USING STRUCTURAL SIMILARITY QUALITY METRICS TO EVALUATE IMAGE COMPRESSION TECHNIQUES

*Alan C. Brooks**

Defensive Systems Division
Northrop Grumman Corporation
Rolling Meadows, IL 60008, United States

Thrasyvoulos N. Pappas

Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208, United States

ABSTRACT

Perceptual image quality metrics have explicitly accounted for perceptual characteristics of the human visual system (HVS) by modeling sensitivity to subband noise as the just-noticeable threshold of distortion. While such metrics can successfully account for contrast and luminance masking, they are quite sensitive to spatial shifts, intensity shifts, contrast changes, and scale changes. In contrast, the recently proposed Structural SIMilarity (SSIM) metrics account for perception more implicitly with the assumption that the HVS is adapted for extracting structural information (relative spatial covariance) from images. As such, they have the potential to be much more effective in quantifying suprathreshold compression artifacts than traditional perceptual metrics, as such artifacts tend to distort the structure of an image. We use a (perceptually) weighted variation of the Complex Wavelet SSIM (CWSSIM) to evaluate standard image compression techniques such as JPEG, JPEG 2000, SPIHT, and the Safranek-Johnston perceptual image coder. Our experimental results indicate that the weighted CWSSIM generally agrees with subjective evaluations.

Index Terms— perceptual quality, structural similarity, image compression, JPEG, JPEG 2000, perceptual subband image coder

1. INTRODUCTION

Perceptual image quality metrics have explicitly accounted for perceptual characteristics of the human visual system (HVS) by modeling sensitivity to subband noise as the just-noticeable threshold of distortion [1]. While these metrics were developed for near-threshold applications, their use has been extended to suprathreshold applications [2, 3]. More systematic studies of the suprathreshold case have been conducted by Hemami's group [4–7]. However, while perceptual metrics can successfully account for contrast and luminance masking, they are quite sensitive to spatial shifts, intensity shifts, contrast changes, and scale changes. Moreover, Chen *et al.* [3] found that perceptual metrics based on a given subband decomposition (DCT, wavelet, generalized quadrature-mirror filters) are inherently biased towards the coders that use the same decomposition.

Another class of quality metrics, known as Structural SIMilarity (SSIM) [8], account for perception more implicitly with the assumption that the HVS is adapted for extracting structural information (relative spatial covariance) from images. While the structural similarity metrics have been shown to have a number of desirable properties, there has been no systematic study in the context of image compression. In this paper, we use a perceptually weighted variation of the most sophisticated and effective of the structural similarity metrics, the Complex Wavelet SSIM (CWSSIM), to evaluate standard image compression techniques such as JPEG, SPIHT [13], the Safranek-Johnston perceptual image coder [12], and JPEG2000. Since the SSIM metrics are derived from assumptions about the high-level functionality of the HVS, they have the potential to be much more effective in quantifying suprathreshold compression artifacts, as such artifacts tend to distort the structure of an image. Our experimental results indicate that the (perceptually) weighted CWSSIM (WCWSSIM) generally agrees with subjective evaluations.

*The first author performed the work while at Northwestern University

2. STRUCTURAL APPROACH TO IMAGE QUALITY MEASUREMENT

2.1. SSIM review

The motivation behind the structural similarity approach of measuring image quality is the concept that the human visual system is not built for detecting absolute, exact intensities. Instead, the HVS is adapted to help us navigate the three-dimensional space we live in and, consequently, the ability to quickly perceive the connectedness or *structure* of natural images is evolutionally advantageous. Our visual system is the preprocessor for one of our greatest strengths: visual pattern recognition. Our recognition system is robust to many changes — we can accurately recognize faces from many angles, under bright or dim lighting, and with partial obscuration because our HVS is very good at extracting structure from the underlying plenoptic light data produced by the interaction of light and the objects we are observing.

