SCALABLE IMAGE EMBEDDINGS FROM ARBITRARY WAVELET-BASED PERCEPTUAL MODELS

Matthew Gaubatz and Sheila S. Hemami

School of Electrical and Computer Engineering
Cornell University, Ithaca, New York 14853
{gaubatz,hemami}@ece.cornell.edu

ABSTRACT

Embedded image coding is a compression technique that yields bit streams robust to truncation. The first embedded wavelet-based compression algorithms were designed to optimize for mean-squared-error (MSE). More recent algorithms incorporate a variety of perceptually-based optimization criteria, which are implemented by selecting subband quantization step-sizes based on a target quality (and therefore rate). Since these step-sizes vary with rate, a bit stream optimized for a specific rate, if truncated, will not necessarily reflect the perceptual quality achievable at this lower rate. Constraints are given under which a wavelet-based perceptual model can generate an embedded image representation yielding perceptually derived results over a range of rates. This kind of embedding is implemented by strategically selecting the order in which subband bit planes are coded. When truncated, the resulting embedded streams provide perceptually tuned images.

1. INTRODUCTION

Embedded image coding is a compression technique that yields bit streams robust to truncation. One of the most commonly used embedded coding techniques is bit plane coding. Traditional bit plane coders transmit subband bit planes in order of bit significance (most-significant to least-significant), which minimizes MSE due to truncation. In this way, all subbands are treated with equal importance. It is well accepted, however, that minimizing MSE does not maximize visual quality. Many perceptual models have been applied to non-embedded coding, but are not suited for embedded applications; models typically generate subband quantization step-sizes for a single target rate, while embedded streams are decodable at a multitude of rates. Ideally, a perceptual embedded stream will produce perceptually-tuned images equivalent to those generated by a fixed-rate perceptual coder.

Many efforts in unifying embedding and perceptual optimization have involved modifying well-understood algorithms to incorporate perceptual considerations. For instance, locally rescaling wavelet coefficients prior to zero-tree coding can be used to improve perceptual quality [1, 2] of embedded streams. One of the more flexible schemes for embedded perceptual coding was proposed by Li [3], in which an adaptive version of this weighting scheme can be implemented by selecting the order in which subband components, such as bit planes, are coded. This visual progressive coding scheme, denoted VIP, was adopted into the JPEG 2000 standard. It was not shown, however, how to construct an optimized ordering based on an arbitrary perceptual model.

This work presents a method for creating an embedded image representation ("embedifying") based on a general class of perceptual models. Given that the model satisfies a small set of constraints, explained in section 2, a perceptual subband bit plane coding order can be derived. An algorithm for determining the order is given in section 3. This scheme can be applied to any perceptual model (as described in section 2) and can be combined with any bit plane coding technique. Bit streams generated in this fashion yield perceptually derived results, even when truncated. Section 4 illustrates this method for two perceptual models, and compares the results with traditional MSE-optimized bit plane coded streams.

2. PERCEPTUAL MODELS AND EMBEDDED CODING

Many perceptual models are implemented either 1) by tailoring subband quantization step-sizes to retain the most perceptually salient information [4–6] or 2) by performing rate-distortion optimization with a subband weighted-MSE distortion metric [7]. Note that a method of one type can be converted to the other, since for each subband MSE value, there is an associated quantization step-size that will induce the same amount of MSE, and vice versa. Any model that gives a *monotonic* relationship between subband quantization step-sizes (or subband MSE proportions) and visual quality can be used to derive a bit plane coding order.

