or only a very limited number of natural images. Thus, in this chapter, we describe two approaches to improve the current CGC model: (1) the CGC model parameters are optimized by training on the large dataset of local masking in natural scenes; and (2) the CGC model is incorporated by a structural facilitation (SF) model which better captures the reduced masking observed in structured regions.

3.2.1. Watson-Solomon contrast gain control (CGC) model

The Watson and Solomon model [20] is a model of V1 simple-cell responses that includes CGC from neighboring neurons. **Figure 2** shows a block diagram of the model. The model takes two images as input: (1) the mask image (original image), and (2) the mask+target image (distorted image). Both of these images are then subjected to the following stages:

1. A spatial filter designed to mimic the human contrast sensitivity function (CSF).
2. A local spatial-frequency decomposition designed to mimic the initially linear response properties of individual V1 neurons.
3. Excitatory and inhibitory nonlinearities designed to mimic the nonlinear response properties of individual V1 neurons.
4. Divisive inhibition designed to mimic the interactions among groups of V1 neurons.

***Steps* 1 *and* 2**: For Step 1, we use the CSF filter specified in [32, 33]. For Step 2, we use a log-Gabor filterbank consisting of six scales and six orientations. The center radial frequencies of the filters are 0.3, 0.61, 1.35, 3.22, 7.83, 16.1 c/deg, each with a radial-frequency bandwidth of

2.75 octaves. The center orientations of the filters are 0°, ±30°, ±60°, 90°, each with an orientation bandwidth of 30°.

***Steps* 3 *and* 4**: Let $c(x_0, y_0, f_0, \theta_0)$ denote the output of the log-Gabor filter with a center of radial frequency f_0, an orientation θ_0, and at the spatial location (x_0, y_0). This filter output represents the initially linear response of the neuron. To obtain the nonlinear neural response, $R(x_0, y_0, f_0, \theta_0)$, we perform Steps 3 and 4 via the following equation:

$$R(x_0, y_0, f_0, \theta_0) = g \cdot \frac{c(x_0, y_0, f_0, \theta_0)^p}{b^q + \sum_{(x,y,f,\theta)\in I_N} (c(x,y,f,\theta))^q} \quad (1)$$

Here, g is an output gain factor (we use $g = 0.1$). The parameters p and q are the exhitatory and inhibitory exponents which impose the nonlinearities (we use $p = 2.4$ and $q = 2.35$). The parameter b is a constant designed to prevent division by zero (we use $b = 0.035$). The division simulates inhibition from neighboring neurons; these neurons constitute the so-called inhibitory pool, and they are neighbors in space, radial frequency, and orientation. In Eq. (1), the inhibitory pool is represented by the set of spatial and spatial frequency coordinates I_N. The neighbors come from a 3 × 3 surround in space, a ±0.7 octave bandwidth surround in radial frequency, and a ±60° bandwidth surround in orientation.

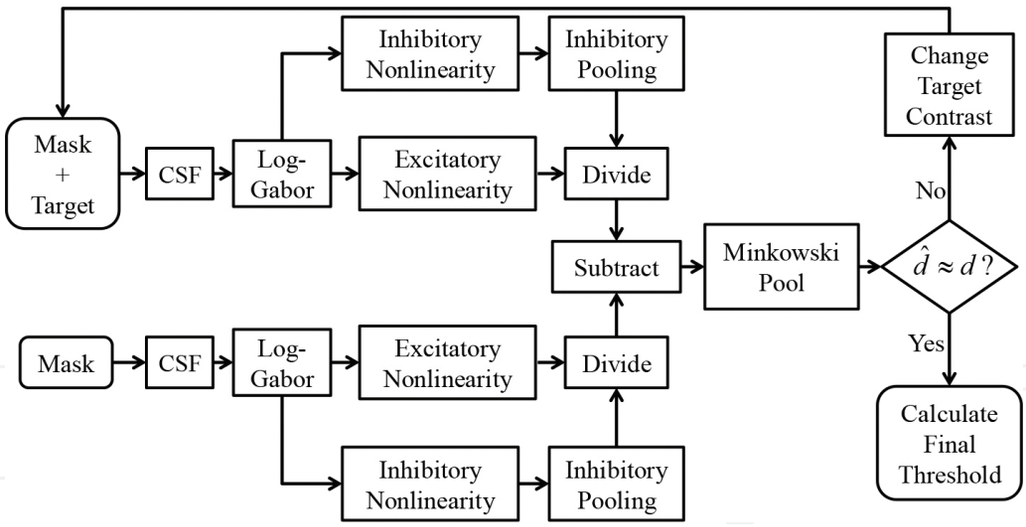

Figure 2. Flow of Watson and Solomon contrast gain control model.

All of the abovementioned parameters (g, p, q, b, and I_N) were chosen via a brute-force search to provide the best overall fit to the thresholds from our database, under the condition that the parameters remain within biologically plausible ranges [3]. The radial frequency bandwidth and center radial frequencies were chosen in this way as well. The other parameters of the model were either set as specified in [20] or were chosen based on our prior related modeling efforts [3].

Comparing the responses: Step 4 results in two collections of responses: One collection of responses to the mask, and another set of responses to the mask+target. The target is deemed visible if these collections of responses are sufficiently different from each other; thus, indicating a visible difference in the two stimuli (i.e., that the distortions are visible). To determine whether this condition is met, the collections of responses are subtracted from each other, then collapsed via Mikowski sum [20], and then this scalar difference ($\hat{d}$) is compared to a predefined "at-threshold" difference value ($d = 1$). We used a Minkowski exponent of 2.0 to collapse across space, and an exponent of 1.5 to collapse across radial frequency and orientation. The contrast of the target is iteratively adjusted until $\hat{d} \approx d$. When this condition is met, the contrast of the target is deemed to be the at-threshold contrast (i.e., the contrast detection threshold).

We refer interested readers to [6] for more specific details of the database and model.

3.2.2. Structure facilitation (SF) model

Using the optimized parameters described in the previous subsection, our implementation of the Watson and Solomon CGC model is quite accurate in predicting detection thresholds. On our database, the model is able to achieve a Pearson correlation coefficient (PCC) of 0.83 between the ground-truth and predicted thresholds. Generally, the model works best on regions containing textures and is worst on regions containing more complex structure. In particular, the model tends to overestimate thresholds for regions containing recognizable structure. This notion is demonstrated in **Figure 3**, which shows the ground-truth and predicted thresholds for two images; observe that the model predict the thresholds to be higher than ground-truth near the top of the gecko's body and in the child's face.