The suggestion that useful image quality metrics can be created based on the idea that the HVS extracts structural information is developed and explained in [8]. It is desirable for an image quality measurement system to be able to account for a wide variety of possible image distortions in a way that agrees with the human perception. Some of the most common distortions that the HVS encounters are due to changes in lighting [9]. Our visual system adapts to a very wide range of lighting changes without any conscious intervention from the perceiver. We consider it useful that the structural similarity approach is mostly *insensitive* to the distortions that lighting changes create: changes in the mean and contrast of an image. It also makes sense that structural approaches are *sensitive* to distortions that break down natural spatial correlation of an image: blur, blocking, ringing, and noise fall into this category.

As described in [8], the structural philosophy can be implemented using a set of equations defining the Structural SIMilarity (SSIM) quality metric. Luminance, contrast, and structure are measured separately. Given two images (or image patches) $\mathbf{x}$ and $\mathbf{y}$ to be compared, *luminance* is estimated as the mean μ of each image, *contrast* is estimated as the standard deviation σ, and *structure* ς is estimated

from the image vector $\mathbf{x}$ by removing the mean and normalizing by the standard deviation.

Then, the measurements $\mu_x, \mu_y, \sigma_x, \sigma_y, \varsigma_x, \varsigma_y$ are combined using a luminance comparison function $l(\mathbf{x}, \mathbf{y})$, contrast comparison function $c(\mathbf{x}, \mathbf{y})$, and structure comparison function $s(\mathbf{x}, \mathbf{y})$ to give a composite measure of structural similarity:

$$SSIM(\mathbf{x}, \mathbf{y}) = l(\mathbf{x}, \mathbf{y})^{\alpha} \cdot c(\mathbf{x}, \mathbf{y})^{\beta} \cdot s(\mathbf{x}, \mathbf{y})^{\gamma} \quad (1)$$

where α, β, γ are positive constants used to weight each comparison function.

Using the comparison functions defined in [8] and setting $\alpha = \beta = \gamma = 1$ gives the specific SSIM quality metric

$$SSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

2.2. CWSSIM review

As suggested in [10], it is straightforward to implement a structural similarity metric in the complex wavelet domain. As more wavelet-based image and video coding techniques are coming into use, it makes sense to be able to implement image quality metrics in this domain. In addition, if an application requires an image quality metric that is unresponsive to spatial translation, this extension of SSIM can be adapted in a way such that it has low sensitivity to small translations. This requires an overcomplete transform such as steerable pyramid decomposition where *phase information is available*.

Given complex wavelet coefficients $\mathbf{c}_x$ and $\mathbf{c}_y$ that correspond to image patches $\mathbf{x}$ and $\mathbf{y}$ to be compared, the complex wavelet structural similarity (CWSSIM) is given by:

$$CWSSIM(\mathbf{c}_x, \mathbf{c}_y) = \frac{2|\sum_{i=1}^{N} c_{x,i} c_{y,i}^*| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K} \quad (3)$$

where K is a small positive constant set to 0.03 in this paper. This equation differs from (2) because the wavelet filters we use are bandpass (i.e. they have no response at zero frequency) forcing the mean of the wavelet coefficients to zero ($\mu_x = \mu_y = 0$). This fact acts to cancel the $(2\mu_x\mu_y + C_1)$ and $(\mu_x^2 + \mu_y^2 + C_1)$ terms of (2).

In [10], Wang and Simoncelli note that wavelet coefficient phase is the key factor that determines the structural distortion results of this metric, emphasizing that "the structural information of local image features is mainly contained in the relative phase patterns of the wavelet coefficients". Linear scaling of the coefficients corresponds to lighting (brightness and contrast) distortions to which CWSSIM is not very sensitive because the *structure* is not perturbed (it is sensitive in a power-law sense similar to Weber's law). Consistent phase shift of the coefficients corresponds to spatial translation, another distortion to which CWSSIM is not strongly sensitive. Phase changes that vary irregularly from one coefficient to the next produce structural distortion to which CWSSIM is very sensitive.

