

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.

VISUAL OPTIMIZATION TOOLS IN JPEG 2000

Wenjun Zeng, Scott Daly and Shawmin Lei

Sharp Laboratories of America

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

ABSTRACT

In this paper, we review the various tools in JPEG-2000 that allow the users to take advantages of the various properties of the human visual system such as spatial frequency sensitivity and the visual masking effect. We show that the visual tool sets in JPEG-2000 are much richer than what was available in JPEG, where only locally invariant frequency weighting can be exploited

1. INTRODUCTION

It is obvious that the human visual system (HVS) plays a key role in the final perceived quality of the compressed images. It is therefore desirable to allow system designers and users to take advantage of the current knowledge of visual perception and models. In the Sydney JPEG meeting (where initial JPEG 2000 proposals were made), the contribution from Sharp Labs of America [1] demonstrated the impressive visual gain that frequency weighting can offer, particularly at display resolutions greater than 200 dpi (127 $\mu\text{m}/\text{pixel}$). Since then, the JPEG committee working on JPEG-2000 has been aggressively pursuing the goal of removing perceptual irrelevancy, in addition to statistical redundancy, of the image data.

The most common method of visually-optimizing compression algorithms is to transform the amplitudes of the image to a perceptually uniform domain. Since the visual system's gray scale behavior is approximately characterized by a cube-root amplitude nonlinearity, the theory is to convert the image to an inverse domain, such as cubic, and then quantize. This technique forms part of nearly all video [2], with the exception that the power function of 3 is replaced by values around 2.2 to 2.4; this domain is generally referred to as gamma-corrected. Most compression algorithms do this by default, as a consequence of compressing images represented in the format of video standards. The advantage of this approach is so substantial that it is essentially de facto in any compression algorithm. The key remaining dimensions that can be visually optimized are along spatial frequencies and the visual masking by the image content.

JPEG-2000 [3] is a wavelet-based bit-plane coder where coefficients in each wavelet sub-band are divided

into blocks of same size (called code-block) and each code-block is embedded coded independently. It introduces the concept of abstract quality layers that allows a 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. As a by-product, this structure enables the incorporation of some visual optimization tools into the system. In this paper, we review the tools in JPEG-2000 that allow the users to take advantages of the various properties of the HVS such as spatial frequency sensitivity and the visual masking effect. We will show that the visual tool sets in JPEG-2000 are much richer than what was available in JPEG, where only locally invariant frequency weighting can be exploited.

2. VISUAL FREQUENCY WEIGHTING

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 frequencies [4][5], as shown in Fig. 1. In general, human eyes are less sensitive to high frequency errors than to low frequency errors. The CSF can be used to determine the relative accuracies needed across differing spatial frequencies, where the term *weight* is used to describe the desired proportional accuracy. In using the CSF, which is described in visual frequencies of cycles/degree (cpd), it must be mapped to the compression domain of digital frequencies such as cycle/pixel. The design of the CSF weights is an encoder issue and depends on the specific viewing condition under which the decoded image is to be viewed [4]. Recent studies [9] suggest that it may also depend on the distortion/bit-rate of the compressed image.

2.1 Fixed frequency weighting

In general, the CSF curve is a continuous function of the spatial frequency. However, for a discrete wavelet transform, it is common that only one CSF weight is chosen for each subband to facilitate the implementation. For example, based on the specific viewing condition, the weight corresponding to the sensitivity of the mid-

frequency of a sub-band could be chosen for that particular subband [4]. This way of applying visual frequency weighting is referred to as fixed *frequency weighting*. The set of CSF weights can be incorporated in one of two ways in JPEG-2000, as described in the following. In both cases, the CSF weights do not need to be explicitly transmitted to the decoder.

