ISO/IEC JTC1/SC29/WG1 N1716
22 June 2000

ISO/IEC JTC 1/SC 29/WG 1
(ITU-T SG8)

Coding of Still Pictures

JBIG	**JPEG**
Joint Bi-level Image Experts Group	Joint Photographic Experts Group

TITLE: Report on CE V1 (Evaluation of the Distortion-adaptive Progressive CSF Weighting Technique)

SOURCE: Wenjun Zeng (zengw@sharplabs.com)
Sharp Labs. of America, Inc., 5750 NW Pacific Rim Blvd, Camas, WA 98607
(Cross-verified by Troy Chinen, FUJIFILM Software)

PROJECT: JPEG 2000

STATUS: Information

REQUESTED ACTION: For information only

DISTRIBUTION: WG 1 Mailing List

Contact:
ISO/IEC JTC 1/SC 29/WG 1 Convener - Dr. Daniel T. Lee
Hewlett-Packard Company, 11000 Wolfe Road, MS42U0, Cupertino, California 95014, USA
Tel: +1 408 447 4160, Fax: +1 408 447 2842, E-mail: Daniel_Lee@hp.com

Core Experiment Description/Results Summary on VM 7.0

Experiment Name: Evaluation of the distortion-adaptive CSF weighting technique

Sub-Group: ____Visual__________ Number: ___V1_____

Description:

Core experiment **partners**	Sharp Labs of America, FUJIFILM Software
Core experiment **objective**	To investigate the performance of the distortion-adaptive CSF weighting technique proposed in the Tokyo meeting
JPEG 2000 **Requirement** Focus	Visual quality
What will **change** fro[m] Verification Model 7.[0]	Modify the distortion metric to incorporate different CSF weighting sets at different visually normalized distortions.
Key **Benefit** of chang[e]	A finer and better adjustment of CSF weights to improve the visual quality at different distortion (or bit rates), especially in a visual progressive weighting environment
Related Experiments	None
Expected Memory Decrease/increase	Negligible
Expected Complexity Decrease/increase	Negligible
Other expected results	N/A

Core experiment detai Description	In the current VM, visual progressive coding is realized by specifying different CSF weighting tables for different bit rate ranges. The set of appropriate weighting tables for different bit rate ranges have not been carefully studied. A distortion-adaptive CSF weighting technique was proposed by Sharp Labs of America in the Tokyo meeting. Preliminary study using the existing visual progressive weighting mechanism in the VM to approximate the proposed scheme showed that this technique has the potential of achieving good visual quality at different distortion/bit rates. This core experiment will further investigate the performance of this technique by exactly deriving the effective weights based on the instant distortion.

Results:
Web Document Number: ____N1716_____________

VM Mode Used in Experiment	"-Fweights", "-Cdavpw", "-Clayers" options

What has **changed** from Verification Model 7.x (provide level of integration)	The visual progressive weights for each quality layer have been set based on the distortion of the previous layer and the traditional weighting table passed through the "-Fweights" argument.

Was this experiment **performed** on the VM or in a testbed	VM7.0

Key **Benefit** of change	Improved visual quality at lower bit rates
Key **Cost** of change	Slight increase of computation

Other performance **results**	Better visual quality at different bit rates of an embedded/non-embedded bit-stream

1. Background

JPEG-2000 allows the implementation of *visual progressive weighting,* where different sets of CSF weights can be applied at different stages of the embedding [1]. In particular, to implement the visual progressive weighting, the JPEG-2000 VM [3] (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 visual frequency weights targeted for different bit rate ranges. However, it is not clear what weighting table should be applied for a specific bit rate range.

In traditional psychophysical experiments, the amplitude of each frequency basis function is increased until it reaches a just noticeable frequency threshold (JND) where people can detect the existence of the signal under a specific viewing condition [4][5]. These frequency JNDs are then used to generate the CSF curve to represent the relative visual significance of each frequency component. Typically, $w_i=k/T_i$, where w_i and T_i, respectively, are the CSF weight and the frequency detection threshold for the ith frequency basis function, and k is a constant normalization factor. Previous works on perceptual coding usually implicitly assume that the relative weights will remain unchanged for different distortions/bit-rates.