Fig. 1. Original grayscale images used for coder comparisons.

The structural similarity metric gives a result in the range from 0.0 to 1.0, where zero corresponds to a loss of all structural similarity and one corresponds to having an exact copy of the original image. Images with lighting-related distortions alone give high SSIM while other distortions result in low similarity, corresponding well with the intuitive perception of quality.

We use the WCWSSIM, a form of CWSSIM that uses weighted results from multiple subbands, where the weights are derived from the HVS contrast sensitivity function [14]). We found that this modification can better handle local mean shift distortions [14]).

3. USING SSIM TO ASSESS IMAGE QUALITY OF JPEG, JPEG2000, SPIHT, & PIC CODERS

As we discussed above, perceptual image quality metrics have been based on the human visual system's sensitivity to just-noticeable distortions. When images are compressed beyond the threshold of distortion, the perceptual metrics fail to provide meaningful measurements of the HVS's response to the severe artifacts [2]. While perceptual metrics have been used in suprathreshold applications [6], they do not account for the wide variety of compression artifacts generated by image and video coders at low bit rates. Moreover, as pointed out in [1], they tend to be biased towards coders that have the same subband structure. SSIM-based image quality metrics conveniently avoid this problem by focusing on the top-down image formation concept that the local structure of images is the most important aspect of image quality. The ability of structural similarity to distinguish between structural and non-structural distortions leads to results that agree with perception for severely distorted images.

In order to explore the utility of structural similarity metrics for evaluating compression algorithms, we test a set of coders similar to the work in [3]. The coders tested are: the standard JPEG algorithm with a perceptually weighted quantization table optimized for 6 image heights [11]; the Safranek and Johnston Perceptual subband Image Coder (PIC) [12] based on a perceptual masking model (two versions of this coder were used, with 4×4 and 8×8 subband decompositions); the Said-Pearlman SPIHT zero-tree wavelet coder [13] with perceptual weighting added; and finally the baseline JPEG2000 coder which is also wavelet-based.

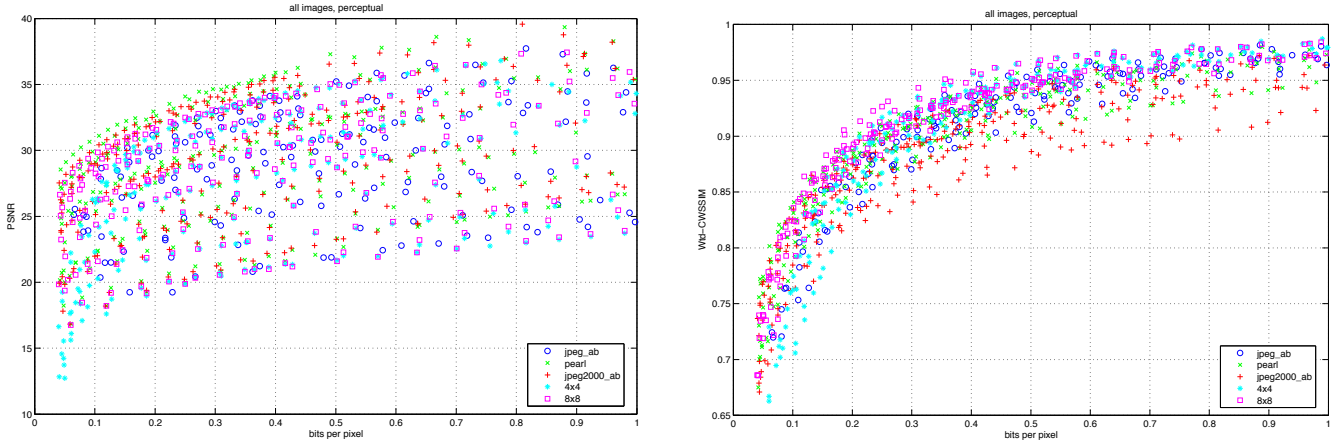
We developed a small database of approximately 1800 compressed images created from 13 original natural images displayed in Figure 1 with compression ranging from 0.1 to 2.0 bits per pixel (b/p) for JPEG, SPIHT, JPEG2000, PIC 4x4, and PIC 8x8. The libjpeg [1] and JasPer [2] reference software was used to create the JPEG and JPEG2000 images, respectively. The SPIHT and PIC images were created using software provided by the corresponding authors.

