

Copyright 2000 IEEE. Published in the 2000 International Conference on Image Processing (ICIP-2000), scheduled for September 10-13, 2000 in Vancouver, BC. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: Manager, Copyrights and Permissions / IEEE Service Center / 445 Hoes Lane / P.O. Box 1331 / Piscataway, NJ 08855-1331, USA. Telephone: + Intl. 908-562-3966.

POINT-WISE EXTENDED VISUAL MASKING FOR JPEG-2000 IMAGE COMPRESSION

Wenjun Zeng, Scott Daly and Shawmin Lei

Sharp Laboratories of America

Emails:{zengw, daly, shawmin}@sharplabs.com

ABSTRACT

One common visual optimization strategy for image compression is to exploit the visual masking effect where artifacts are locally masked by the image acting as a background signal. In this paper, we present a point-wise extended visual masking approach that nonlinearly maps the wavelet coefficients to a perceptually *uniform* domain prior to quantization by taking advantages of both self-contrast masking and neighborhood masking effects, thus achieving very good visual quality. It is essentially a coefficient-wise adaptive quantization without any overhead. It allows bitstream scalability, as opposed to many previous works. The proposed scheme has been adopted into the working draft of JPEG-2000 Part II.

1. INTRODUCTION

One major goal of image compression is to remove the statistical redundancy in the image data. Given a target bit rate, image compression techniques try to minimize the distortion (usually measured in mean square error (MSE)). Another goal of image compression is to try to remove the perceptual irrelevancy. It is well known that MSE is usually not a good measure for visual quality. It is therefore important that the compression scheme takes into account the properties of the human visual systems (HVS) in the optimization process.

One common visual optimization strategy for compression is to make use of the contrast sensitivity function (CSF) that characterizes the varying sensitivity of the visual system to 2D spatial frequency [1][2][3]. The advantage of this technique, however, becomes less noticeable for lower resolution display and closer viewing distance, since the CSF curve tends to be flat under those viewing conditions.

Visual masking is a perceptual phenomenon where artifacts are locally masked by the image acting as a background signal. Early work [4] scaled the overall quantization values as a function of local image variance. Such adaptive methods required overhead to tell the decoder what quantizer was used to encode a local region.

To avoid excess overhead, these schemes often significantly restrict local adaptation [4][5].

It is understood nowadays that the masking property of human vision primarily occurs *within* spatial frequency channels that are each limited in radial frequency as well as orientation. It is then possible to exploit the masking effects by *nonuniform* quantization which quantizes more coarsely as a function of the activity in spatial frequency and spatial location, as opposed to overtly adaptive techniques [4]. Since these masking effects are approximately the same in each channel, once normalized, the same masking procedure could be used in each channel without incurring any overhead [6]. One way [6] to exploit this masking effect (we will refer to this as self-contrast masking effect) is to let the CSF-normalized transform coefficients go through a *point-wise* nonlinear transducer function such as a power function, prior to *uniform* quantization. This effectively results in a non-uniform quantization of the original coefficients where coefficients with larger amplitude are more coarsely quantized.

In [7], an algorithm that locally adapts the quantizer step size at each coefficient according to an estimate of the masking measure is presented. To eliminate the overhead, it exploits the self-contrast masking based on an estimate of the current coefficient from neighboring already coded coefficients. The estimate, however, may not be accurate given that the coefficients are pretty much de-correlated. It is not amenable to scalable coding.

