SPATIALLY-SELECTIVE QUANTIZATION AND CODING FOR WAVELET-BASED IMAGE COMPRESSION

Matthew D. Gaubatz, Damon M. Chandler and Sheila S. Hemami

School of Electrical and Computer Engineering
Cornell University, Ithaca, New York 14853

ABSTRACT

Recent developments in psychovisual modeling have led to improvements in wavelet-based coder performance. A spatially-selective quantizer based on texture masking sensitivities is introduced, which hides distortion in high-contrast portions of images. Unlike other spatial quantization schemes, this method requires explicit side information to convey step-sizes. A simple coder is presented which leverages this side information to reduce the rate required to code the quantized data. Side information coding is also discussed. With respect to visual quality, this compression scheme performs competitively with a CSF-optimized JPEG-2000 coder at equivalent rates.

1. INTRODUCTION

Modern image compression algorithms attempt to mimic the multi-channel nature of the human visual system (HVS) by utilizing a discrete wavelet transform (DWT) front-end which separates the image into spatial-frequency and orientation components. A key advantage of this representation is that it facilitates integration of HVS properties into the quantization and encoding stages. Many wavelet-based quantization strategies are not spatially dependent, that is, a subband-specific step-size is used to quantize each wavelet band [1, 2]. Perceptual models that describe human sensitivity to subband quantization distortions distribute distortion across scales and orientations in a manner pleasing to the eye. This idea can be extended to place distortion non-uniformly across the space of one subband to further optimize visual quality.

A variety of spatially-selective quantization schemes have been proposed for compression [3, 4]. In these works, parameters used to determine step-sizes are causally predicted from previously decoded subband data, such that no side information is required to transmit the step-sizes. The EBCOT coder [5] effectively implements anti-causal spatial subband quantization, but since the step-sizes often apply to large groups of wavelet coefficients (32x32 or 64x64), side information required by the coder is minimal. A method involving more side information (and therefore a finer spatial granularity of step-sizes) is presented herein; step-sizes are uniquely designed for each wavelet coefficient, based on a texture masking model. At threshold, for the images tested, the cost of this information is roughly 0.04-0.07 bpp.

This paper is organized as follows. Section 2 outlines a perceptual test measuring human sensitivity to a variety of masked subband distortions. Section 3 discusses how the outcome is applied to quantization. A coder for compressing data quantized in this manner is presented in Section 4. Results and conclusions follow in Section 5.

2. SPATIALLY-SELECTIVE TEXTURE MASKING

A spatially-selective approach to quantization requires an understanding of the effect of natural images on the detectability of wavelet distortions. The following experiment investigates how various background textures mask quantization distortion as a function of spatial-frequency. Using a two-alternative forced-choice paradigm, contrast thresholds were measured for detection of 4.6, 9.2, and 18.4 cycles/degree wavelet subband quantization distortions presented against three texture maskers (grass, burlap, and bricks). Following an initial adaptation period, observers concurrently viewed two adjacent images placed upon a uniform background; one of the images was the mask alone; the other image additionally contained the target. Observers indicated which of the two images contained the target. Target contrasts were varied throughout the procedure and threshold was defined as the 82-percent-correct point on a Weibull function, fitted to the data following each series of trials.

Figure 1 depicts the contrast detection thresholds measured for one observer in the form of threshold-versus-contrast (TvC) curves. Data points indicate contrast detection thresholds and error bars indicate standard errors of the means over trials. Average thresholds over all textures for the minimum contrast condition are consistent with the thresholds measured previously for the same targets in the unmasked condition [6]. A two-way repeated-measures ANOVA revealed a significant effect of both texture type ($p < 0.01$) and texture contrast ($p < 0.000001$) on threshold; however most of the variability in the data was explained by texture