The compressed images were evaluated for image quality using PSNR and the WCWSSIM described in Section 2.2. The focus of this paper is in comparing WCWSSIM to mean squared error (PSNR). Comparison with perceptually weighted metrics is beyond the scope of this paper. For some interesting comparisons with perceptually weighted metrics, see [1].

Figure 2 shows scatter plots of the quality metrics for the entire image database. As might be expected, PSNR varies considerably based on image content alone. Even at the relatively high bitrate of 1.0 b/p, PSNR ranges from 25 dB to 37 dB. The WCWSSIM scatter plot is more tightly clustered with no obvious banding due to particular image content, therefore, it is measuring the structural distortion due to compression artifacts and is much less lasensitive to

[1] libjpeg is available at http://www.iig.org/

[2] JasPer is available at http://www.ece.uvic.ca/ mdadams/jasper/



(a) PSNR vs b/p

(b) WCWSSIM vs b/p

Fig. 2. PSNR and WCWSSIM metrics for JPEG, JPEG2000, SPIHT ("pearl" in legend), PIC 4x4, and PIC 8x8 coders. This plot shows combined results from the entire database of 13 different images compressed with each coder from 0.05 to 1.0 bits per pixel. Plot (a) exhibits banding due to variation from image to image, while plot (b) shows consistent results indicating that the numeric results of WCWSSIM have more meaning in the absolute sense.

Fig. 3. "Rose" image compressed to 0.5 bits per pixel with JPEG2000 (left) and PIC 8x8 (right). The JPEG2000 image has PSNR=23.2 and WCWSSIM=0.88 while the PIC 8x8 image has PSNR=21.6 and WCWSSIM=0.94.

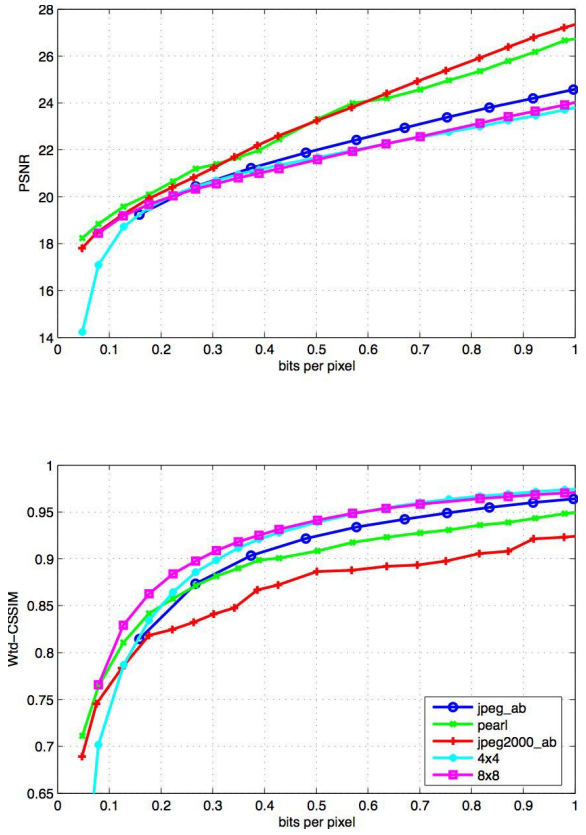


Fig. 4. PSNR (top) and WCWSSIM (bottom) for "rose" compressed with JPEG, JPEG2000, SPIHT ("pearl" in legend), PIC 4x4, and PIC 8x8 coders.

intra-image differences. These results indicate that the WCWSSIM numbers are more meaningful in comparisons across images.