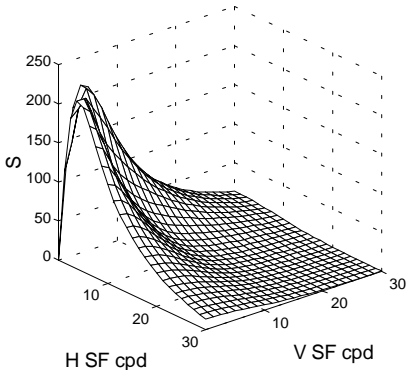


Figure 1. A general un-sampled 2D CSF

Modify the quantization step size

At the encoder, the quantization step size q_i of the transform coefficients of subband i is adjusted to be inversely proportional to the CSF weight w_i . The CSF-normalized quantization indices are then treated uniformly in the R-D optimization process. The CSF weighting information is reflected in the quantization step sizes that are explicitly transmitted for each subband. This approach needs to explicitly specify the quantizer so it may not be suitable for embedded coding from lossy all the way to lossless. This implementation can be invoked in the VM software [6] by supplying the same file of visual weights to both “-Fsteps” and “-Fweights” arguments.

Modify the embedded coding order

In this implementation, the quantization step sizes are not modified, but the distortion weights fed into the R-D optimization are altered instead, based on the CSF weight for each sub-band. This effectively controls the relative significance of including different numbers of bit-planes from the embedded bitstream of each code-block. This implementation can be invoked through the “-Fweights” option in the VM software, and is recommended since it produces similar results as the first implementation and is compatible with lossless compression. This strategy is also used for the visual progressive weighting to be described next.

It is also possible to do cell-adaptive CSF weighting [7], which allows a better adaptation of the CSF weight to the signal spectrum in a sub-region of a subband. It has

been shown [7] that the advantage of this strategy over the above mentioned fixed frequency weighting is small for the compression of natural images, but it might be of bigger impact for images of non-natural scenery.

2.2 Visual progressive weighting

JPEG-2000 allows the implementation of *visual progressive weighting*, where different sets of CSF weights can be applied at different stages of the embedding [8]. In particular, to implement the visual progressive weighting, the JPEG-2000 VM (using the “-Cvpw” argument) changes, on the fly, the order in which code-block sub-bitplanes should appear in the overall embedded bitstream based on several sets of frequency weights targeted for different bit rate ranges.

The initial motivation for visual progressive weighting is that “as the embedded bitstream may be truncated later, the viewing conditions for different stages of embedding may be very different” [8]. Visual progressive weighting thus allows the use of different sets of CSF weights that correspond to different viewing distances at different stages of the embedding. However, it remains unclear what viewing distance should be considered for a specific bit rate range, or if that is entirely application dependent.

Recent studies [9] have shown that even with a fixed viewing distance, a more aggressive weighting usually results in a better visual quality than the “matched” weighting at lower bit rates. A distortion-adaptive visual weighting strategy, based on a visual signal estimation approach (in addition to the traditional visual signal detection approach), has been proposed [9] to address visual weighting at low bit rates for both fixed frequency weighting and visual progressive weighting.

3. VISUAL MASKING

Frequency weighting is usually very effective for applications with a high-resolution display or large viewing distance. In both cases, the viewing distance expressed in units of pixels will be greater than around 1500. 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. In this case, visual masking provides more leverage for improving the visual quality.

Visual masking is a perceptual phenomenon where artifacts are locally masked (i.e., hidden) by the explicit image. The image acts as a background signal that reduces the visibility of the false signals generated by the distortion. The visual masking approaches in JPEG-2000 allow bitstream scalability, as opposed to many previous works.

3.1 Self-contrast masking

It is understood nowadays that the masking property of human vision primarily occurs locally *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 [10], as opposed to overtly adaptive techniques such as [11][12][13]. 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 [14]. One way [14] to exploit this masking effect (which will be referred to 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 (see Fig. 2). The exploitation of self-contrast masking could be invoked in the VM software using the “-Xmask” option with the parameter β set to 0.

