Improving JPEG2000's Perceptual Performance with Weights Based on Both Contrast Sensitivity and Standard Deviation

K. Varma
Virginia Tech
Email: kvarma@vt.edu

A. E. Bell
Virginia Tech
Email: abell@vt.edu

***Abstract*— The JPEG2000 standard includes a perceptual weighting method based on the contrast sensitivity function (CSF). Recent literature shows that perceptual methods based on subband standard deviation are also effective in image compression. We present two new perceptual weighting methods that combine information from both the human contrast sensitivity function as well as the standard deviation within a subband or code-block. We compare these two new sets of perceptual weights to the JPEG2000 CSF weights. The results indicate that our new weights performed better than the JPEG2000 CSF weights for high frequency images. Weights based solely on subband standard deviation are shown to perform worse than JPEG2000 CSF weights for all images at all compression ratios.**

I. Introduction and Background

Improving the perceptual quality of compressed images has received considerable attention in the literature. Perceptual methods that exploit the varying sensitivity of the human visual system to different spatial frequencies are among the most effective techniques. Recent research indicates that there have been three primary approaches to the effective design of frequency-based perceptual methods. The first method uses a contrast sensitivity function (CSF) to design perceptual weights as a function of spatial frequency [1], [2], [3]. Each weight is multiplied with all of the discrete wavelet transform (DWT) coefficients in the subband with the corresponding spatial frequency: the DWT coefficients are increased at the more sensitive frequencies and are decreased at the less sensitive frequencies. The JPEG2000 standard [4] includes perceptual weights derived from a CSF [3].

The second method is based on the maximum amount of quantization that can be present in a subband without being detected by humans [5]. This dictates the maximum quantization step size for each subband that will result in visually lossless compression. However, this method does not allow direct control of the bit rate. The method can be adapted to include multiplication factors on the quantization step sizes in order to achieve a prescribed bit rate; unfortunately, iteration is required to find the multiplication factors that give the desired bit rate for a particular image.

The third approach showed that the multiplication factors in the second method were sub-optimal when quantization artifacts were visible at very low bit rates [6]. This third method uses subband standard deviation (SSD) to derive a weight for each subband. Here the weights are not only frequency dependent, but also orientation dependent (distinguishing it from the previous two methods). This approach has not been compared to weights derived from a CSF.

This material is based upon work supported by the National Science Foundation under Grant 0218672.

II. New Perceptual Methods and Comparisons

In this paper we develop two new sets of perceptual weights: C-SSD, which combines the weights from CSF and subband standard deviation; and, C-CSD which combines the weights from CSF and code-block standard deviation. We compare the performance of our new weights to the JPEG2000 CSF weights. We also compare the SSD weights [6] to JPEG2000's CSF weights. We expect that one of our new methods that combines the CSF and standard deviation weights will improve the perceptual quality of the compressed images over JPEG2000's CSF weights.

JPEG2000 allows the perceptual weights to be directly applied to the mean-squared-error (MSE) obtained for each MQ-encoded code-block. These weighted MSEs, defined as:

$$WMSE(s,c) = MSE(s,c)W_i^2(s,c) \tag{1}$$

alter JPEG2000's subsequent optimal bit allocation. (s,c) refers to functions of subband s and/or code-block c. $WMSE(s,c)$ is the weighted MSE and $MSE(s,c)$ is the unweighted MSE for code-block c belonging to subband s. $W_i(s,c)$ refers to the subband or code-block weights of method i; the four methods are C for CSF, SSD for subband standard deviation, C-SSD for CSF and subband standard deviation, and C-CSD for CSF and code-block standard deviation.

The first set of weights is obtained from the standard deviations of the subbands [7]:

$$W_{SSD}(s) = l\sigma_s^{(1-b)} \tag{2}$$

where $b = 0.3432$, σ_s is the standard deviation of subband s, and l is a normalization factor that makes the largest weight unity. These weights have been demonstrated to improve subjective performance over no weights using SPIHT for quantization. However, $W_{SSD}(s)$ has not been compared with weights derived from a CSF. In this paper, we compare the performance of $W_{SSD}(s)$ with JPEG2000's CSF weights, $W_C(s)$, using the JPEG2000 EBCOT quantization algorithm. In our evaluations $W_C(s)$ represents the JPEG2000 CSF perceptual weights for a viewing distance of 1700 pixels [8].

The second set of weights is based on our new method that combines the CSF weights with subband standard deviations:

$$W_{C-SSD}(s) = l\sigma_s^{(1-b)} W_C(s). \quad (3)$$

We compare the performance of $W_{C-SSD}(s)$ with $W_C(s)$.