Another notable observation is that PSNR favors the SPIHT and JPEG2000 images because they are optimized toward minimizing mean-squared error. Figure 4 shows the relative performance of different compression techniques according to PSNR and WCWSSIM metrics for the "rose" image. Note that, according to WCWSSIM, JPEG2000 has the worst performance especially at high bit rates, generally agreeing with perceived quality in informal subjective evaluations. This can be accounted for in part by the fact that the baseline JPEG2000 implementation uses no perceptually weighted quantization table, while all the other techniques do.

A specific example is shown in Figure 3, where compressed "rose" images are displayed at 0.5 b/p for JPEG2000 and PIC 8x8 coders. WCWSSIM indicates that the PIC 8x8 image has significantly higher quality than the JPEG2000 image. Indeed, the perceptual quality of these images correlates well with WCWSSIM's predictions. Especially in the pavement region, the JPEG2000 blur distortions are quite noticeable. WCWSSIM detects this loss of structure, giving a much higher quality score to PIC 8x8. These results are very similar for the rest of the images in the database, especially in the suprathreshold range of 0.3 to 1.0 b/p.

Overall, our results indicate that WCWSSIM provides results that agree well with informal subjective evaluation of images with suprathreshold levels of distortion. In addition, it provides an unbiased metric for comparison across coders and across images. As such, we believe that it holds great potential for future coder development and evaluation.

4. REFERENCES

[1] T. N. Pappas, R. J. Safranek, and J. Chen, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing, 2nd Ed.*, A. C. Bovik, Ed. Academic Press, 2005.

[2] T. N. Pappas, T. A. Michel, and R. O. Hinds, "Supra-threshold perceptual image coding," 1996, vol. I of *Proc. Int. Conf. Image Processing (ICIP-96)*, pp. 237–240.

[3] J. Chen and T. N. Pappas, "Perceptual coders and perceptual metrics," in *Human Vision and Electronic Imaging VI*, San Jose, CA, Jan. 2001, Proc. SPIE Vol. 4299, pp. 150–162.

[4] S. S. Hemami and M. G. Ramos, "Wavelet coefficient quantization to produce equivalent visual distortion in complex stimuli," in *Human Vision and Electronic Imaging V*, San Jose, CA, Jan. 2000, Proc. SPIE Vol. 3959, pp. 200–210.

[5] M. G. Ramos and S. S. Hemami, "Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *J. Opt. Soc. Am. A*, Oct. 2001.

[6] D. M. Chandler and S. S. Hemami, "Additivity models for suprathreshold distortion in quantized wavelet-coded images," in *Human Vision and Electronic Imaging VII*, San Jose, CA, Jan. 2002, Proc. SPIE Vol. 4662, pp. 105–118.

[7] D. M. Chandler and S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," *J. Opt. Soc. Am. A*, vol. 20, no. 7, July 2003.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[9] J. Lubin, "The use of psychophysical data and models in the analysis of display system performance," in *Digital Images and Human Vision*, A. B. Watson, Ed., pp. 163–178. The MIT Press, 1993.

[10] Z. Wang and E. P. Simoncelli, "Translation insensitive image similarity in complex wavelet domain," in *IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, PA Philadelphia, Ed., 2005, vol. II of *Proc. IEEE*, pp. 573–576.

[11] H. A. Peterson, A. J. Ahumada Jr., and A. B. Watson, "Improved detection model for dct coefficient quantization," 1993, vol. 1913 of *Proc. Int. Conf. Human Vision, Visual Processing, and Digital Display (HVEI-99)*.

[12] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantization data compression," 1989, vol. 3 of *Proc. ICASSP-89*, pp. 1945–1948.

[13] A. Said and W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 6, pp. 243–250, 1996.

[14] A. C. Brooks and T. N. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distortions," in *Human Vision and Electronic Imaging XI*, San Jose, CA, Jan. 2006, Proc. SPIE Vol. 6057.