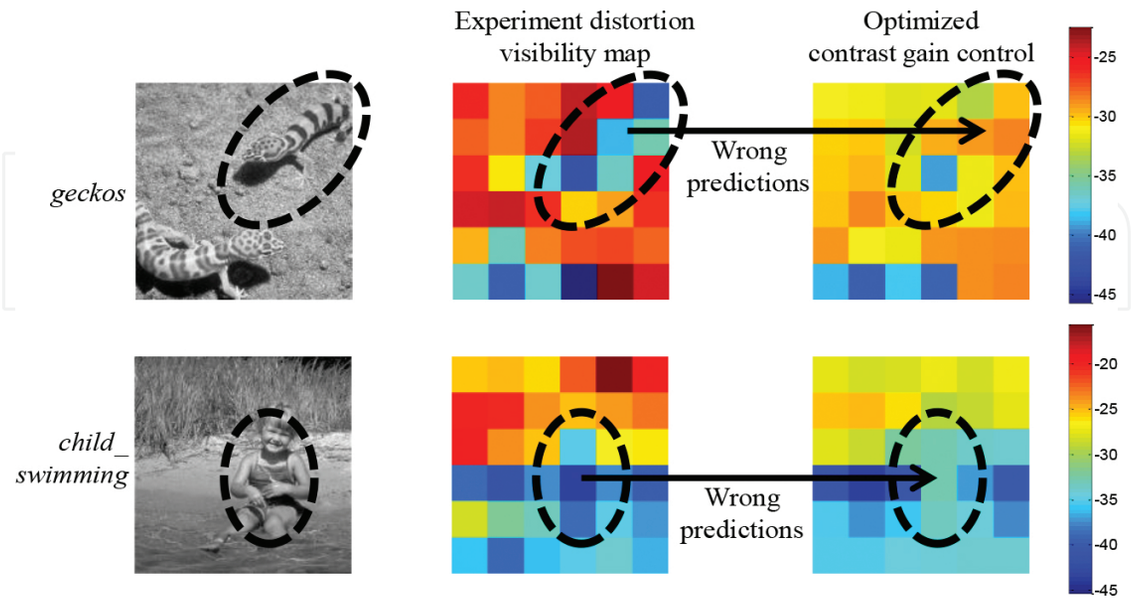

Figure 3. Examples of the Watson and Solomon model overestimating thresholds for distortion in some image regions that contain recognizable structures.

As we mentioned in [3], recognizable structures within the local regions of natural scenes facilitate (rather than mask) the distortion visibility. Thus, to model this "structure facilitation," we employ an inhibition modulation factor (λ_s) in the gain control equation:

$$R(x_0, y_0, f_0, \theta_0) = g \cdot \frac{c(x_0, y_0, f_0, \theta_0)^p}{b^q + \lambda_s \sum_{(x,y,f,\theta)\in I_N} \left(c(x, y, f, \theta)\right)^q} \tag{2}$$

where we adjust λ_s depending on the strength of structure within an image. Although the specific amount of inhibition modulation remains an open area of research, we have found the following sigmoidal relationship between λ_s and estimated structure strength to be quite effective (shown in **Figure 4**):

$$\lambda_{s,i} = \begin{cases} 1 - 80 \sum_{x,y=1,1}^{M,N} \left[1 / \left(1 + \exp\left(-\frac{S_i(x,y) - p(S,80)}{0.005} \right) \right) \right], & \max(S) > 0.04 \;\&\; \mathrm{Kurt}(S) > 3.5 \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

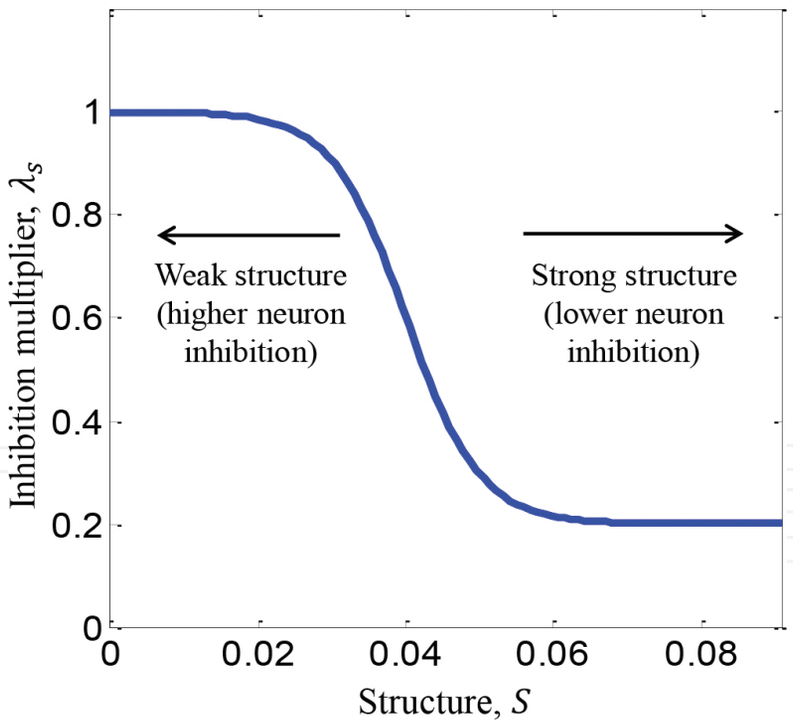

Figure 4. The inhibition multiplier λ_s varies depending on structure strength. Strong structures give rise to lower inhibition to facilitate the distortion visibility.

Observe that the inhibition modulation is applied in a block-based fashion. Here, $\lambda_{s,i}$ denotes the inhibition modulation factor for the i^{th} block of size $M \times N$.

The variable S in Eq. (3) is a map which denotes the local structure strength (described next), and S_i is a block of S corresponding to the i^{th} block of the image. The inhibition modulation for each block is further adjusted based on 80% largest values of S, denoted by the variable $p(S, 80)$. Furthermore, if the largest value of S is small, or if the kurtosis of S is small, then there is either no sufficient structure (e.g., the image is mostly textured or smooth), or the structure is not locally concentrated. In this case, no inhibition modulation is applied (i.e., $\lambda_{s,i} = 1$, for all blocks) (**Figure 4**).

The structure map S of an image is generated via the following equation which uses different feature maps:

$$S = L_n \times Sh_n \times E_n \times \left(1 - D_{\mu n}\right)^2 \times \left(1 - D_{\sigma n}\right)^2. \tag{4}$$

Here, L_n, Sh_n, and E_n denote maps of local luminance, local sharpness [29], and local first-order Shannon entropy, respectively. The values $D_{\mu n}$ and $D_{\sigma n}$ denote, respectively, maps of the average and the standard deviation of fractal texture features [34] computed for each local region. All features were computed for 32 × 32 blocks with 50% overlap between neighboring blocks. Each feature map was then normalized to the range [0, 1] and then resized to match the input image's dimensions. **Figure 5** shows some examples.

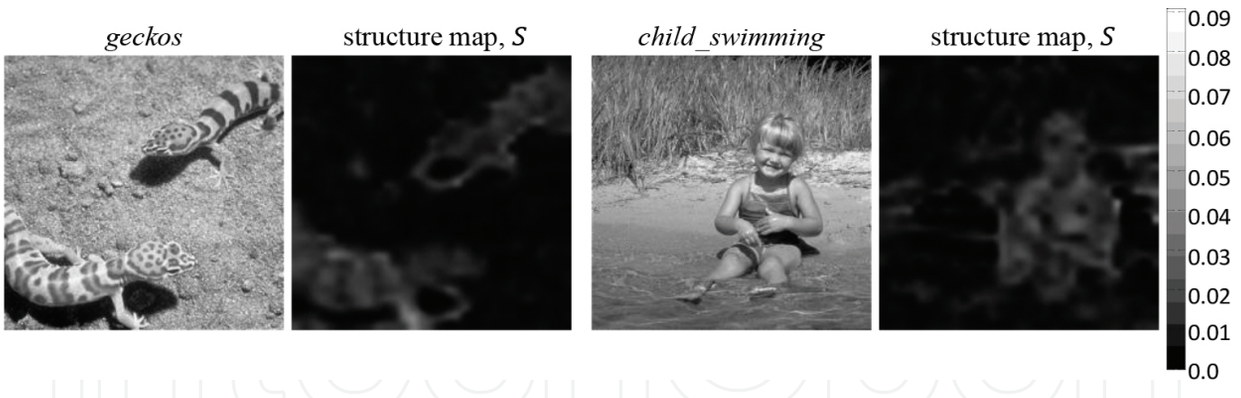

Figure 5. Structure maps of two example images. The color bar at right denotes the structure strength at each spatial location of the structure map.