The third set of weights is based on our new method that combines the CSF weights at the subband level with the standard deviations at the code-block level:

$$W_{C-CSD}(s,c) = l\sigma_c^{(1-b)} W_C(s) \quad (4)$$

σ_c is the standard deviation of code-block c in subband s.

III. Testing Methodology

Eight images from the standard image set were used in our subjective tests [9]. 15 512×512 sub-images were chosen from these 8 images in order to represent a wide range of characteristics: frequency (smooth areas–highly detailed/textured areas), color (many colors–grayscale), and content (aerial photos of buildings–people). The sub-images were classified as high, medium or low frequency by thresholding the average number of edges detected by a Sobel operator.

The 15 images were compressed with JPEG2000's *Kakadu* coder Version 2.2 [8] using one of the four sets of perceptual weights: $W_C(s)$, $W_{SSD}(s)$, $W_{C-SSD}(s)$ and $W_{C-CSD}(s,c)$. The images were compressed using 5 levels in the JPEG2000 lossy coder (biorthogonal-9/7 wavelet); 5 compression ratios were examined: 32:1, 50:1, 70:1, 90:1, and 120:1. This resulted in 75 (15 images $\times$ 5 compression ratios) test images for each perceptual weighting method. Again, the 3 comparisons performed in our subjective tests were: $W_C(s)$ vs. $W_{SSD}(s)$, $W_C(s)$ vs. $W_{C-SSD}(s)$, and $W_C(s)$ vs. $W_{C-CSD}(s,c)$.

We derived our testing methodology from the stimulus comparison testing procedure outlined in the standard for subjective assessment of the quality of television pictures [10]. During a subjective test session, the participant was asked to compare the perceptual quality of two candidate images. Both candidates were the same image compressed at the same compression ratio; however, each candidate was compressed using one of the two perceptual weighting methods under comparison: $W_i(s)$ or $W_C(s)$. Throughout each test session the image/compression ratio combination was randomly chosen from the set of 75 candidate images.

The two images were shown one at a time; the participant could switch between them as often as desired, but a gray, blank screen was displayed for 2 seconds during the switch. The participant was asked to fill the blank: "Image 1 is ____ compared to Image 2". The choices were (1) worse, (2) slightly worse, (3) same, (4) slightly better, or (5) better. (Note: the assignment of $W_i(s)$ and $W_C(s)$ to Image 1 and Image 2 was random throughout each test session).

When $W_i(s)$ was used in Image 1, choices (1)-(5) translated to scores of $-2, -1, 0, +1, +2$. However, when $W_C(s)$ was used in Image 1, choices (1)-(5) translated to scores of $+2, +1, 0, -1, -2$. This scoring method generates positive values if the participant decided that $W_i(s)$ performed better than $W_C(s)$; conversely, negative values indicate that the participant decided that $W_C(s)$ performed better than $W_i(s)$.

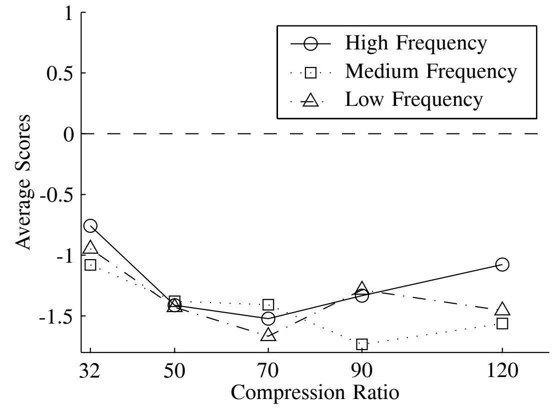

Fig. 1. Average test scores for the $W_C(s)$ vs. $W_{SSD}(s)$ comparison. $W_C(s)$ outperforms $W_{SSD}(s)$ for all image types at all compression ratios.

Each participant viewed 25 comparisons; the first 5 decisions were thrown away to account for the learning process. 45 participants took our test—15 participants for each perceptual weight comparison. Thus, each of the 3 perceptual weighting comparisons was evaluated using 300 decisions spread randomly across 75 image/compression ratio combinations.

IV. Results

A. Comparison 1: $W_C(s)$ vs. $W_{SSD}(s)$

Figure 1 shows that the subjective quality with the $W_{SSD}(s)$ perceptual weights was lower than with the $W_C(s)$ perceptual weights. For all image types (high, medium and low frequency) at all compression ratios, the average test scores are always below zero—indicating that compression using $W_{SSD}(s)$ was consistently perceived as worse than compression using $W_C(s)$.