Experiments have shown [2] that the traditional CSF weights do not seem to work well in low bit rate scenarios. This is not necessarily surprising because the traditional CSF curve is derived based on detection threshold under visually near-lossless condition. At lower bit rates, the distortion is very large and the visual effect has not been studied sufficiently in the literature. In our previous core experiment [2], we proposed a distortion-adaptive CSF weighting strategy to address the visual frequency sensitivity under the condition of large distortion.

We argued that [2], for low bit rates, the effect of visual distortion is an estimation problem rather than just a detection problem. In other words, it becomes important to estimate the amount of visual distortion of each frequency component perceived by the human eyes when measuring the frequency sensitivity. For wavelet based systems, different basis functions usually have different spatial supports, and different non-flat envelops. In general, low frequency basis functions have a larger spatial support than high frequency basis functions. In visually near lossless scenario, the side lobes of the basis function remain largely undetected. So the spatial support of the basis function is not of significant impact on the perception. However, at low bit rates, the distortion signal strength is increased and the side lobes of the basis function become visible. The spatial support of the basis function starts affecting the perception of the distortion.

We have proposed [2] to use the following measure to compensate for the "side lobe effect". Let $f_i(x)$ denote the basis function with unit peak-to-mean amplitude for the ith subband. Assume the distortion to each basis function is $d\,f_i(x)$ where d is the normalized peak-to-mean amplitude (in the unit of T_i). The normalization is with respect to the frequency detection threshold T_i of each basis function. It accounts for the visual sensitivity to spatial frequency. We define the "effective" basis distortion function $g_i(x; d)$ as

$$g_i(x; d) = \begin{cases} |d\,f_i(x)|, & \text{if } |d\,f_i(x)|>1 \\ 0, & \text{otherwise} \end{cases} \qquad (1)$$

The compensation factor λ_i that account for the "side lobe effect" can be defined as

$$\lambda_i(d) = (\int_{-\infty}^{+\infty} | g_i(x;d) |^p \, dx)^{1/p}, \quad \text{if } d > 1 \quad (2)$$

where $0 \le p < \infty$. If $d \le 1$, $\lambda_i(d)$ will all be set to 1. Therefore, if d is less than 1 (or equivalently, the peak-to-mean amplitude of the distortion to each basis function is less than the frequency detection threshold T_i), there is no compensation for the "side lobe effect". If the actual peak-to-mean amplitude of the basis distortion function is greater than T_i, then the portion of the basis function whose amplitude exceeds the thredshold T_i will contribute to the visual distortion. A special case is that p=2, and $d \to \infty$. In this case, the compensation factor λ_i is in fact the square root of the energy of the basis function with unit peak-to-mean amplitude, subject to a constant factor that is common to all basis functions. In general, low frequency basis function with unit peak-to-mean amplitude will have larger energy than high frequency basis funtion with unit peak-to-mean amplitude. This suggests that low frequency basis function is more sensitive to distortion than high frequency basis function, thus demanding more protection than what the traditional CSF curve suggests that only accounts for frequency sensitivity, but not the "side-lobe effect".

The final effective CSF weight for a distortion with normalized peak-to-mean amplitude of d should be

$$w_i' = w_i \lambda_i, \quad (3)$$

subject to a constant normalization factor.

The model described above tries to characterize the different amounts of distortion perceived by the human eyes when the distortion signal for each frequency has an amplitude that is d times of its frequency detection threshold T_i. Note that previous works on perceptual coding usually assume that these visual distortions are the same. The proposed model therefore provides a fine adjustment of the frequency weights based on the instant normalized peak-to-mean amplitude of the distortion signal during the embedded coding process.