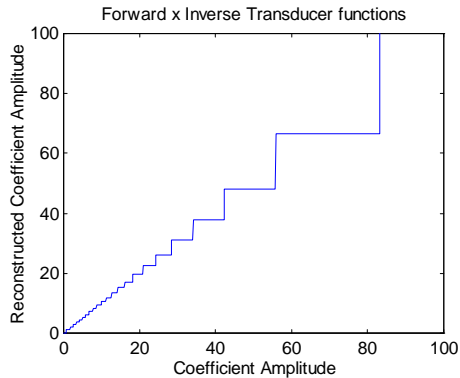


Figure 2. Nonuniform quantization for self-contrast masking.

3.2 Neighborhood masking

Another way of exploiting visual masking is through control of individual code-block contribution [15]. In this approach, the embedded coding of each code-block is performed without considering visual masking effect. However, in the post-compression rate-distortion (R-D) optimization process, 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. We will refer to this type of masking as block-based neighborhood masking. This approach adjusts only the distortion metric at the encoder, which is an advantage from an implementation point of

view. The masking effect exploited can also be spatially extensive which is not exploited in the above self-masking approach. Its weakness is that it can only adjust the truncation points of each code-block, which is a spatially coarser adjustment than the sample-by-sample compensation offered by the self-masking approach [14]. This neighborhood masking is accessible via the “-Cvis” option in the JPEG-2000 VM software.

3.3 Point-wise extended masking

A more comprehensive visual masking approach has been developed [16] that extends the point-wise “non-linearity” of self-masking [14] to an “extended non-linearity”. This also takes care of the masking effect and spatial summation contributed from spatially neighboring coefficients. This is to overcome the over-masking problem of the self-masking approach [14] 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, such as texture. This feature will assure the good visual quality of simple edges in a smooth background, which is often critical to the overall perceived quality.

This point-wise extended masking approach treats visual masking as a combination of two separate processes. The first step is to apply a point-wise power function to the original coefficient x_i , i.e., $x_i \rightarrow y_i = \text{sign}(x_i) |x_i|^\alpha$. This is basically to account for the self-masking effect. In a real image, 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 [14]. To further exploit this neighborhood masking effect, the second step normalizes y_i by a neighborhood masking factor that is a function of the amplitudes of the neighboring signals. A good model that has been adopted by the JPEG-2000 standard is to use the non-linear transform

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

where $|\phi_i|$ denotes the size of a causal neighborhood, a is a normalization factor, \hat{x}_k denotes the quantized neighboring coefficients (that only retain the first few most significant bits of the quantization index to allow for embedded coding). The parameters β and $|\phi_i|$ are used to control the degree of neighborhood masking. The z values are then subject to uniform quantization. The inverse is performed at the decoder. This is essentially a coefficient-wise adaptive quantization without any overhead.

4. DISCUSSION AND CONCLUSION

The various visual optimization tools in JPEG-2000 have their own merit and weakness. The visual frequency weighting is usually very effective for large viewing distances or high-resolution displays, but it is tied to a specific viewing condition. Under different viewing conditions, the perceived quality can vary a lot. In other words, the weights used at the encoder have to match the viewing condition under which the image is to be viewed. When using a viewing distance for an application or image study, it is important to use a frequency weighting set for the closest distance expected. Three sets of CSF weights have been recommended in JPEG-2000 for some common viewing/printing scenarios. These are csf1000, csf2000, and csf4000, where 1000, 2000 and 4000 refer to the viewing distance in pixels. Unlike the JPEG default, these are based solely on the CSF and hence, do not include any display MTF effects, such as the CRT MTF implicitly occurring in the JPEG default tables. We decided to omit the display MTF in the default, since with today's technology it is equally as likely that the display will be an LCD, FED, DLP projector, or hardcopy as it will be a CRT.