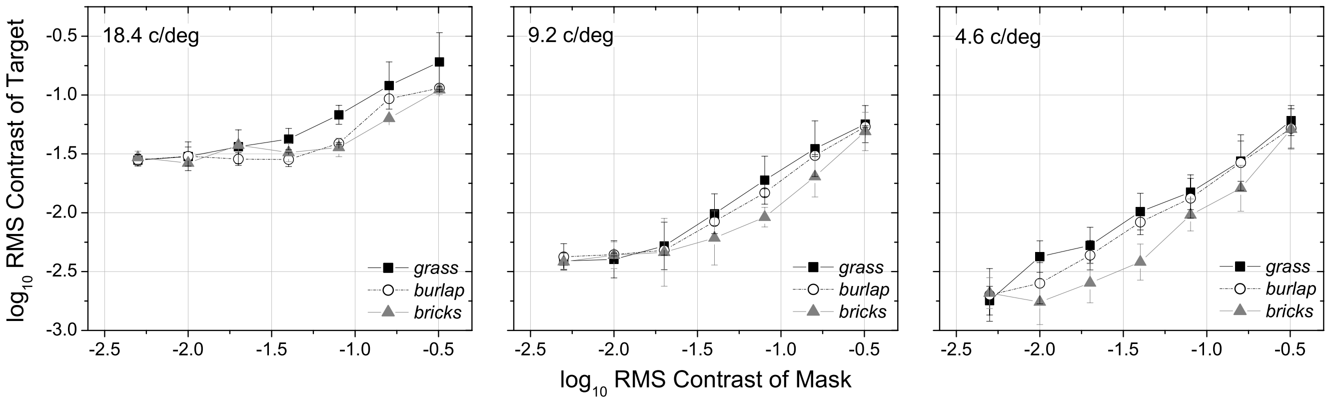


Fig. 1. Contrast thresholds for detection of wavelet subband quantization distortions presented against three texture maskers (grass, burlap, and bricks). The horizontal axis denotes the RMS contrast of the mask. The vertical axis denotes the RMS contrast of the distortions. Data points indicate contrast thresholds for one observer; error bars indicate standard errors of the means over trials.

contrast ($F > 100$ for contrast, $F < 10$ for type). The TvC curves reveal: (1) an increase in detection threshold as the contrast of the texture increased, a finding which is consistent with previous masking studies; and (2) that these textures, which contain substantially different semantic content and energy distributions, impose similar elevations in threshold at most masker contrasts.

These data were used to derive a spatially adaptive compression algorithm which estimates detection thresholds for individual image patches and then quantizes each subband coefficient such that the resulting distortion at the corresponding spatial location exhibits the predicted threshold contrast. The masked thresholds (averaged over all textures) and the previously assessed unmasked thresholds were used as part of a spatially adaptive quantization algorithm in which individual wavelet coefficients are quantized such that the resulting distortions are at the threshold of detection. The algorithm assumes that any given image patch is either (1) reasonably well-modeled as a texture, or (2) contains a low-contrast subregion to which the unmasked thresholds apply.

3. CODING STRATEGY

A common way to code quantized data involves decomposing quantization indices into several parts: a *significance map*, a map of the most-significant-bit (MSB) for each coefficient, a set of *refinement* bits, the bits in the quantization index other than the MSB, and a set of sign bits. Many wavelet coders store the significance map with bit-plane coding. With spatial quantization, this scheme is still effective, and yields an embedded bit-stream. There exists, however, a correlation between the significance map, and the step-sizes used to quantize the coefficients. The correlation is most meaningful given the locations on non-zero quantization indices. Thus, a (non-embedded) coder that first specifies the locations of these indices can more efficiently use the side information to reduce the compressed rate.

The following coder is proposed for this task. Locations of non-zero coefficients are bit-plane coded with an arithmetic coder, such as in [7]. Given this information, the significance map is entropy coded conditioned on the step-sizes (which are associated with each coefficient). Refinement bits are entropy coded as a Bernoulli process and sign bits are inserted into the stream uncoded. When creating compressed images at threshold, this maneuver (non-embedding + conditioning) can save over half the rate spent on coding the side information. Table 1 compares the rate required to code the quantized data with (1) an embedded bit-plane coder, (2) a non-embedded coder (similar to the coder above, which codes the significance map without any conditioning) and (3) the proposed coder. An embedded Tarp-filter based coder is chosen as a basis for a comparison with the state-of-the-art, and the coder without conditioning is used to accurately illustrate the effect of the conditioning. It is more effective at higher rates (which in the table correspond to compression at threshold), and decreases the compressed rate by up to as much as the rate of side information.

The side information itself consists of a map of contrasts used to generate the step-sizes. Ideally, a contrast value is associated with each spatial location (pixel) in the image. For efficiency, however, it is necessary to use average contrasts for local groups of coefficients. The collection of these average contrasts is denoted the *contrast map*. There is a trade-off between the rate required to code the contrast map, and the visual quality of an image quantized by step-

Table 1. Coder efficiency comparison between an embedded coder (Tarp), non-embedded coder (ML), and non-embedded coder conditioned on the quantization step-sizes (ML|Q).

	rate (bpp)			
image	Tarp	M—L	M—L,Q	side info
horse, 0.638 bpp	0.621	0.608	0.596	0.041
horse, 1.125 bpp	1.121	1.109	1.109	0.041
rhino, 1.239 bpp	1.221	1.214	1.167	0.072
rhino, 1.876 bpp	1.889	1.897	1.804	0.072

Table 2. Comparison between spatially-quantized and J2K coded images. The first column compares the proposed method to J2K w/VFW, while the second column compares it to standard J2K.