Different perceptual models analyze quality with differ-

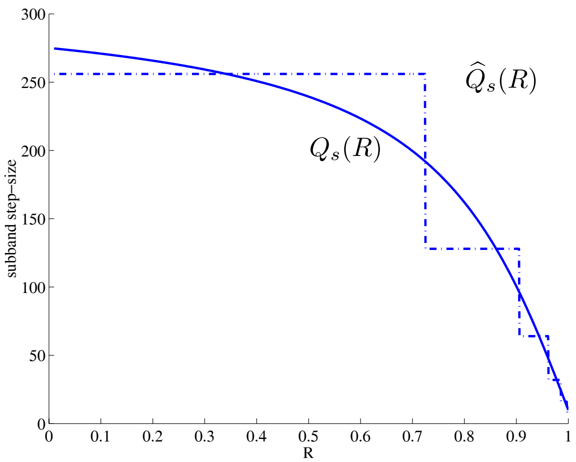

Fig. 1. Example perceptual and truncation mappings for a subband s.

ent measures of visual distortion, such as MSE or contrast of the distortion image. The relationship between quality and step-size can be re-parameterized in terms of rate. In particular, rate can be estimated from the subband step-sizes, using any number of modeling techniques. One such convenient method is given in [8], where each subband is considered a quantized version of Laplacian process. The re-parameterization places all models/distortion measures on a standard setting, simplifying discussion and analysis.

Let R represent *normalized rate*, i.e. a fraction of the overall rate required to losslessly represent an image with a given coder. For an image composed of subbands with all coefficients quantized to zero, $R = 0$. For the unquantized wavelet coefficients, $R = 1$, since the image is already fully described. Let s denote a wavelet subband. Any perceptual model, as described above, will yield a *perceptual mapping* between R and quantization step-size, $Q_s(R)$ for each band. In a compression system with a 5-level wavelet transform, a usable perceptual model will generate 16 such scalar maps, one for the low-frequency band, and 15 for all the high-frequency bands.

3. BIT PLANE CODING ORDER

Embedded bit plane coders naturally yield mappings between normalized rate and quantization step-size. These functions are denoted $\widehat{Q}_s(R)$ and are referred to as *truncation mappings*. Truncating an embedded image representation effectively induces a quantization step-size on each band. The functions $\widehat{Q}_s(R)$ are monotonically decreasing and piecewise-dyadic-constant. Specifically, for every $R_1 < R_2$,

$$\widehat{Q}_s(R_1) = \widehat{Q}_s(R_2) \cdot 2^k, k \geq 0 \in \mathbb{Z}. \quad (1)$$

Note $\widehat{Q}_s(0) = 2^K \cdot \delta_s, K \in \mathbb{Z}^+$, where $\delta_s \in \mathbb{R}^+$ is the base quantization-step size. For $\delta_s \neq 1$, the subband is first (lossily) quantized with a step-size of δ_s. Example perceptual and truncation mappings are given in Figure 1.

A set of truncation mappings describe a subband bit plane coding order completely. As an image is coded, R increases from 0 to 1, and the order in which the discontinuities appear over all s is the order that subband bit planes are coded. The relative positioning of these discontinuities and not the absolute relationship between $\widehat{Q}_s(R)$ and R determines this order. Thus, every collection of truncation mappings (for all subbands), $\{\widehat{Q}_s(R)\}$, can be associated with a coding order.

The problem of determining a perceptual coding order can now be more formally stated. Given a perceptual model, for each subband s and associated perceptual mapping, construct a truncation mapping, which minimizes the average distance between $Q_s(R)$ and $\widehat{Q}_s(R)$. This criterion is designed to minimize per subband degradation in perceptual quality due to truncation. Mathematically, this task corresponds to solving the following, for each s:

$$\widehat{Q}_s(R)^* = \underset{\widehat{Q}_s(R)}{\operatorname{argmin}} \; ||Q_s(R) - \widehat{Q}_s(R)||_p, \quad (2)$$

where $||\cdot||_p$ denotes the p-norm. Assuming that $\forall k \in \mathbb{Z}, \exists R$ such that $Q_s(R) = \widehat{Q}_s(R) = 2^k \cdot \delta_s$, it can be shown that $\widehat{Q}_s(R)^*$ is the same using any p-norm. For any set of mappings, the solution for each $\widehat{Q}_s(R)$ is equal to $Q_s(R)$ rounded to the nearest dyadic multiple of the base quantizer step-size:

$$\widehat{Q}_s(R) = \delta_s \cdot 2^{\lfloor log_2(\frac{4 \cdot Q_s(R)}{3 \cdot \delta_s}) \rfloor}, \quad (3)$$