Another way of exploiting visual masking in the context of JPEG-2000 [8] - an emerging wavelet-based standard for still image compression, is through control of individual code-block contribution [9]. This approach takes advantage of the structure of JPEG-2000 - coefficients in each wavelet subband are divided into blocks of same size (called code-block) and each code-block is embedded coded independently. In this approach, the embedded coding of each code-block (usually with size no less than 32x32) is performed without considering visual masking effect. However, in the post-compression optimization process where sub-bitstreams from each code-block are assembled in a rate-distortion (R-D) optimized order to form the final bitstream, the distortion

metric is modified to take into account the visual masking effect. More specifically, the distortion of each coefficient is weighted by a visual masking factor that is in general a function of the neighboring coefficients in the *same* subband (we will refer to this type of masking as block-based neighborhood masking). The weakness of this block-based neighborhood masking approach is that it can only adjust the truncation points of the bit-stream of each code-block, which is a spatially coarser adjustment than the sample-by-sample compensation offered by the self-masking approach [6].

In this paper, we present a point-wise extended masking approach [10] that has been recently adopted into the working draft of JPEG-2000 Part II. The proposed masking approach nonlinearly maps the wavelet coefficients to a perceptually *uniform* domain prior to quantization by taking advantages of both self-contrast masking and neighborhood masking effects, thus achieving very good visual quality. It is essentially a coefficient-wise adaptive quantization without any overhead. It allows bitstream scalability, as opposed to many previous works [4][5][7]. Furthermore, our approach allows optimally distributing available bits across space and spatial frequency to minimize visual distortion, as opposed to many previous works that generally focus on perceptually lossless compression.

2. POINT-WISE “EXTENDED NON-LINEARITY”

In this section, we describe our proposed point-wise extended masking approach [10] that exploits both self-contrast-masking and neighborhood-masking. We develop a visual masking model that extends the point-wise “non-linearity” of [6] to a point-wise “extended non-linearity” that also takes care of the masking effect contributed from spatially neighboring coefficients. This is to overcome the over-masking problem of the self-masking approach [6] that occurs at diagonal edges. The main advantage of this strategy is its ability to distinguish between large amplitude coefficients that lie in a region of simple edge structure and those in a complex region. This feature will assure the good visual quality of simple edges in a smooth background, which is often critical to the overall perceived quality.

The proposed point-wise extended masking approach treats visual masking as a combination of two separate processes, i.e., self-contrast masking and neighborhood masking. The first step is to apply a point-wise non-linear transducer function $f(\cdot)$ such as a power function to the original coefficient x_i , i.e.,

$$x_i \rightarrow y_i = f(x_i) \quad (1)$$

This step assumes each signal with which a coefficient is associated is lying on a common flat background. Under this assumption, $\{y_i\}$ are perceptually uniform. In a real

image, however, this is usually not the case. Each signal is superimposed on other spatially neighboring signals. There is some masking effect contributed from spatially neighboring signals due to the phase uncertainty, receptive field sizes, as well as possible longer range effects (“pooling”). To further exploit this neighborhood masking effect, the second step normalizes y_i by a neighborhood masking weighting factor w_i which is a function of the amplitudes of the neighboring signals, i.e.,

$$y_i \rightarrow z_i = y_i / w_i = f(x_i) / g(N_i(\{x_k\})) \quad (2)$$

where w_i is a function $g(\cdot)$ of the neighboring signals denoted in a vector form as $N_i(\{x_k\})$. The neighboring coefficients are in the same subband. They could also be coefficients around the same spatial location but in other frequency bands. This, however, is not in current implementation. As discussed above, the second step is especially important for wavelet based systems where over-masking may result from the first step.

Fig. 1 shows the system diagram for the point-wise extended masking approach. To avoid overhead and to make sure the inverse process is feasible, quantized versions of the neighboring coefficients that are also available at the decoder will be used, and the neighborhood has to be *causal* in the sense that each coefficient x_k in this neighborhood has to be recovered before the current one x_i so that the decoder can perform exactly the same operation to reconstruct w_i . A good model we have developed and has been adopted by JPEG 2000 is to use the non-linear transform

$$z_i = \text{sign}(x_i) |x_i|^\alpha / (1 + a \sum_{\{k \text{ near } -i\}} |\hat{x}_k|^\beta / |\phi_i|)$$