	# preferred			
image	proposed	J2K w/VFW	proposed	J2K
horse, 0.638 bpp	6	2	6	2
horse, 1.125 bpp	8	0	5	3
rhino, 1.239 bpp	6	2	6	2
rhino, 1.876 bpp	7	1	5	3

sizes associated with the thresholds derived from the map: the finer the granularity of a map, the more effectively spatial properties are leveraged to improve visual quality, but at the expense of a higher rate. Experimentally, associating one contrast value with each 8×8 blocks of pixels provides a reasonable trade off between step-size accuracy and rate required to communicate the side information.

A threshold must be associated with each wavelet coefficient to generate a step-size. Since each subband is a different size, the contrast map is resampled (averaged and downsampled, or interpolated) to provide a match (see Figure 2). This approach is reasonable since in the more detailed bands, the contrast thresholds generally do not vary greatly over local regions (provided that the contrast map granularity is not too coarse). The map itself is bit-plane coded after quantization in the wavelet domain. The encoder uses the reconstructed threshold map to generate step-sizes and spatially quantize individual wavelet coefficients. This information usually coded with around 5 percent of the rate required to code quantized coefficients (see Table 1).

4. RESULTS

A perceptual test was conducted to compare the performance of the proposed coder with that of JPEG-2000, with and without visual frequency weighting optimization (VFW). Test subjects were given the original image and asked which of the spatially-quantized image and JPEG-2000 (J2K) versions exhibits more perceptual distortion, when compressed at equivalent rates. Images were compared under set lighting conditions at a fixed viewing distance (three image heights). Comparisons were made at threshold, and supra-threshold, to emphasize differences between spatial and non-spatial perceptual compression.

The results of this test for several images are listed in Table 2. In both cases, the higher rate version of each image was quantized at threshold. Clearly, the proposed method is preferred in all cases. Note that JPEG-2000 with VFW tends to show slight improvements in performance at lower rates, while the standard JPEG-2000 shows slight improvements at higher rates, which is expected. The spatially quantized image pushes distortion into regions with masking textures, (see Figure 4) saving bits for details such as the horse's shoulder and harness, as well as on the fence post and in the tree. The smooth regions in Figure 4 correspond to where details of *horse* have been preserved, i.e. areas that will not effectively mask wavelet distortions.

5. REFERENCES

[1] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Transactions on Image Processing*, vol. 6, pp. 1164–1175, 1997.

[2] D. M. Chandler and S. S. Hemami, "Contrast-based quantization and rate control for wavelet-coded images," in *Proceedings on the International Conference on Image Processing, 2002*, 2002, pp. 233–236.

[3] S. Daly W. Zheng and S. Lei, "Point-wise extended visual masking for jpeg-2000 image compression," in *Proceedings on the International Conference on Image Processing, 2000*, September 2000, vol. 1, pp. 657–660.

[4] I. Honstch and L. J. Karam, "Locally adaptive perceptual image coding," *IEEE Transactions on Image Processing*, vol. 9, pp. 1472–1483, September 2000.

[5] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Transactions on Image Processing*, vol. 9, pp. 1158–1170, July 2000.

[6] 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*, vol. 20, pp. 1164–1180, 2003.

[7] P. Simard, D. Steinkraus, and H. Malvar, "On-line adaptation in image coding with a 2-d tarp fi lter," in *Data Compression Conference, 2002 Proceedings*, 2002, pp. 23–32.