The prediction performance of the Watson and Solomon CGC model can be greatly improved when the structure facilitation is taken into account [as specified in Eq. (2)]. As demonstrated in **Figure 6**, the proposed SF model was able to improve the CGC model's prediction performance in local image regions that contain recognizable structures, while not adversely affecting the prediction results of the others. For example, near the top of the gecko's body and in the child's face, the contrast detection thresholds predicted using the combined CGC+SF model match the ground-truth thresholds better than using the CGC model. Furthermore, the Pearson

correlation coefficients between the CGC+SF model predictions and ground-truth thresholds also improved as compared to using the CGC model alone.

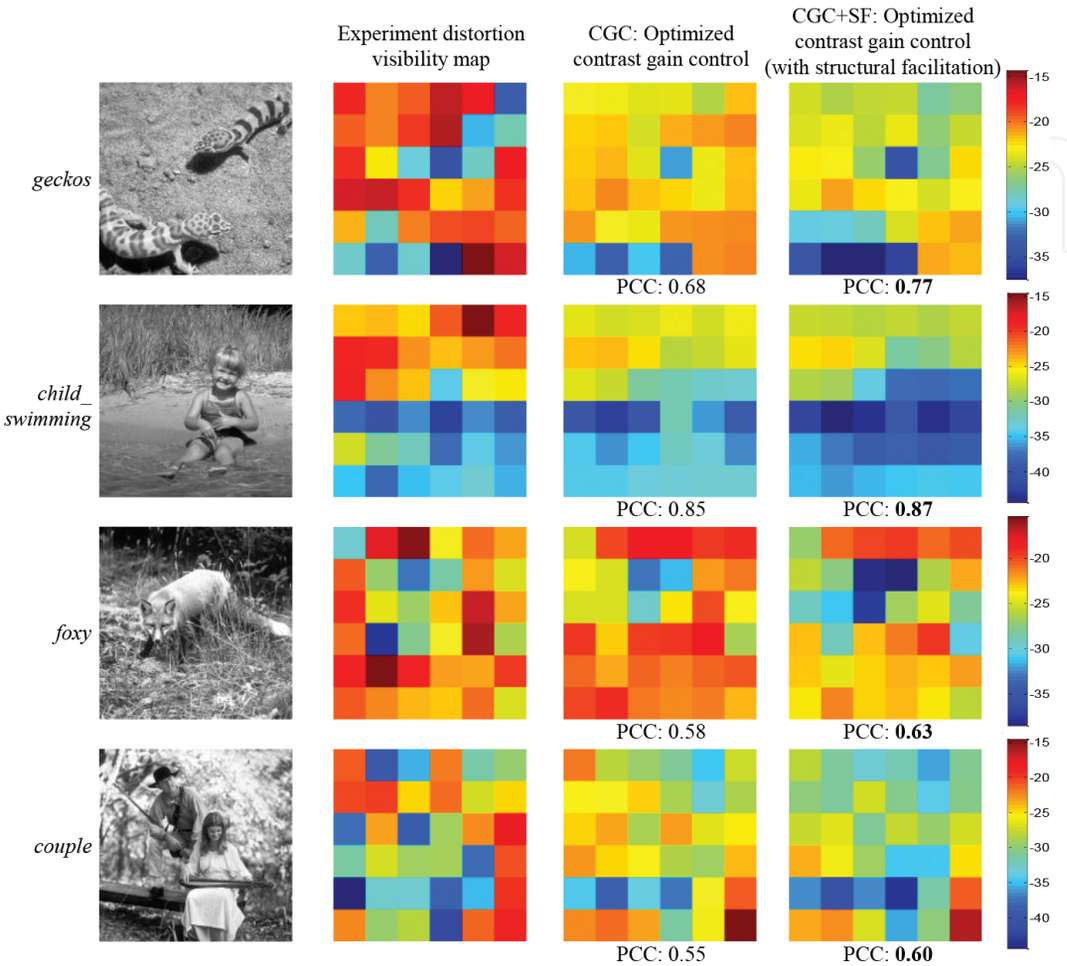

Figure 6. Structural facilitation improves the distortion visibility predictions in local regions of images containing recognizable structures. Pearson correlation coefficient (PCC) of each map with the experiment map is shown below the map.

4. Application of the masking model to compression

The masking model described in the previous section provides a way of predicting a masking map for any given input image. In this section, we show how to use this masking map to achieve visually lossless compression. In particular, we describe two different ways of incorporating the masking maps into an HEVC image coder: (1) by adjusting the QP values in HEVC on a per-block basis; and (2) by pre-adjusting the image's pixel values prior to the HEVC compression, and post-adjusting the pixel values of the decompressed image following HEVC decompression.

Similar to H.264/AVC, HEVC employs a uniform reconstruction quantizer for the transform coefficients. It is the quantization stage that introduces distortions; thus, to generate visually lossless results requires direct or indirect modification of the quantization step sizes (Q_{step} values) or quantization parameters (QP values). Previous efforts toward improved quantization have aimed at achieving higher PSNR values (e.g., [35, 36]) or other visual quality measures (e.g., [37, 38]). However, for visually lossless compression, we argue that the use of masking maps is a much better and logical alternative.

Our approach assumes that each local area within an image should have its own QP based on the amount of masking induced in that region. Note that the larger QP value is, the greater the contrast of the distortions. Therefore, the first step of our method is to predict a QP map consisting of block-based QP values, such that the resulting distortions in each corresponding block exhibit a contrast at the contrast threshold C_T. Furthermore, as we mention later in Section 5, because the predicted C_T values are underestimates of thresholds for normal viewing conditions (as opposed to the highly controlled viewing conditions used in the psychophysical experiment), we aim for QP values required to generate slightly greater than C_T (greater by at most 10 dB).

4.1. Local QP estimation from the masking map

Let QP_i denotes the QP value for the i^{th} block, and let C_i denotes the contrast of the resulting distortions. Our objective is to employ a QP_i for the i^{th} block such that the C_i for that block is given by $C_i = C_{T,i}$, where $C_{T,i}$ denotes the contrast threshold for the i^{th} block. That is, we seek the QP_i value for each block required to make the block's distortions at the threshold of visibility.

The primary difficulty in determining the relationship between C and QP is that the relationship changes depending the patch. In our previous work [39], we used a regression model to predict the relationship between QP and C on a per-block basis using statistical properties of each block as regressors. Although that approach was extremely fast, it suffered from a significant number of mispredicted QP values and thus induced distortions with incorrect contrasts. Here, we present a much more accurate solution based on the use of a pre-compression lookup table.

