A FULL-REFERENCE COLOR IMAGE QUALITY MEASURE IN THE DWT DOMAIN

Devon Gayle, Hazem Mahlab, Yuksel Ucar, Ahmet M. Eskicioglu[1]
Department of Computer and Information Science, CUNY Brooklyn College
2900 Bedford Avenue, Brooklyn, NY 11210, USA

ABSTRACT

In this paper, we present a new objective quality measure for color images. In any frequency domain transform, the coefficients in different frequency bands have different magnitudes. The 2-dimensional Discrete Wavelet Transform (DWT) separates a given image into four bands: LL, HL, LH, and HH. After applying the DWT to both the original and degraded images, we compute the absolute value of the difference of the magnitudes in each band, and obtain the standard deviation (SD) of these differences. The proposed measure is defined as the mean of four SD values. Correlation of the subjective ratings and objective scores gives the performance of the measure. A comparison with the peak signal-to-noise ratio (PSNR), and two state-of-the-art metrics, Q and MSSIM, shows that our results correlate better with the judgment of human observers.

1. INTRODUCTION

An important criterion used in the classification of image quality measures is the type of information needed to evaluate the distortion in degraded images. Measures that require both the original image and the distorted image are called “full-reference” or “non-blind” methods, measures that do not require the original image are called “no-reference” or “blind” methods, and measures that require both the distorted image and partial information about the original image are called “reduced-reference” methods.

Although no-reference measures are needed in some applications in which the original image is not available, they can be used to predict only a small number of distortion types. In the current literature, a few papers attempt to predict JPEG compression artifacts [1,2,3,4], and others blurring and JPEG 2000 artifacts [5,6]. Reduced-reference measures are between full-reference and no-reference measures; [7] evaluates the quality of JPEG and JPEG2000 coded images whereas [8] provides assessment for JPEG and JPEG200 compressed images, images distorted by white Gaussian noise, Gaussian blur, and the transmission errors in JPEG2000 bit streams. The applicability of full-reference measures is much wider. They can be used to estimate a spectrum of distortions that range from blurriness and blockiness to several types of noise. Recent examples of such measures are given in Table 1.

Table 1. Full-reference image quality measures

Publication	**Domain type**	**Type of distortion predicted**
Wang and Bovik [9]	Pixel	Impulsive salt-pepper noise, additive Gaussian noise, multiplicative speckle noise, mean shift, contrast stretching, blurring, and JPEG compression
Wang, Bovik, Sheikh and Simoncelli [10]	Pixel	JPEG compression, JPEG 2000 compression
Van der Weken, Nachtegael and Kerre [11]	Pixel	Salt and pepper noise, enlightening, and darkening
Beghdadi and Pesquet-Popescu [12]	Discrete Wavelet Transform (DWT)	Gaussian noise, grid pattern, JPEG compression

Two of the state-of-the-art image quality metrics are the universal image quality index (Q) [9] and the Structural Similarity Index (SSIM) [10]. The universal image quality index, Q, is defined as

$$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)]},$$

where $x_i, y_i, i = 1,...n$, represent the original and distorted signals, respectively, $\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i$,

[1] Corresponding author’s email address: eskicioglu@sci.brooklyn.cuny.edu

$$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)^2,\ \sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \mu_y)^2,$$

$$\text{and } \sigma_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y).$$

The dynamic range of Q is [-1,1], with the best value achieved when $y_i = x_i$, $i = 1,2,\ldots,n$. The index is computed for each window, leading to a quality map of the image. The overall quality index is the average of all the Q values in the quality map:

$$Q = \frac{1}{M}\sum_{j=1}^{M} Q_j,\ M = \text{total number of windows.}$$

Q produces unstable results when either $(\mu_x^2 + \mu_y^2)$ or $(\sigma_x^2 + \sigma_y^2)$ is very close to zero. In order to circumvent this problem, the measure has been generalized to the Structural Similarity Index (SSIM):

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Q is a special case of SSIM that can be derived by setting C_1 and C_2 to zero. As in the case of Q, the overall image quality MSSIM is obtained by computing the average of SSIM values over all windows:

$$\text{MSSIM} = \frac{1}{M}\sum_{j=1}^{M}\text{SSIM}_j$$

In this paper, we propose a new image quality measure in the DWT domain using the magnitudes of DWT coefficients.

2. NEW IMAGE QUALITY MEASURE: M-DWT

For a given frequency domain transform (e.g., DCT, DWT, and DFT), the coefficients in different frequency bands have different magnitudes. The process of separating the frequency bands using the DWT is well-defined. In two-dimensional DWT, each level of decomposition produces four bands of data denoted by LL, HL, LH, and HH.