The masking approaches usually are less sensitive to the viewing condition. The self-masking approach usually protects the fine texture well, which is especially suitable for high quality photographic images that contain human faces. It, however, may have some problems with sharp edges, especially at low bit rates. The block-based neighborhood masking approach usually tends to smooth out the fine texture, but protects high contrast edges well. It also has some limitations for 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 self-masking and neighborhood masking, thus resulting in mutual synergism.

The various visual optimization tools can in fact be combined together to maximize the visual performance. It has been observed that, for some complex images with diverse content, the visual improvement can be equivalent to a saving of up to 50% in bit-rate.

5. ACKNOWLEDGEMENTS

We would like to thank Jin Li of Microsoft Research China, David Taubman of University of New South Wales, Marcus Nadenau and Julien Reichel of EPFL, Troy Chinen of FUJIFILM Software, Tom Flohr of SAIC, Alan Chien of Eastman Kodak, Margaret Lepley of MITRE, and many others in the JPEG committee for their contributions and support to the visual optimization work in JPEG-2000.

6. REFERENCES

- [1] J. Li, W. Zeng and S. Lei, "Sharp rate-distortion optimized embedded wavelet coding—an algorithm proposal for JPEG 2000" ISO/IEC JTC1/SC29/WG1 N621, Sydney, Australia, Oct. 1997.
- [2] C. Poynton, "A Technical Introduction to Digital Video". Pg. 92, John Wiley and Sons, NY, 1996
- [3] "Information Technology – JPEG 2000 Image Coding System," ISO/IEC FCD15444-1: 2000 (V1.0, Mar. 2000).
- [4] P. Jones, S. Daly, R. Gaborski and M. Rabbani, "Comparative study of wavelet and DCT decompositions with equivalent quantization and encoding strategies for medical images," *SPIE Proceedings of Conference on Medical Imaging*, vol. 2431, pp. 571-582, 1995.
- [5] Watson, Yang, Solomon and Vilasenor, "Visibility of wavelet quantization noise," *IEEE Tran. Image Proc.*, vol. 6, No.8, pp. 1164-1175, 1997.
- [6] "JPEG 2000 Verification Model 7.0 Software", ISO/IEC JTC1/SC29/WG1 N1685, April 2000.
- [7] M. Nadenau and J. Reichel, "Report on CE V2 (Performance analysis of csf-filtering compared to fixed weighting using recommended parameter set)," ISO/IEC JTC1/SC29/WG 1 N1470, Maui, Hawaii, Dec. 1999.
- [8] J. Li, "Visual progressive coding," in *Proc. IS&T/SPIE Conf. Visual Communications and Image Processing*, vol. 3653, Jan. 1999.
- [9] W. Zeng and S. Lei, "Report on CE V1 (CSF weighting strategy for visual progressive coding)," ISO/IEC JTC1/SC29/WG1 N1584, Tokyo, March, 2000.
- [10] A. B. Watson, "Efficiency of an image code based on human vision", *JOSA A V. 4*, pp. 2401-2417, 1987.
- [11] M. R. Civanlar, S. A. Rajala, and W. M. Lee "Second generation hybrid image coding techniques", *SPIE V. 707 VCIP*, pp. 132- 137. 1986
- [12] 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.
- [13] T. N. Pappas, T. A. Michel, and R. O. Hinds, "Suprathreshold perceptual image coding" *IEEE, ICIP*, pp. 237-240, 1996.
- [14] S. Daly, W. Zeng, J. Li, and S. Lei, "Visual masking in wavelet compression for JPEG2000," in *Proc. IS&T/SPIE Conf. Image and Video Communications and Processing*, vol. 3974, Jan. 2000.
- [15] David Taubman, "High performance scalable image compression with EBCOT", to appear in *IEEE Transactions on Image Processing*, 2000.
- [16] W. Zeng, S. Daly and S. Lei, "Point-wise extended visual masking for JPEG2000 image compression," in *Proc. IEEE Inter. Conf. Image Proc.*, Sept. 2000, Vancouver, Canada.