Specifically, prior to using HEVC, we perform the following steps:

STEP 1. Divide the image into 32 × 32 blocks (the maximum block size for HEVC).

STEP 2. Compute the 2D DCT of each block.

STEP 3. Iterate over a QP range from 1 to 51…

a. Quantize the block using a corresponding Q_{step} value given by $Q_{step} = (2^{1/6})^{QP-4}$ as specified in [40].

b. Perform an inverse 2D DCT of each block.

c. Measure and record the contrast of the resulting distortions.

In this way, for each block, we record a table that can be used to look up the closest QP_i value required to achieve $C_i = C_{T,i}$. **Figure 7** shows the lookup table values in the forms of plots (QP vs. C) for eight different image blocks. Generating the lookup table requires only a small fraction of the total time required to encode the image because only a series of inverse 2D DCTs and contrast measurements in required. Most importantly, this technique provides extremely accurate selection of the QP values.

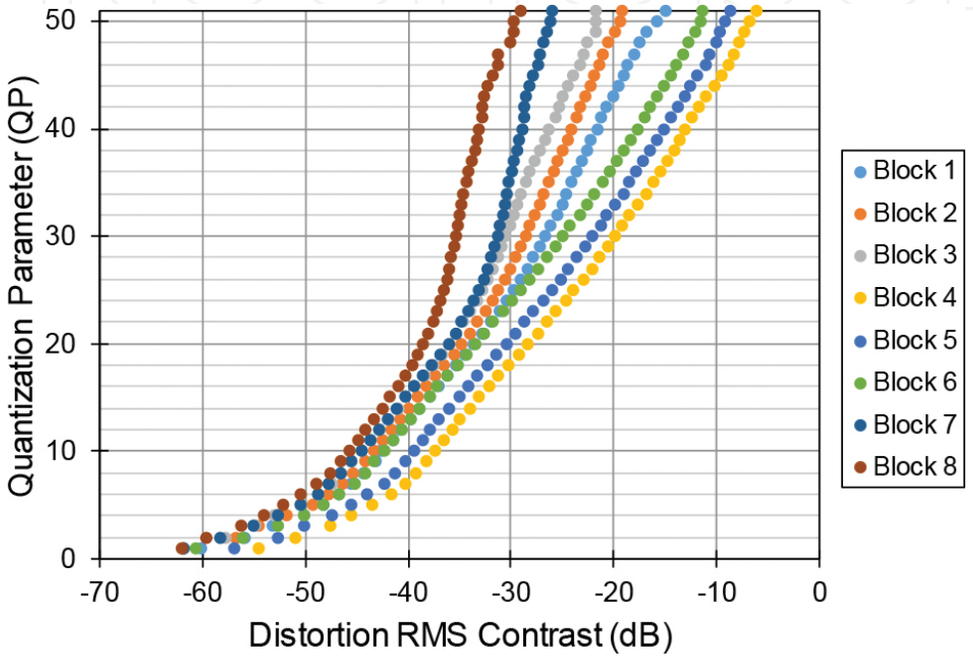

Figure 7. The relationship between distortion contrast C (in dB) and the QP used to generate that distortion for eight blocks from an image. Observe that the QP vs. C relationships are patch-specific; thus, we generate these curves (in the form of lookup tables) for all blocks prior to the compression.

4.2. Spatially adaptive quantization using the QP map

Given the QP map, we present two approaches to implement the compression. The first approach, which is the more direct approach, assigns different QP values for each 64 × 64 block. This approach was implemented by modifying the reference HEVC profile to explicitly use a separate QP value for each 64 × 64 coding unit. This approach is straightforward to implement, but it lacks some flexibility.

The other approach, which can be used with any lossy compression algorithm, effects the spatially adaptive quantization using pre-processing and post-processing stages. Let x_1 and x_2 denote the two image pixels and their corresponding quantization step sizes are denoted by Q_{s1} and Q_{s2}, respectively. The quantized values of the two pixels (denoted by $\hat{x}_1$ and $\hat{x}_2$) are then given by

http://dx.doi.org/10.5772/65362

$$\hat{x}_1 = \left\lfloor \frac{x_1}{Q_{s1}} + \frac{1}{2} \right\rfloor \cdot Q_{s1} = \beta \cdot \left\lfloor \frac{x_1/\beta}{Q_{s1}/\beta} + \frac{1}{2} \right\rfloor \cdot \frac{Q_{s1}}{\beta} \tag{5}$$

$$\hat{x}_2 = \left\lfloor \frac{x_2}{Q_{s2}} + \frac{1}{2} \right\rfloor \cdot Q_{s2} = \left\lfloor \frac{x_2 \cdot \frac{Q_{s1}}{Q_{s2}}}{Q_{s1}} + \frac{1}{2} \right\rfloor \cdot \frac{Q_{s2}}{Q_{s1}} \cdot Q_{s1} = \frac{1}{\alpha} \cdot \left\lfloor \frac{x_2 \cdot \alpha}{Q_{s1}} + \frac{1}{2} \right\rfloor \cdot Q_{s1} = \frac{\beta}{\alpha} \cdot \left\lfloor \frac{x_2 \cdot \alpha/\beta}{Q_{s1}/\beta} + \frac{1}{2} \right\rfloor \cdot \frac{Q_{s1}}{\beta} \tag{6}$$

where $\alpha = Q_{s1}/Q_{s2}$ is a scaling factor; β is a factor that normalizes the scaled pixel value (e.g., $x_2 \cdot \alpha$) into [0, 255]. Eqs. (5) and (6) indicate that different local image areas can have different quantization parameters even though the whole image is quantized using one uniform QP, as long as different image pixels are scaled properly.

For standard HEVC, the quantization step sizes relate to the QP values via $Q_{step} = \left(2^{1/6}\right)^{QP-4}$. However, in our second approach, because pixel values are quantized, we relate the quantization step to QP value through

$$Q_{step} = f\left(QP\right) = A \cdot QP^t + B, \tag{7}$$

where t is a nonlinear coefficient which aims at increasing/decreasing the QP value range within a QP map; A and B are the ratio and offset parameters which adjust the quantization step size after the nonlinear transform. The block diagram of the second approach is shown in **Figure 8**. Specifically, in the pre-processing stage, the luma channel of an image is first multiplied by a scaling map (dented by U) and then divided by β to have a range of [0, 255]. The scaling map is given by