2. Implementation and performance evaluation

The traditional CSF curve usually has a dip at very low frequencies and reaches the peak value at some mid-frequency f_{peak}. In practice, the weights are usually set to 1 for all frequencies no larger than f_{peak}. For a reasonable compenstation of the side-lobe effect, we assume that the peak will be assumed at the next lower frequency level. For example, in Table 1, the original CSF weights have a peak at level 3. The effective weights will then have a peak at level 2.

The distortion-adaptive progressive CSF weighting strategy was implemented based on VM7.0 [3]. Basically, for each quality layer, the instant T_i-normalized average distortion of the whole image after encoding the last quality layer will be used to calculate the compensation factors and update the effective weights. In fact, we also tested using the distortion corresponding to the current encoded bit planes to update the effective weights. Similar performance was found.

Fig. 1 illustrates how λ_i changes as a function of the subband index for different normalized peak-to-mean distortion amplitude *d* for the Daubechies 9/7 filter. The lowest LL subband is indexed as 0, the lowest HL, LH and HH bands are indexed as 1, 2, and 3, and so on. When *d* is smaller than 1, λ_i is 1 for all subbands. When *d* increases, λ_i tends to become smaller for the higher frequency bands.

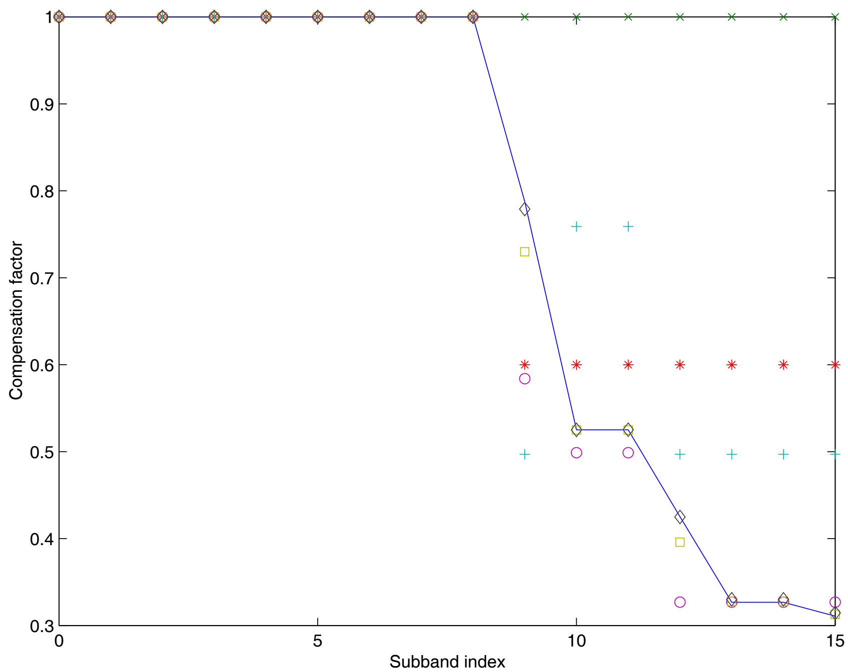


Fig. 1: Compensation factors as a function of the subband index for different distortions. "x", "*","+", "o", "□", "◊", and "solid-line"represent the factors for the normalized peak-to-mean distortion amplitude *d* of 1, 1.18, 1.3, 1.82, 4.05, 8.85, and ∞ , respectively.

1000.tbl				**1000_∞.tbl**			
(1 1)				(1 1)			
1.0000	0.5608	0.5608	0.2841	1.0000	0.1833	0.1833	0.0884
1.0000	1.0000	1.0000	0.7271	1.0000	0.5251	0.5251	0.3092
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7876
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 1: two sets of weights for visual progressive weighting under 1000-pixel viewing distance condition. Left: basic weights; Right: effective weights with *p*=2 and *d*=∞.