The YUV color model is a linear transformation between the gamma-corrected RGB components that produces a luminance signal and a pair of chrominance signals. A common approach employed in developing a quality measure for color images is to use only the luminance signal.

Our proposed algorithm, M-DWT, is as follows:

1. Apply DWT to the luminance layer of the original image.
2. Apply DWT to the luminance layer of the degraded image.
3. For each frequency band, perform the following operations:
   a. Obtain the magnitudes M_{oi}, i=1,…,n of original DWT coefficients.
   b. Obtain the magnitudes M_{di}, i=1,…,n of degraded DWT coefficients.
   c. Compute the absolute value of the differences: $|M_{oi}-M_{di}|$, i=1,…,n.
   d. Compute the standard deviation of the differences.
4. Obtain the mean of four standard deviations.

The measure was applied to a full color, 24-bit version of 512x512 Lena. Table 2 shows the tools and parameters for six degradation types, and five degradation levels. Note that all of these degradations were performed in the pixel domain.

Table 2. Distortion types and distortion levels

Type \ Level	Level 1	Level 2	Level 3	Level 4	Level 5
JPEG (XnView)	20:1	40:1	60:1	80:1	100:1
JPEG2000 (XnView)	20:1	40:1	60:1	80:1	100:1
Gaussian blur (Photoshop)	1	2	3	4	5
Gaussian noise (Photoshop)	3	6	9	12	15
Sharpening (XnView)	10	20	30	40	50
DC-shifting (Programming	4	8	12	16	20

High quality print-outs of 30 distorted full color images were subjectively evaluated by 14 observers. The printer was a Hewlett-Packard printer with model number "hp color Laserjet 4600dn." The 8-2/16"x8-2/16" images were printed on 8.5"x11" white paper with the basis weight 20lb and brightness 84. The observers were chosen among the graduate students and instructors from the Department of Computer and Information Science at Brooklyn College. About half of the observers were familiar with image processing, and the others only had computer science background. They were asked to rate the images using a 50-point scale in two ways: Within a given distortion type (i.e, rating of the 5 distorted images), and across six distortion types (i.e., rating of the 6 distorted images for each distortion level). As the proposed measure is not HVS-based, no viewing distance was imposed on the observers in the experiment. Grade 1 was assigned to the best image, and grade 50 was assigned to the worst image.

The subjective ratings and objective scores are given in Tables 3 and 4, respectively. MOS is the average of the ratings by human observers.

Table3. Mean opinion score (MOS)

	JPEG	**JPEG 2000**	**Blur**	**Noise**	**Sharp**	**DC-shift**
Level 1	5.267	4	13.2	9.467	2.133	1.8
Level 2	10.533	7.8	24.133	16.667	3.933	3.4
Level 3	19.467	11.667	34.2	21.8	5.9333	5.067
Level 4	28.467	15.333	42.4	27.6	8.867	6.867
Level 5	35.2	20.067	49.933	32.867	11.667	8.733

Table 4. Objective scores

	JPEG	**JPEG 2000**	**Blur**	**Noise**	**Sharp**	**DC-shift**
Level 1	6.297	5.183	9.898	8.192	1.046	0.231
Level 2	8.291	6.996	14.795	14.026	1.608	0.373
Level 3	9.659	8.075	19.914	17.991	2.205	0.741
Level 4	11.043	9.088	23.327	22.453	2.942	1.250
Level 5	12.582	9.877	24.811	22.457	3.865	1.862

In the Video Quality Experts Group (VQEG) Phase I and Phase II testing and validation, a nonlinear mapping was used between the objective model outputs and subjective quality ratings [13]. The performance of each proponent model was evaluated after compensating for the nonlinearity. To establish a nonlinear mapping, we followed the same procedure by fitting the logistic curve

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1+\exp(\tau x)}, \ \tau: \text{a constant parameter.}$$

We will compare the performance of M-DWT with PSNR, and two state-of-the-art metrics, Q and MSSIM.

Figure 1 shows the scatter plots for the four measures using the luminance layer of Lena for thirty images. Each marked point represents the corresponding values in Tables 3 and 4.