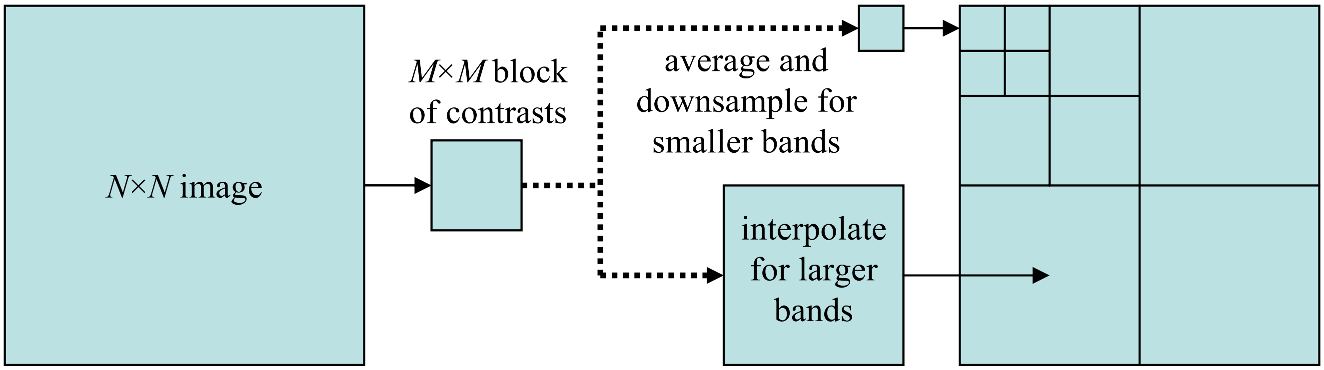
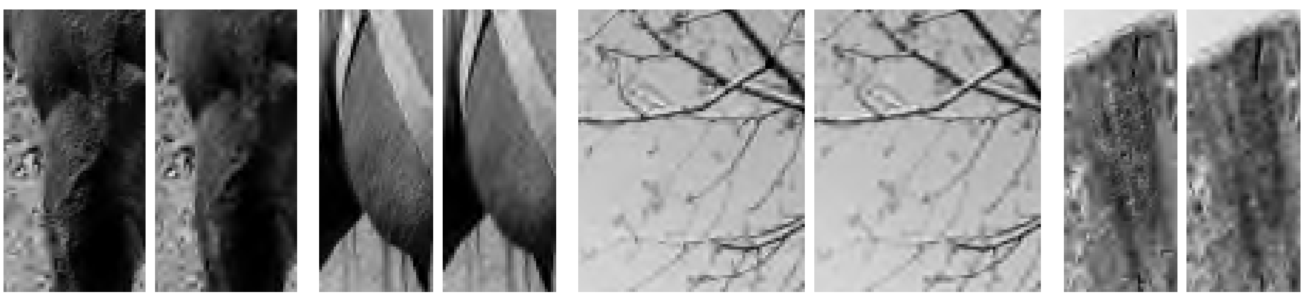
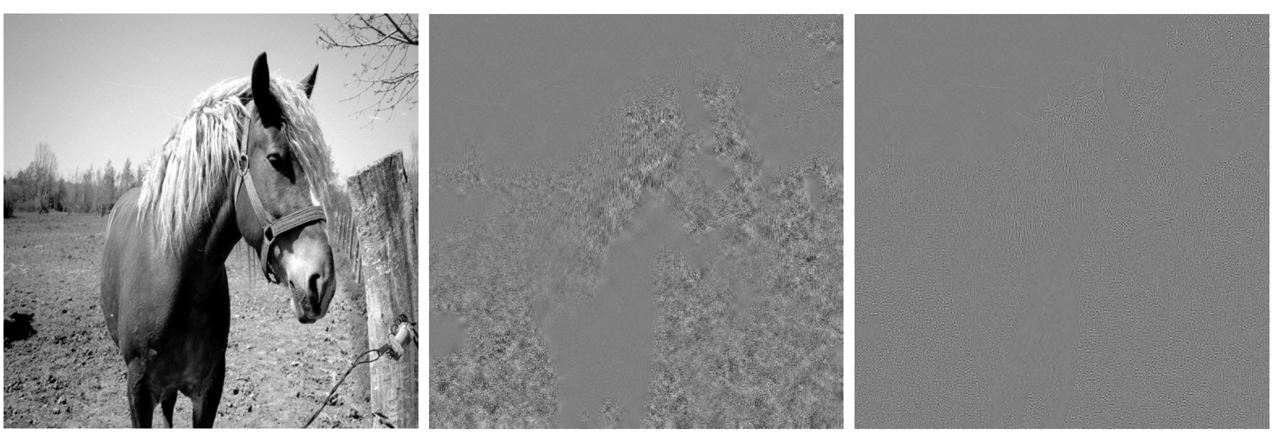


Fig. 2. Diagram of relationship between contrast map and wavelet subband coefficients. In order to associate each coefficient in a subband with a contrast, and therefore a threshold (which is needed to predict the step-size for the coefficient), the contrast map is resized in order to match the number of map entries with subband coefficients.



(a) horse's shoulder (b) harness (c) tree (d) fence post

Fig. 3. Zoomed in comparison of the shoulder 3(a), harness 3(b), tree 3(c) and fence post 3(d), regions of *horse* coded with the proposed method at 0.7 bpp. The images corresponding to the proposed compression scheme are on the left, and the JPEG-2000 images are on the right. Note that the shoulder and fence post retain more texture, the strap has less aliasing, and the branches have sharper edges in the spatially-quantized images.



(a) original *horse* (b) spatial quantization residual (c) optimized JPEG-2000 residual

Fig. 4. Residual of the spatially-quantized image (left) v. residual of the JPEG-2000 reconstruction. Note that the spatially quantized-image allocated more distortion to textured regions of the image.