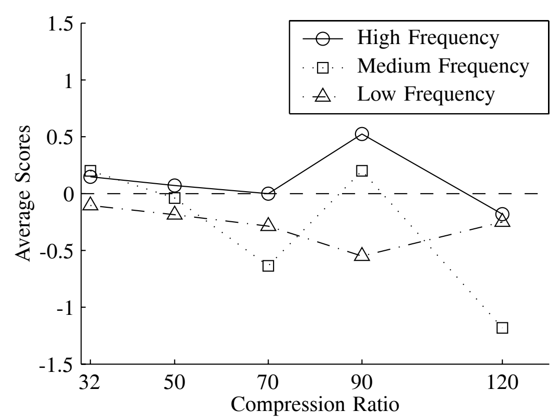

Fig. 2. Average test scores for the $W_C(s)$ vs. $W_{C-SSD}(s)$ comparison. $W_{C-SSD}(s)$ improves subjective quality for the high frequency images, but $W_C(s)$ provides better quality for low frequency images.

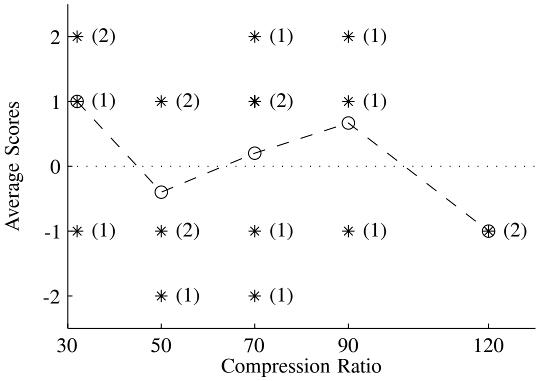

Fig. 3. Individual (*) and average test scores (dashed line) for the $W_C(s)$ vs. $W_{C-SSD}(s)$ comparison using the newspaper/silverware sub-image from Bike. There was a lack of consensus on perceived quality.

B. Comparison 2: $W_C(s)$ vs. $W_{C-SSD}(s)$

Figure 2 depicts varying subjective quality results for the different image types when the $W_{C-SSD}(s)$ perceptual weights are compared with the $W_C(s)$ perceptual weights. For high frequency images, the average test score is positive (except at 120:1 where images compressed with any method are significantly degraded); this indicates that compression using $W_{C-SSD}(s)$ was consistently perceived as better than compression using $W_C(s)$. For low frequency images, the average test score is negative which means that compression using $W_{C-SSD}(s)$ was consistently perceived as worse than compression using $W_C(s)$.

Figure 2 exhibits mixed results for medium frequency images. One test sub-image was taken from the newspaper/silverware section of Bike. The subjective test scores for each participant are depicted in Figure 3. The participants' evaluations are positive and negative for this image; thus, there was a lack of consensus on perceived quality.

In general, for medium frequency images at low to moderate compression ratios (32:1 and 50:1), $W_{C-SSD}(s)$ improved the subjective quality over $W_C(s)$. $W_{C-SSD}(s)$ is better able to preserve the high frequencies (edges, texture) due to the SSD part of the weight that gives more emphasis to the higher frequencies than the CSF part of the weight.

At higher compression ratios ($\geq$ 70:1), $W_{C-SSD}(s)$ continues to preserves edges better than $W_C(s)$. Figure 4 illustrates how $W_C(s)$ does not reconstruct the edges of the silverware as sharply as $W_{C-SSD}(s)$ in Figure 5. However, $W_{C-SSD}(s)$ tends to introduce shadow artifacts at the edges of smooth areas juxtaposed against detailed areas. For example, compare the smooth top of the newspaper in Figure 6 compressed using $W_C(s)$ with the same area in Figure 7 compressed using $W_{C-SSD}(s)$. Figure 6 is perceptually better.

A participant who focused more on the silverware than the newspaper would have contributed to the positive scores at 70:1 in Figure 3. On the other hand, a participant who focused more on the newspaper than the silverware would have contributed to the negative scores at 70:1 in Figure 3.

Fig. 4. The silverware edges compressed at 70:1 using $W_C(s)$. The edges are not as sharp as with $W_{C-SSD}(s)$ in Figure 5.

Fig. 5. The silverware edges compressed at 70:1 using $W_{C-SSD}(s)$.

Fig. 6. The newspaper compressed at 70:1 using $W_C(s)$.

C. Comparison 3: $W_C(s)$ vs. $W_{C-CSD}(s)$

The subjective quality with the $W_{C-CSD}(s)$ perceptual weights was lower than with the $W_C(s)$ perceptual weights

Fig. 7. The newspaper compressed at 70:1 using $W_{C-SSD}(s)$. The smooth areas have a shadow artifact not present with $W_C(s)$ in Figure 6.