where $|\phi_i|$ denotes the size of the causal neighborhood, a is a normalization factor, \hat{x}_k denotes the quantized neighboring coefficients, and the neighborhood contains coefficients in the same band that lie within an $N \times N$ window centered at the current coefficient and also appear earlier than the current coefficient in the raster scan order (see Fig. 2 for an example). The neighborhood does not include the current coefficient itself so that an explicit solution for the inverse process is available. α assumes a value between 0 and 1, and is used to control the degree of self-masking. A typical value of α is 0.7. β assumes a positive value, and, together with N , are used to control the degree of neighborhood masking. β and N play important roles in differentiating coefficients around simple edge from those in the complex area. N controls the degree of averaging; β controls the influence of the amplitude of each coefficient. It is important that β be chosen as a value much smaller than 1. A good value of β is 0.2. This is quite different from some previously proposed variance-based neighborhood activity measure [4][7]. It helps to protect coefficients around simple sharp edges, since the coefficients around sharp edges usually

have high values. A variance-based measure may not be able to distinguish a local sharp edge area (with few large coefficients and all the rest close to zero) from a local complex area (with many mid-amplitude coefficients). This is because the large coefficients, although few, in a local sharp edge area could contribute significantly to the overall variance, due to the square operation. Note that masking is lower than expected near sharp edges (as opposed to textures) due to “pooling”. A small value of β suppresses the contribution of a few large coefficients around sharp edges to the masking factor, thus implicitly distinguishing coefficients around sharp edges from coefficients in a complex region. For example, two neighborhood sets of $\{5, -5, 5, -5, 5, -5, 5, -5\}$ and $\{0, 0, 0, 10, -10, 0, 0, 0\}$ have the same variance of 5. But their “0.2-norms” are 1.38 and 0.40, respectively.

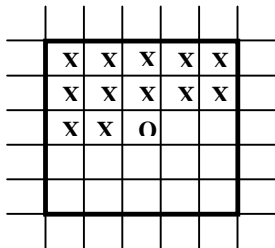


Fig. 2: Causal neighborhood ($N=5$, $|\phi_i|=12$). “o”: current coefficient; “x”: coefficients in the causal neighborhood.

Note that quantized neighboring coefficients will be used at the encoder to ensure the invertibility at the decoder. For non-scalable coding, the encoder will just use the final quantized neighboring coefficients to calculate the neighborhood masking factor. For embedded coding, unfortunately, this is not feasible because the nonlinear transform is performed prior to scalable compression, and the decoder can have any bitstream that has a lower rate than the final rate. Nevertheless, the discrepancy of the values of w_i calculated at the encoder and the decoder can be completely eliminated or reduced by a conservative strategy (“worst case” consideration) where only the *same* very coarsely quantized coefficients are used to calculate w_i at both the encoder and the decoder. For example, after z_k is quantized (for bit-plane coding), the n most significant bits of the quantization index will be retained (the rest replaced with 0). This modified quantization index is then dequantized and used for calculating w_i . As long as n is small enough (with respect to the available bits at the decoder), the decoder will be able to get exactly the same quantized version of the neighboring coefficients. The compromise here is a coarser granularity of w_i which may slightly affect the accuracy of the masking model. But our experiments show that the performance usually is not very sensitive to the accuracy of the quantized neighboring coefficients.

3. EXPERIMENTAL RESULTS

The point-wise extended masking approach has been implemented in the JPEG-2000 VM software [11]. The power functions in the “extended non-linearity” are realized using look-up-tables to reduce the complexity. To compare the performance of different visual masking approaches, no CSF weighting is used in the testing. This is reasonable for monitor display condition where the viewers are allowed to zoom in to see the details.