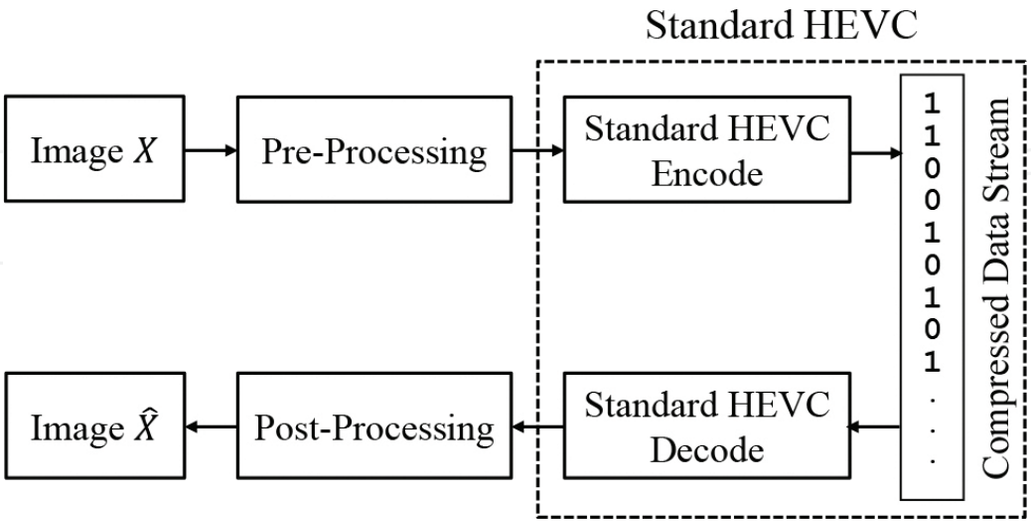

Figure 8. Block diagram of the second approach to achieve spatially adaptive quantization. Although in this chapter, we show results using HEVC as the encoder and decoder, this second approach can be used with any image compression algorithm.

$$U = \left[\frac{A \cdot Q_m + B}{A \cdot Q_1^t + B}, \frac{A \cdot Q_m + B}{A \cdot Q_2^t + B}, \ldots, \frac{A \cdot Q_m + B}{A \cdot Q_N^t + B} \right], \tag{8}$$

where Q_1, Q_2, ..., Q_N denote the QP values for N different local image areas; Q_m denotes the average value of Q_1^t, Q_2^t, $\cdots$, Q_N^t [i.e., $Q_m = \left(Q_1^t + Q_2^t + \cdots + Q_N^t\right)/N$];$\beta$ is given by

$$\beta = \max \left\{ \frac{A \cdot Q_m + B}{A \cdot Q_1^t + B} \cdot x_1, \frac{A \cdot Q_m + B}{A \cdot Q_2^t + B} \cdot x_2, \ldots, \frac{A \cdot Q_m + B}{A \cdot Q_N^t + B} \cdot x_N \right\} / 255. \tag{9}$$

In this chapter, we set $t = 2/3$, $B = 0$. Thus, U and β can be written as

$$U = \left[\frac{Q_m}{Q_1^t}, \frac{Q_m}{Q_2^t}, \ldots, \frac{Q_m}{Q_N^t} \right], \tag{10}$$

$$\beta = \max \left\{ \frac{Q_m}{Q_1^t} \cdot x_1, \frac{Q_m}{Q_2^t} \cdot x_2, \ldots, \frac{Q_m}{Q_N^t} \cdot x_N \right\} / 255. \tag{11}$$

In the post-processing stage, an inverse scaling map (dented by V) is applied to convert the scaled luminance to the original value:

$$V = \left[\frac{\beta \cdot Q_1^t}{Q_m}, \frac{\beta \cdot Q_2^t}{Q_m}, \ldots, \frac{\beta \cdot Q_N^t}{Q_m} \right]. \tag{12}$$

In standard HEVC stage, the global QP is computed by

$$QP = round \left[\left(\lambda_1 \frac{Q_m}{\beta} + \lambda_2 \right)^{\frac{1}{t}} \right], \tag{13}$$

where λ_1 and λ_2 are the linear coefficients which adjust the RMS contrast of the distortions in the compressed image to be near or below the threshold. We estimated their values by fitting the model to the 30 images in the CSIQ database, and thus, we set $\lambda_1 = 0.8$, $\lambda_2 = 2.4$.

http://dx.doi.org/10.5772/65362

Two problems can occur with this approach. First, the QP map may possibly contain zero values, in which case the above equations are not valid. Second, the predicted block-based QP maps often contain abrupt changes of QP values on the patch edges, which may possibly deteriorate the qualities of the compressed images by producing the ringing or blocking artifacts especially at lower bit compression. To solve these two problems, we first set the local zero QP values to be the minimum value among all the extra QP values within the image and then applied a Gaussian filter to the modified QP maps. As we have observed, for most natural images, the image contrast should change smoothly, not abruptly, and consequently, the resulting QP maps should also be smooth. **Figure 9** shows the *1600* image compressed using the QP map with and without the Gaussian filtering. Observe that the blocking artifacts occur in the compressed image (**Figure 9a**) if the original QP map was used; these blocking artifacts disappear when the QP map is smoothed by a Gaussian filter (**Figure 9b**).

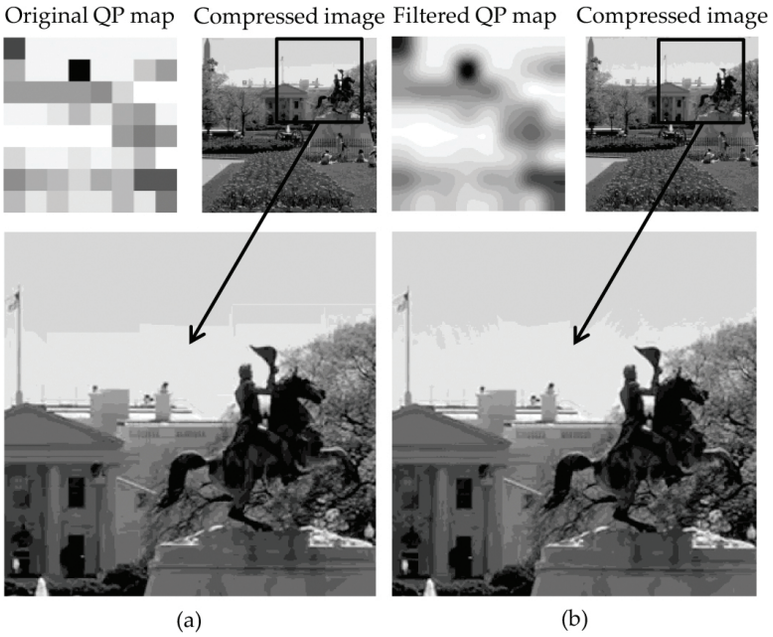

Figure 9. Gaussian filtering of the QP map improves the perceived quality of the compressed image.

In the following section, we show qualitative and quantitative results of using these two schemes with HEVC.

5. Results and discussion

In this section, we analyze the performance of the proposed visually lossless image coding algorithm. For this task, all 30 reference images in the CSIQ database were compressed at visually lossless rates using the proposed method and compared against standard HEVC. The main difference is that standard HEVC employs a uniform QP for coding the whole image, whereas our approach uses spatially adaptive QP values based on masking.

Furthermore, we have found that it is possible to induce distortions at up to 10 dB above the predicted C_T values while still yielding images which are visually lossless under normal viewing conditions. The contrast thresholds measured in the aforementioned experiment and thus the contrast thresholds predicted by the CGC+SF model are accurate for the highly controlled viewing conditions; yet, they are quite conservative for normal, everyday viewing.

5.1. QP maps

The CGC+SF model takes the 64 × 64-pixels image patch as input and predicts the distortion contrast threshold (C_T) and the corresponding threshold QP map. **Figure 10** shows the QP maps generated from the CGC+SF model for eight images in the CSIQ database.