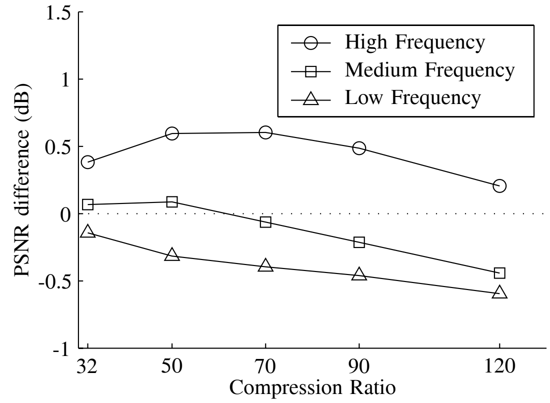

Fig. 8. Average difference in PSNR between $W_{C-SSD}(s)$ and $W_C(s)$; positive values represent a higher PSNR for $W_{C-SSD}(s)$.

for low and medium frequency images. This differs slightly from the $W_{C-SSD}(s)$ performance which showed improved perceptual quality for medium frequency images at 32:1 and 50:1. Also unlike the $W_{C-SSD}(s)$ results, $W_{C-CSD}(s)$ generated similar perceptual quality as $W_C(s)$ for high frequency images. Although the average test score for all images at all compression ratios is negative for $W_{C-CSD}(s)$—indicating that compression using $W_{C-CSD}(s)$ was consistently perceived as worse than compression using $W_C(s)$—it did not perform as poorly as $W_{SSD}(s)$.

D. Additional Metrics: PSNR and WVDP

Although the focus of our study was on subjective quality evaluated by test participants, we also examined peak signal-to-noise ratio (PSNR); the results are depicted in Figure 8. Interestingly, the PSNR values were higher with $W_{C-SSD}(s)$ than with $W_C(s)$ for the high frequency images (and lower for the low frequency images). For medium frequency images, PSNR was higher with $W_{C-SSD}(s)$ at 32:1 and 50:1, but lower for 70:1 and higher. Our perceptual evaluations were in line with PSNR—an uncommon occurrence.

We also calculated the wavelet visible difference predictor (WVDP) metric [11]. This performance measure is supposed to be able to capture subjective quality by computing a single number. Our results indicated that the WVDP for $W_{C-SSD}(s)$ was worse than the WVDP for $W_C(s)$ for all images at all compression ratios. This does not correlate with the results of our subjective tests.

V. Conclusion

We presented two new perceptual weighting methods that combine information from both the human contrast sensitivity function as well as the standard deviation within a subband or code-block. We compared these two new sets of perceptual weights to the CSF perceptual weights that are part of the JPEG2000 standard. The perceptual weights based on the CSF and subband standard deviation, $W_{C-SSD}(s)$, performed better than the JPEG2000 CSF weights, $W_C(s)$, for high frequency images (but they performed worse for low frequency images. Weights based solely on subband standard deviation, $W_{SSD}(s)$, performed worse than the JPEG2000 CSF weights for all images at all compression ratios. This suggests that the underlying human contrast sensitivity function plays a more important role in perceptual quality than the amount of activity in a subband.

JPEG2000's tiling feature can be used to divide an image into high, medium and low frequency tiles. Our results suggest that the low frequency tiles be compressed using $W_C(s)$ and the high frequency tiles be compressed using $W_{C-SSD}(s)$. For the medium frequency tiles JPEG2000's visual progressive weights feature could be used to apply $W_{C-SSD}(s)$ at low to moderate compression ratios and then switch to $W_C(s)$ at higher compression ratios.

References

[1] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Trans. Inform. Theory*, vol. 20, no. 4, pp. 525–36, July 1974.

[2] K. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *Journal of Physiology*, vol. 359, pp. 381–400, 1985.

[3] M. Nadenau and J. Reichel, "Opponent color, human vision and wavelets for image compression," in *Proc. of the Seventh Color Imaging Conference*, 1999, pp. 237–42.

[4] *ITU T.800: JPEG2000 image coding system Part 1*, ITU Std., July 2002.

[5] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Processing*, vol. 6, no. 8, pp. 1164–75, August 1997.

[6] S. S. Hemami and M. G. Ramos, "Wavelet coefficient quantization to produce equivalent visual distortions in complex stimuli," *Proc. of SPIE Human Vision and Electronic Imaging V*, vol. 3959, pp. 200–10, January 2000.

[7] ——, "Perceptual quantization for wavelet-based image coding," in *Proc. IEEE Int. Conf. Image Proc.*, vol. 1, September 2000, pp. 645–48.

[8] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Norwell, MA: Kluwer Academic Publishers, 2002.

[9] *ITU T.24: Standardized digitized image set*, ITU Std., June 1998.

[10] *ITU-R BT.500-11: Methodology for the subjective assessment of the quality of television pictures*, ITU Std., 2002.

[11] A. P. Bradley, "A wavelet visible difference predictor," *IEEE Trans. Image Processing*, vol. 8, no. 5, pp. 717–30, May 1999.