We tested on a set of 14 test images, among which “bike” and “woman” are in the JPEG-2000 standard test image set, while the rest are high quality photographic face images. The image size ranges from 512x640 to 2048x2560. We compare the performances of no masking, self-masking (with $\alpha=0.7$), block-based neighborhood masking [9] (with $\rho=0.5$), and the proposed point-wise extended masking (with $\alpha=0.7$, $\beta=0.2$, $N=13$, $n=9$). For all four approaches, scalable coding based on deadzone quantizer [11] is tested, i.e., the bitstream is encoded at 2 bpp, and truncated to get decoded images at 0.25, 0.5, 0.75 and 1 bpp. Note that for all three masking methods, the parameters used are the default or recommended parameters. Different choices of the parameters may result in slightly different performances.

The independent visual evaluation [10] suggests that for all test images at different bit rates, the proposed point-wise extended masking provides the best overall visual quality. The difference is significant in many cases. For almost all test images, self-masking provides slightly better visual quality than the block-based neighborhood masking for middle to high range bit rates (≥ 0.75 bpp). For relatively low rates (0.25-0.5 bpp), there are some artifacts generated by self-masking that make it often inferior to block-based neighborhood masking. The self-masking approach usually protects the fine texture well, but has problems with sharp edges, especially at low bit rates. The block-based neighborhood masking approach tends to smooth out the fine texture, but protects the areas immediately surrounding the high contrast edges well. It also does not seem to work well on relatively small images, mainly due to its block-based nature. However, it has successful performance for large images with diverse content. The point-wise extended masking approach combines the strength of both, thus resulting in mutual synergism. Fig. 3 shows some examples for the performance comparisons. The performance of the point-wise extended masking method is quite consistent, and is much better for this particular case. For some complex images with diverse content such as “woman”, the visual

improvement over no masking case can be equivalent to a saving of up to 50% in bit-rate. For more details about the performance of different masking methods for different bit rates and different images, see [10].

The proposed point-wise extended masking can also be used in conjunction with other quantization schemes such as TCQ. Our preliminary study showed that TCQ alone can achieve similar feature as self-masking (protect fine texture). In light of this property, the neighborhood masking component is more important than the self-masking when TCQ is used (thus α could be set to 1 to de-emphasize self-masking).

4. REFERENCES

- [1] S. Daly, "Application of a noise-adaptive contrast sensitivity function to image data compression," *Optical Engineering*, vol. 29, pp.977-987, 1990.
- [2] H. A. Peterson, A. J. Ahumada, Jr., and A. B. Watson, "Improved detection model for DCT coefficient quantization," *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, pp.191-201, Feb. 1993.
- [3] Watson, Yang, Solomon and Vilasenor, "Visibility of wavelet quantization noise," *IEEE Tran. Image Proc.*, vol. 6, No.8, pp. 1164-1175, 1997.
- [4] R. J. Safranek and J.D. Johnston, "A perceptually-tuned sub-band image coder with image dependent quantization and post-quantization data compression ", *IEEE Inter. Conf. Acoustic, Speech, and Signal Process.*, pp. 1945-1948, 1989.
- [5] A. B. Watson, "DCT quantization matrices visually optimized for individual images," *Proc. SPIE Conf. Human Vision, Visual Processing, and Digital Display IV*, vol. 1913, pp. 202-216, 1993.
- [6] S. Daly, W. Zeng, J. Li, and S. Lei, "Visual masking in wavelet compression for JPEG2000," in *IS&T/SPIE Conf. Image and Video Communications and Processing*, vol. 3974, Jan. 2000.
- [7] I. Hontsch and L. Karam, "APIC: adaptive perceptual image coding based on subband decomposition with locally adaptive perceptual weighting," *Proc. IEEE Inter. Conf. Image Proc.*, pp. 37-40, 1997.
- [8] "Information Technology – JPEG 2000 Image Coding System," ISO/IEC FCD15444-1: 2000 (V1.0, Mar. 2000).
- [9] David Taubman, "High performance Scalable Image Compression with EBCOT", to appear in *IEEE Transactions on Image Processing*, 2000.
- [10] W. Zeng, S. Daly and S. Lei, "Report on CE V1 (Improve Visual Quality by Making Use of Both Self-Contrast-Masking and Neighborhood-Masking)," ISO/IEC JTC1/SC29/WG 1 N1452, Maui, Hawaii, Dec. 1999.
- [11] "JPEG 2000 Verification Model 7.0 Software", ISO/IEC JTC1/SC29/WG1 N1685, April 2000.