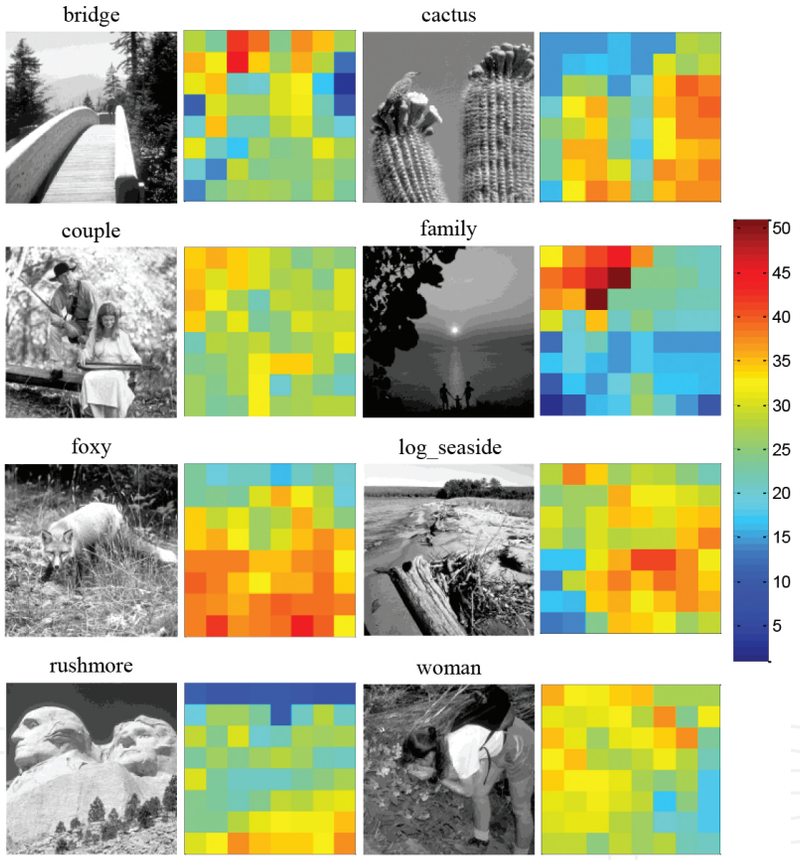

Figure 10. Eight sample reference images in CSIQ and their corresponding QP maps estimated based on CGC+SF model.

Observe that the QP maps are indeed image-adaptive; that is, the pattern of how quantization step sizes are varied across space adapts based on the image content (which is itself based on the masking model and the relationships between QP and C). In general, the QP maps specify larger quantization step sizes for regions that can mask the resulting distortions, and small quantization step sizes for regions with less masking. For example, in the *cactus* image, the bodies of the cacti impose great masking, the bird and boundaries of the cacti impose much

http://dx.doi.org/10.5772/65362

less masking, and sky has almost no masking. Accordingly, the QP values are smallest for the sky, larger for the bird and cacti boundaries, and largest for the bodies of the cacti.

Again, we remind the reader that the QP maps alone can provide only a rough gauge of how the distortions will be distributed across space. Recall from **Figure 7** that the relationship between QP and the contrast of the resulting distortion C is very much patch-specific. The same QP applied to two different blocks can give rise to vastly different distortion contrasts.

5.2. Distortion contrast maps

The proposed coding approach assumes that to compress an image in a visually lossless manner, the RMS contrast of the distortion in any compressed image region should be near or below the ground truth RMS contrast threshold. Thus, to verify the effectiveness of our proposed approach, **Figure 11** shows the contrast threshold maps (masking maps) for four sample images (as predicted by the CGC+SF model), as well as the resulting distortion contrast maps of the corresponding images coded with standard HEVC and the two proposed approaches. Note that the displayed contrast threshold maps are all 10 dB greater than predicted by the CGC+SF model due to the fact that the experimental contrast thresholds are overly conservative for normal viewing conditions. As we have found in our research, distortions with a contrast up to 10 dB above threshold can still remain visually undetectable under normal viewing conditions. Observe from **Figure 11** that images coded by standard HEVC have quite different contrast patterns with the ground truth, whereas images coded by the proposed approaches appear quite similar in pattern to the masking maps. These figures demonstrate that it is possible to achieve better compression performance than standard HEVC if using QP maps and the proposed adaptive coding scheme. We will quantify the compression performance of each method in the following section.

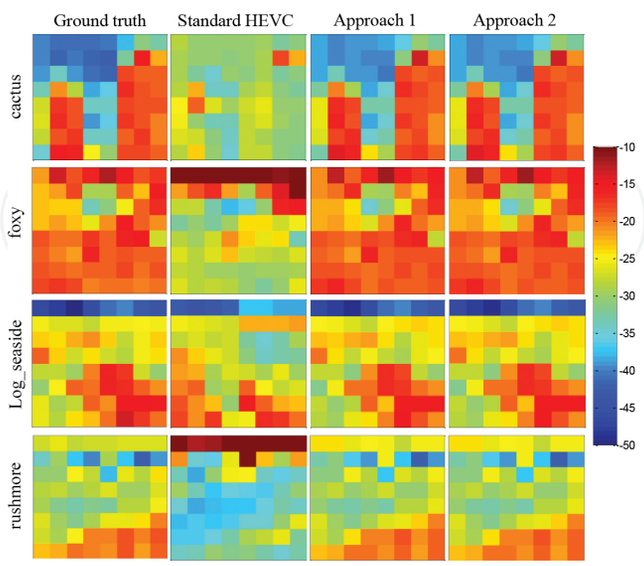

Figure 11. The ground truth RMS contrast threshold maps for four sample images, as well as the RMS contrast maps of their corresponding compressed images coded by standard HEVC, and two proposed approaches.