Figure 2 depicts the relationship between discontinuities in a set of $\widehat{Q}_s(R)$ mappings and the bit plane coding order. Note that to establish the relative order among all bands, in a system with a 5-level wavelet transform, the discontinuities in 16 truncation mappings must be compared. Table 1 compares bit-significance (MSE) coding order with the order depicted in Figure 2. The following algorithm computes the order in which the discontinuities occur as R increases in a complete set of truncation mappings. Note that the actual truncation mappings resulting from this coding order will resemble the desired mappings, though these sets might not be identical. Let M be the largest coefficient magnitude among all subbands:

Algorithm: Determination of bit plane coding order

i) set $R = 0, k_s = \lfloor log_2(M) \rfloor \forall s$
ii) set $\Delta \ll 1$ such that $\forall R, \widehat{Q}_s(R) > \widehat{Q}_s(R + \Delta)$ for at most one subband s
iii) $\forall s$ compute $\widehat{Q}_s(R)$

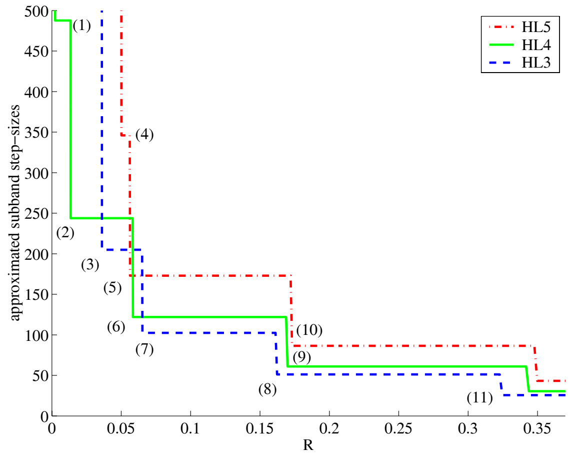

Fig. 2. Example truncation mappings for HL5, HL4 and HL3 bands derived from a distortion-contrast model.

if $\widehat{Q}_s(R) > \widehat{Q}_s(R + \Delta)$ and $k_s > 0$
code bit plane k_s of subband s
set $k_s = k_s - 1$
iv) if $k_s = 0 \; \forall s$
end
else
set $R = R + \Delta$ and goto iii)

A perceptually optimized order can thus be determined with a simple algorithm that runs at the encoder and the decoder. The required coding overhead includes M as well as any other auxiliary information needed by the perceptual model.

4. RESULTS

The bit plane ordering algorithm described in section 3 can be applied to many perceptual models. Here, two examples are given. Recent work [6] involves quantifying the perceived visual response to compound wavelet distortions. This technique maps visual distortion, computed as root-mean-squared error in the luminance domain, to a distribution of MSE among wavelet subbands. The resulting mapping between contrast and subband MSE can be formulated in terms of rate and quantization step-size, as in section 2, to generate $Q_s(R)$. Coding overhead consists of per band standard deviation and kurtosis, which is used at reconstruction to derive $\widehat{Q}_s(R)$. These statistics may be inserted in the bit stream only when it becomes necessary to evaluate $\widehat{Q}_s(R)$ for a new subband.

The JPEG 2000 standard includes the option to scale subband coefficients prior to coding with weights based on a fixed viewing distance and a perceptual contrast sensitivity model [7]. These weights are inversely proportional to quantization step-size, and since rate can be estimated from step-size, each set of weights can be associated with a set of curves mapping rate to step-size. Note, however, that since the weights have been optimized for a specific rate, in a sense, these curves are meaningful around a specific rate. In [3], Li generates images using 2 sets of weights. Below 0.125 bpp, one set of weights is used to control bit plane order, while above this threshold uniform weights are employed. To utilize the algorithm described in section 3, intermediate weights can be effectively implemented for a set of $Q_s(R)$ curves created as linear combinations of the rate v. step-size curves from each set of weights. Though the visual weights used in JPEG-2000 are the same for all images, it is still necessary to transmit the bit plane coding order, since the relationship between rate and step-size varies from image to image.