We tested the performance of the distortion-adaptive visual progressive weighting (DAVPW) strategy on "woman", "bike", "lenna", and "face_guy_down2" images for the on-screen 1000-

pixel viewing distance case. The command lines used are

vm7_compress_32 -i ${file}.pgm –o ${file}_vpw1000.cd2 –rate 2 –Clayers –Fweights 1000.tbl –Cdavpw

vm7_expand_32 -i ${file}_vpw1000.cd2 –o ${file}_vpw1000_{rate}.pgm –trunc ${rate}

We compared the performance of distortion-adaptive visual progressive weighting to that of fixed weighting using the 1000.tbl table. It was observed that the distortion-adaptive visual progressive weighting provides noticeable visual gains over fixed weighting at 0.125 bpp, 0.25 bpp, and 0.5 bpp for the "bike", "woman" and "face_guy_down2" images (see, e.g., Fig. 2). In general, the lower the bit-rate, the larger the visual gain. For the "lenna" image, the visual gain is not quite noticeable. At 0.5 bpp, there is some visual gain around the hat area. At lower bit rates, it is difficult to judge. We have shown in [2] that using the aggressive 1000_∞.tbl table for fixed weighting results in some high frequency artifacts for the "bike" image (e.g., areas around the bottle and the frequency charts) at 0.5 bpp and 0.75 bpp. We also observed that 1000_∞.tbl results in similar high frequency artifacts for the "face_guy_down2" (e.g., face, hair, collar etc.) and "lenna" (e.g., hat area) images at 0.5 and 0.75 bpp, and for the "woman" image at 0.75 bpp (e.g., the fingers, see Fig. 3).

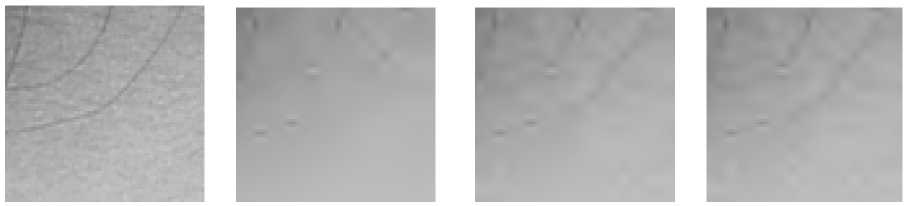
3. Conclusion

We believe that for large signal distortion, the "effective" CSF weighting should take into account the amount of visual distortion each basis function with unit peak-to-mean amplitude contributes, in addition to the visual sensitivity due to signal frequency. Experimental results suggest that our proposed model characterize the psychophysical experience reasonably well, thus providing a valuable guidance as to how to adjust the CSF weights for visual progressive weighting. Other factors may also contribute to the visual gain. For example, since high frequency basis function usually has a smaller support, it can be masked when it is close to a sharp edge, while the side-lobe of a lower frequency basis function can not be masked easily. Further investigation is warranted.

References

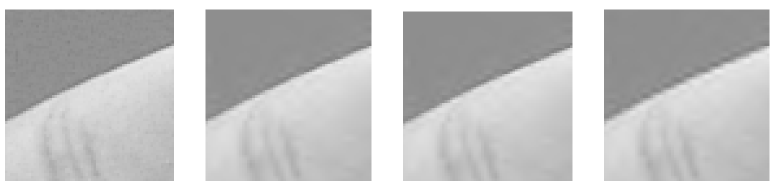
1. J. Li, "Visual progressive coding," in *Proc. IS&T/SPIE Conf. Visual Communications and Image Processing*, vol. 3653, Jan. 1999.
2. W. Zeng and S. Lei, "Report on CE V1 (CSF Weighting Strategy for Visual Progressive Coding)," ISO/IEC JTC1/SC29/WG1 N1584, Tokyo, March 2000.
3. "JPEG 2000 Verification Model 7.0, ISO/IEC JTC1/SC29/WG1 N1684, Feb. 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.



Original | 1000.tbl | DAVPW | 1000_∞.tbl

Fig. 2: foreheads of the “woman” image coded at 0.25 bpp using different weighting strategies.



Original | 1000.tbl | DAVPW | 1000_∞.tbl

Fig. 3: fingers of the “woman” image coded at 0.75 bpp using different weighting strategies.