Image	Standard HEVC				JPEG			JPEG2000			Approach 1			Approach 2		
	QP	bpp	PSNR	Error	bpp	PSNR	Error	bpp	PSNR	Error	bpp	PSNR	Error	bpp	PSNR	Error
1600	24	1.98	40.21	733.48	3.31	45.83	563.06	2.21	45.88	594.24	1.88	39.21	676.20	1.49	41.94	786.39
Aerial_city	23	1.75	40.29	794.63	2.52	42.29	841.94	2.21	45.05	692.72	2.09	38.36	716.19	2.10	46.09	653.83
Boston	20	2.41	43.02	502.98	2.61	42.31	712.80	2.34	45.43	534.92	1.81	38.53	699.00	1.61	44.37	703.05
Bridge	22	1.73	41.97	706.05	1.46	38.41	1017.83	2.09	46.57	575.40	1.85	39.83	689.23	1.67	43.35	645.30
Butter_flower	23	1.13	42.10	1036.55	2.17	47.32	1153.96	0.60	41.29	890.83	0.87	35.58	1005.51	0.71	42.64	1282.70
Cactus	21	2.14	42.64	638.78	1.75	39.16	876.22	2.20	46.97	651.90	1.55	33.40	757.97	1.65	47.20	642.46
Child_swimming	25	2.17	38.35	577.45	3.32	43.14	438.94	2.49	43.83	426.37	1.74	33.67	715.74	1.67	43.80	641.40
Couple	25	1.60	39.22	460.92	2.74	45.07	323.70	1.47	42.61	431.27	1.21	34.76	618.40	1.45	42.21	375.91
Elk	29	1.02	35.60	740.88	1.94	40.37	616.10	1.23	39.80	665.91	1.38	34.09	660.53	1.47	42.92	521.01
Family	22	0.94	42.60	962.12	1.15	43.74	966.63	1.13	47.04	794.78	1.65	45.80	586.11	0.74	51.36	845.01
Fisher	21	1.27	42.25	936.04	1.41	43.27	1034.82	1.16	44.84	953.55	1.63	43.02	789.65	0.93	46.91	995.65
Foxy	26	2.43	37.42	415.59	3.00	39.33	427.49	2.48	42.08	434.99	1.62	29.47	680.18	2.14	42.65	374.69
Geckos	28	1.57	35.70	546.44	2.03	37.41	593.12	2.48	43.83	242.22	1.48	31.78	696.98	2.05	39.93	294.62
Lady_liberty	22	0.70	43.55	948.07	2.36	50.94	691.58	0.52	44.55	1027.35	1.26	39.56	767.69	0.58	50.48	932.29
Lake	23	3.14	40.59	433.81	2.26	34.64	937.05	3.99	48.82	420.27	2.24	31.84	634.87	2.42	39.93	517.14
Log_seaside	25	2.53	38.40	582.71	2.34	36.49	867.01	3.99	50.95	359.16	2.15	33.74	651.16	2.52	41.75	462.66
Monument	22	1.49	41.51	687.93	2.01	43.03	741.39	1.52	44.17	692.42	1.48	36.64	696.29	1.12	42.84	737.65
Native_american	23	1.57	40.63	801.26	1.60	41.28	911.87	1.98	46.68	633.27	1.70	38.14	713.70	1.62	43.99	676.87
Redwood	22	1.97	41.59	675.00	1.57	38.18	942.47	2.34	46.40	539.17	1.81	35.81	710.60	1.33	46.47	778.18
Roping	23	1.67	41.63	573.60	1.57	40.85	737.25	1.47	43.79	682.79	1.10	32.22	653.10	1.42	40.88	586.43
Rushmore	21	2.88	42.16	558.05	3.34	42.79	627.87	2.48	43.70	622.51	2.29	34.39	726.84	2.89	44.70	461.22
Shroom	19	2.35	43.94	410.47	1.52	42.36	630.61	1.28	43.89	549.87	1.25	36.79	610.10	0.79	43.71	683.36
Snow_leaves	28	1.08	37.26	675.15	1.58	38.92	787.26	1.47	41.99	612.13	1.31	35.25	548.25	1.35	40.54	519.24
Sunset_sparrow	23	1.56	40.27	1012.00	1.37	40.27	1142.99	1.52	45.13	907.72	2.04	40.10	721.38	0.91	44.04	1052.54
Sunsetcolor	22	0.32	44.57	1165.65	0.50	46.37	1166.02	0.26	47.32	1137.51	1.42	48.35	810.06	0.60	49.00	1061.84
Swarm	21	1.03	42.60	1001.60	1.42	44.85	1014.01	0.91	45.19	1024.93	1.59	41.05	813.39	0.67	44.27	1066.44
Trolley	22	2.57	41.14	482.71	3.17	41.55	589.34	3.31	47.75	387.67	1.95	34.47	668.14	1.94	43.13	581.13
Turtle	19	1.49	44.15	839.82	0.97	42.85	1038.58	1.16	46.64	866.72	1.46	40.26	778.86	1.06	44.96	880.51
Veggies	24	1.64	40.83	497.75	1.96	41.98	696.02	1.72	44.79	602.85	1.01	31.49	584.03	1.23	41.48	616.87
Woman	24	1.72	39.61	654.98	2.09	41.17	730.93	1.98	44.55	538.89	1.51	36.76	710.15	1.45	43.25	655.77
Average	23	1.73	40.86	701.75	2.03	41.87	793.96	1.86	45.05	649.81	1.61	36.81	703.01	1.45	44.03	701.07

Table 1. Performance comparison of standard HEVC, JPEG, JPEG2000, and the two proposed CGC+SF model based approaches in terms of coded rate (bpp), PSNR, and the absolute RMS contrast error.

5.3. Compression performance

Table 1 shows the compression results of 30 images using standard HEVC, JPEG, JPEG2000, and the two proposed approaches. To compare with the standard HEVC, JPEG, and JPEG2000 coding methods, a visual quality matching experiment was performed by three experienced subjects. The purpose of the experiment was to find at which compression rate, the three reference coding methods (i.e., HEVC, JPEG, and JPEG2000) yielded images with just detectable distortions; the corresponding bit-rates of these "at-threshold" compressed images were then recorded. Note that all these five coding methods only add near or below-threshold distortions, and thus judging the quality of the images is quite difficult. Although the human subjective judgment is a more reliable way for assessing the intensities of the near/below threshold distortions, we also report the PSNRs and the absolute RMS contrast errors between the reference images and the coded images for reference.

From **Table 1**, observe that the second approach of the CGC+SF model demonstrates a reduction in coded rate (bpp) by an average factor of about 16% as compared with standard HEVC, while still maintaining relatively higher PSNR values and equivalent RMS contrast errors. In comparison, the first approach seems to work less effectively. This might due to the fact that fixed local QP values are applied to the local image areas, but some local QP values are improperly estimated because of the much complex image patches and potential model limitations. However, this straightforward approach still performs competitively well, considering the relatively smaller errors it produces. For the second approach, we employed additional parameters, which indirectly adjust the coded rate to meet the visually lossless requirement. Note that for each method, the average total error is around 700 dB, which means that for each block there is an approximately 10 dB RMS contrast error (each image contains 64 blocks) compared with the ground truth. This is also attributed to the three-alternative forced-choice procedure that has been used in the experiment and mentioned in Section 5.2. Also, it should be noted that we generated the QP maps mainly from contrast masking and structural facilitation. Thus, if an image does not contain areas that can sufficiently mask the distortions, using the QP map yields no gain.

6. Conclusion

This chapter described a computational model which predicts masking maps for any given input images, and two approaches which employ the predicted masking map to achieve visually lossless compression. The proposed computational model consists of a contrast gain control model, which was trained on a database of local masking thresholds in natural images, and a structural facilitation model, which was incorporated to take into account the effects of recognizable structures on distortion visibility. Compared with standard HEVC, our approach shows an average of 16% improvement in bit-rate when testing on the CSIQ database

Author details

Yi Zhang[1], Md Mushfiqul Alam[2] and Damon M. Chandler[1*]

*Address all correspondence to: chandler.damon.michael@shizuoka.ac.jp

1 Department of Electrical and Electronic Engineering, Shizuoka University, Hamamatsu, Shizuoka, Japan

2 mPerpetuo, Inc., San Francisco, CA, USA

References

[1] R. L. DeValois and K. K. DeValois. *Spatial Vision*. Oxford University Press, 1990.

[2] D. Regan. *Human Perception of Objects: Early Visual Processing of Spatial Form Defined by Luminance, Color, Texture, Motion, and Binocular Disparity*. Sinauer Associates, Inc., Publishers, Sunderland, Massachusetts, 2000.

[3] D. M. Chandler, M. D. Gaubatz, and Sheila S. Hemami. A patch-based structural masking model with an application to compression. *J. Image Video Process*., 2009:1–22, 2009.