Embedded bit streams are constructed with a 5-level 9/7 wavelet transform of 512×512 greyscale 8-bit images, using the models described above to determine coding order. The bit planes are independently coded with a Tarp-filter based arithmetic coder [9]. MSE-embedded perceptual streams, targeted for quality around 1 bpp, are similarly created based on the same models, where subbands are coded in traditional bit plane (MSE) order. Using each model, perceptually-embedded and MSE-embedded streams are truncated and compared at a set of rates.

Images are compared in a perceptual test under set lighting conditions at a fixed viewing distance (three image heights). Test subjects are given the original image and asked which image (the perceptually-ordered or MSE-ordered represen-

Table 1. MSE-based coding order compared with the perceptual order established in Figure 2.

coding order	MSE			distortion-contrast		
subband	HL5	HL4	HL3	HL5	HL4	HL3
bit plane 1	1	5	9	4	1	3
bit plane 2	2	6	10	5	2	7
bit plane 3	3	7	11	10	6	8
bit plane 4	4	8	12	13	9	11

Table 2. Comparison of number of subjects preferring the perceptually-ordered versus MSE-ordered representations.

coding order	perceptual	MSE-based
distortion-contrast, 0.134 bpp	6	1
distortion-contrast, 0.090 bpp	7	0
CSF-based, 0.134 bpp	5	2
CSF-based, 0.090 bpp	4	3

Fig. 3. embedded (left) and MSE-ordered (right) coding of *barbara* based on a distortion-contrast model at 0.090 bpp.

Table 3. Coder fractional rate comparison (per scale) using *barbara* at 0.090 bpp for perceptual and MSE coding orders.

scale	5	4	3	2	1
perceptual distortion-contrast	.49	.51	.00	.00	.00
MSE distortion-contrast	.34	.47	.19	.00	.00
perceptual CSF-based	.22	.34	.44	.00	.00
MSE CSF-based	.34	.42	.19	.05	.00

tation) displays less perceptual distortion. The results of this test for *barbara* are depicted in Table 2. Clearly, the perceptually-embedded stream is preferred to the MSE-ordered perceptual representation, especially at low rates.

Table 3 compares the distribution of bits among scales for perceptually-ordered and MSE-ordered streams truncated at 0.090 bpp, and Figure 3 illustrates the distortion-contrast comparison at 0.090 bpp. Differences are most noticeable in the tablecloth, the object on the table, the shawl, and the face. The right-hand image contains sporadic detail that almost appears as noise at this low rate, while the image on the left more accurately preserves the overall image structure, especially in the face and the object on the table. The bit distributions reveal that MSE-ordered streams spend bits on higher frequency information, at the cost of blurring various structural elements in the scene. Based on the results of the perceptual test, the structural information plays a more important role in determining visual quality at low rates.

5. REFERENCES

[1] I. Honstch, L. J. Karam, and R. J. Safranek, "A perceptually tuned embedded zerotree image coder," in *Proceedings on the International Conference on Image Processing, 1997*, October 1997, vol. 1, pp. 41–44.

[2] M. Ramos and S. S. Hemami, "Activity selective spiht coding," in *SPIE Visual Communications and Image Processing*, January 1999.

[3] J. Li, "Visual progressive coding," in *SPIE Conference on Visual Communications and Image Processing*, January 1999, vol. 3653, pp. 1143–1154.

[4] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post-quantization data compression," in *Proceedings on Acoustics, Speech and Signal Processing, 1989*, May 1989, pp. 23–26.

[5] 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.

[6] 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.

[7] "Itu-t rec. t.800 — is 15444-1:2000, information technology - jpeg 2000 image coding standard," .

[8] M. J. Gormish and J. T. Gill, "Computation-rate-distortion in transform coders for image compression," in *SPIE Visual Communications and Image Processing*, 1993.

[9] 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.