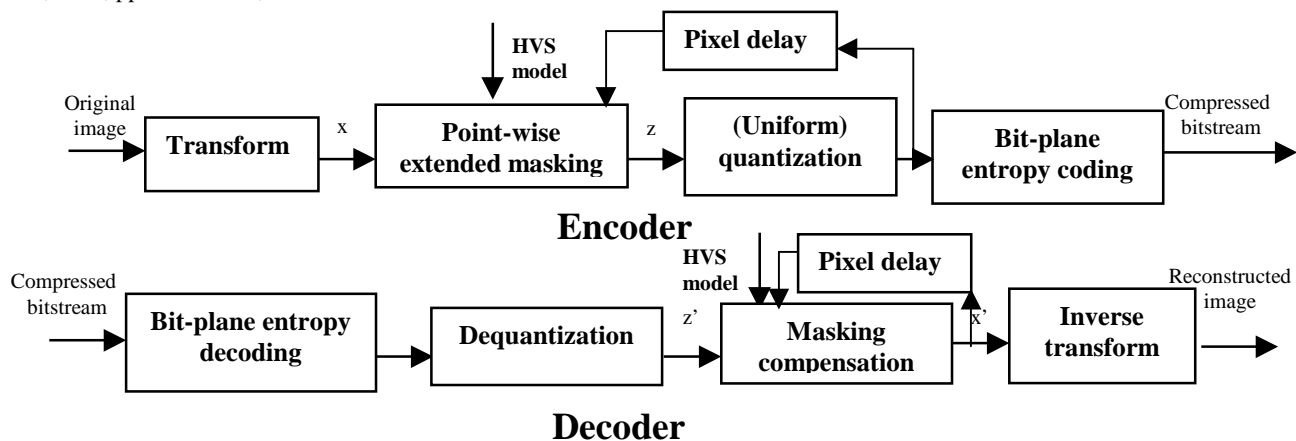


Fig. 1: System diagram for the point-wise extended masking approach.

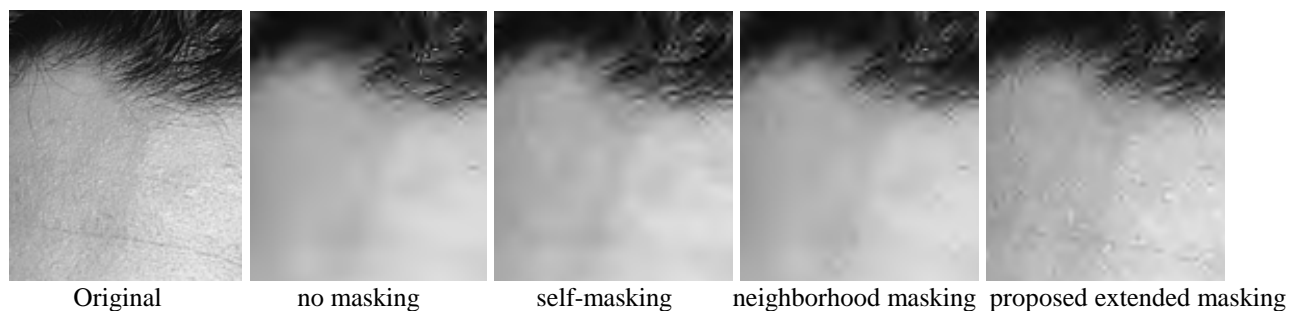


Fig. 3: Comparison of different masking methods at 0.25 bpp. Foreheads of the 512x640 image "face_guy_down2"