[4] S. Winkler and S. Susstrunk. Visibility of noise in natural images. *Electronic Imaging*. International Society for Optics and Photonics, pp 121–129, 2004.

[5] M. J. Nadenau, J. Reichel, and M. Kunt. Performance comparison of masking models based on a new psychovisual test method with natural scenery stimuli. *Signal Process.: Image Commun*., 17(10):807–823, 2002.

[6] M. M. Alam, K. P. Vilankar, D. J. Field, and D. M. Chandler. Local masking in natural images: A database and analysis. *J. Vis*., 14(8):22–22, 2014.

[7] R. J. Safranek and J. D. Johnston. A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression. *Proc. ICASSP*, 3:1945–1948, 1989.

[8] T. N. Pappas, T. A. Michel, and R. O. Hinds. Supra-threshold perceptual image coding. *Proc. ICIP,* vol. 1, pp 237–240, 1996.

[9] I. Höntsch and L. Karam. Locally adaptive perceptual image coding. *IEEE Trans. Image Process*., 9:1472–1483, 2000.

[10] M. Nadenau, J. Reichel, and M. Kunt. Wavelet-based color image compression: Exploiting the contrast sensitivity function. *IEEE Trans. Image Process*., 12:58–70, 2003.

[11] M. G. Albanesi. Wavelets and human visual perception in image compression. *Proc. ICPR*, II:859–863, 1996.

[12] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transforms. *IEEE Trans. Image Process*., 1:205–220, 1992.

[13] T. P. O'Rourke and R. L. Stevenson. Human visual system based wavelet decomposition for image compression. *J. Visual Comm. Image Repr*., 6:109–121, 1995.

[14] Y. K. Lai and C-C. J. Kuo. Wavelet image compression with optimized perceptual quality. *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, International Society for Optics and Photonics, pp 436–447, 1998

[15] A. P. Beegan. Wavelet-based image compression using human visual system models. *PhD dissertation*, Virginia Tech, 2001.

[16] Z. Wei, Y. Fu, Z. Gao, and S. Cheng. Visual compander in wavelet-based image coding. *IEEE Trans. Consumer Elec*., 44:1261–1266, 1998.

[17] Y. Zhang, E. Reinhard, and D. R. Bull. Perceptually lossless high dynamic range image compression with jpeg 2000. In *2012 19th IEEE International Conference on Image Processing*, IEEE, pp 1057–1060, 2012.

[18] A. B. Watson, G. Y. Tangand, J. A. Solomon, and J. Villasenor. Visibility of wavelet quantization noise. *IEEE Trans. Image Process*., 6:1164–1175, 1997.

[19] Z. Liu, L. J. Karam, and A. B. Watson. JPEG2000 encoding with perceptual distortion control. *IEEE Trans. Image Process*.,15(7):1763–1778, 2006.

[20] A. B. Watson and J. A. Solomon. A model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14:2379–2391, 1997.

[21] D. M. Chandler, N. L. Dykes, and S. S. Hemami. Visually lossless compression of digitized radiographs based on contrast sensitivity and visual masking. In M. Eckstein and Y. Jiang, editors, *Proceedings of SPIE Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, vol 5749. pp 359–372, 2005.

[22] H. Oh, A. Bilgin, and M. W. Marcellin. Visually lossless encoding for JPEG2000. *IEEE Trans. Image Process*., 22(1):189–201, 2013.

[23] N. N. Ponomarenko, V. V. Lukin, K. O. Egiazarian, and L. Lepisto. Adaptive visually lossless jpeg-based color image compression. *Signal, Image Video Process*., 7(3):437–452, 2013.

[24] T. Leung, M. W. Marcellin, and A. Bilgin. Visually lossless compression of windowed images. In *Data Compression Conference (DCC), 2013*, IEEE, pp 504–504, 2013.

[25] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–2394, 1987.

[26] N. Graham. *Visual Pattern Analyzers*. Oxford University Press, New York, 1989.

[27] G. E. Legge and J. M. Foley. Contrast masking in human vision. *J. Opt. Soc. Am.*, 70:1458–1470, 1980.

[28] Y. Jia, W. Lin, and A. A. Kassim. Estimating just-noticeable distortion for video. *IEEE Trans. Circuits Syst. Video Technol.*, 16(7):820–829, 2006.

[29] C. T. Vu, T. D. Phan, and D. M. Chandler. A spectral and spatial measure of local perceived sharpness in natural images. *IEEE Trans. Image Process.*, 21(3):934–945, 2012.

[30] J. M. Foley. Human luminance pattern mechanisms: masking experiments require a new model. *J. Opt. Soc. Am. A*, 11:1710–1719, 1994.

[31] J. Lubin. A visual discrimination model for imaging system design and evaluation. *Vision Models for Target Detection and Recognition 2*, World Scientific Publishing Co. Pte. Ltd. pp 245–283, 1995.

[32] S. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*. pp 179–206, 1993.

[33] E. C. Larson and D. M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *J. Electron. Imaging*, 19(1):011006, 2010.

[34] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina. An efficient algorithm for fractal analysis of textures. In *2012 25th SIBGRAPI Conference on Graphics, Patterns and Images*, IEEE, pp 39–46, 2012.

[35] E.-H. Yang and X. Yu. Rate distortion optimization for H. 264 interframe coding: a general framework and algorithms. *IEEE Trans. Image Process.*, 16(7):1774–1784, 2007.

[36] M. Karczewicz, Y. Ye, and I. Chong. Rate distortion optimized quantization. *ITU-T Q*, 6, 2008.

[37] Y. Mo, J. Xiong, J. Chen, and F. Xu. Quantization matrix coding for high efficiency video coding. In *Advances on Digital Television and Wireless Multimedia Communications*. Springer, pp 244–249, 2012.

[38] J. Chen, J. Zheng, F. Xu, and J. Villasenor. Adaptive frequency weighting for high-performance video coding. *IEEE Trans. Circuits Syst. Video Technol.*, 22(7):1027–1036, 2012.

[39] M. M. Alam, P. Patil, M. T. Hagan, and D. M. Chandler. A computational model for predicting local distortion visibility via convolutional neural network trainedon natural scenes. In *2015 IEEE International Conference on Image Processing (ICIP)*. Institute of Electrical & Electronics Engineers (IEEE), 2015.

[40] V. Sze, M. Budagavi, and G. J. Sullivan. High efficiency video coding (HEVC). *Integrated Circuit and Systems, Algorithms and Architectures*. Springer, pp 1–375, 2014.

[41] P. C. Teo and D. J. Heeger. Perceptual image distortion. *Proc. SPIE*, 2179:127-141, 1994.

[42] A. B. Watson. DCTune: A technique for visual optimization of DCT quantization matrices for individual images. *SID International Symposium Digest of Technical Papers*, vol. 24. Society for Information Display, pp 946-946, 1993.

[43] Z. Wei and K. N. Ngan. Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain. *IEEE Trans. Circ. Syst. Video Technol*., 19(3):337-346, 2009.

[44] X. H. Zhang, W. S. Lin, and P. Xue. Improved estimation for just-noticeable visual distortion. *Signal Process*., 85(4):795-808, 2005.

IntechOpen