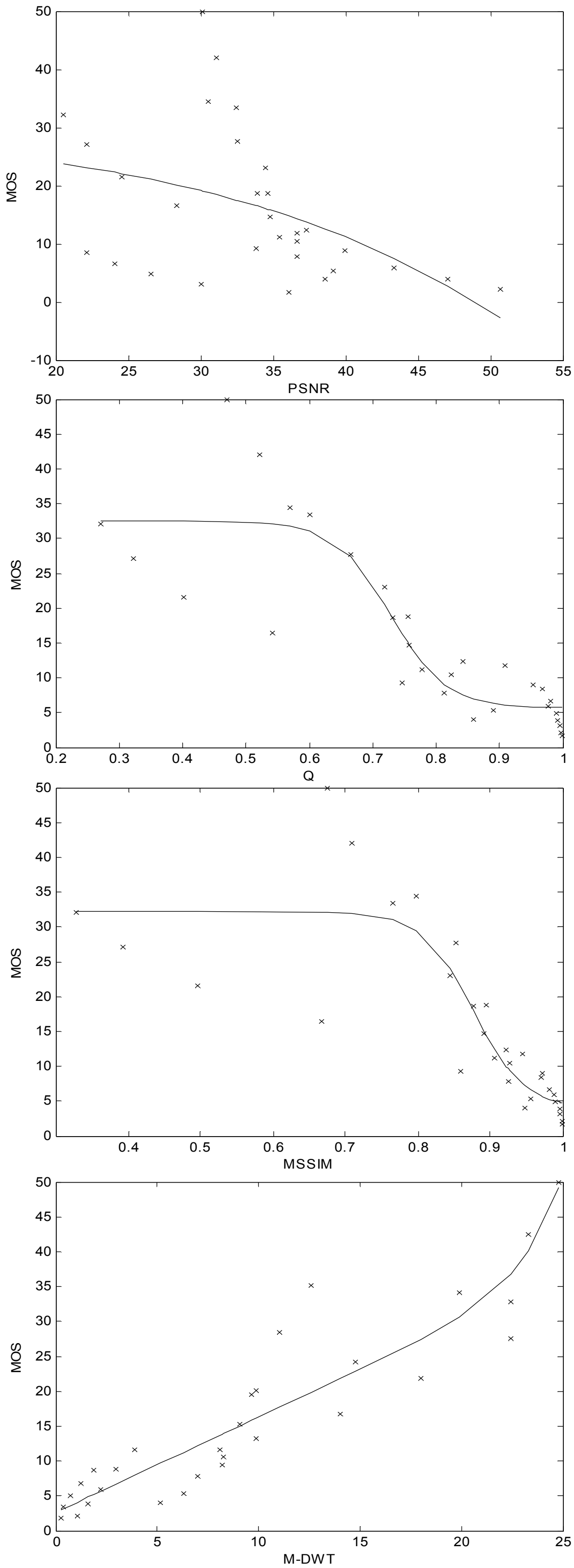

Figure 1. Comparison of scatter plots for PSNR, Q, SSIM, and M-DWT

Table 5 displays the overall performance of each measure using the correlation between MOS and objective scores.

Table 5. Comparison of four measures

Criterion\ Measure	PSNR	Q	MSSIM	M-DWT
Correlation	0.455	0.886	0.873	0.9208

The performance of a good objective measure can be determined by its ability to predict the quality not only within a given distortion type but also across different distortion types. It will be shown that simple quality measures like PSNR perform well for all distortion types. We therefore computed two additional sets of data to compare the performance of the measures: (1) Correlation within each of the 6 distortion types, and (2) Correlation across each of the 5 distortion levels.

These correlations are given in Tables 6 and 7.

Table 6. Performance within each distortion type

Distortion type\Measure	PSNR	Q	MSSIM	M-DWT
Gaussian blur	1.000	0.999	0.999	0.996
Gaussian noise	1.000	0.998	0.996	0.979
JPEG	1.000	1.000	1.000	1.000
JPEG2000	0.999	0.999	1.000	0.999
Sharpening	1.000	0.999	0.998	1.000
DC-shifting	1.000	0.999	0.999	1.000

Table 7. Performance across each distortion level

Distortion level\Measure	PSNR	Q	MSSIM	M-DWT
1	0.485	0.926	0.935	1.000
2	0.439	0.943	0.954	0.998
3	0.259	0.937	0.937	0.947
4	0.163	0.934	0.936	0.938
5	0.194	0.925	0.932	0.950

3. CONCLUSIONS

We presented a new color image quality measure based on the DWT. As it does not incorporate a Human Visual System (HVS) model, we do not use any assumptions regarding the viewing distance. In the experiments, a wide range of distortion types was used. For each distortion type, five levels of distortion were introduced. Although PSNR is still widely used by researchers, our results indicate that it is an unreliable measure, especially for correlation across distortion levels. The performance of M-DWT has a perfect match with the quality perceived by human observers. The proposed measure is also superior to state-of-the-art metrics Q and MSSIM. In future work, we will use more color images, and extend the measure for evaluating the quality of watermarked images and video sequences.

REFERENCES

1. A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 7-11, 2001.
2. Z. Wang, A. C. Bovik and B. L. Evans, "Blind measurement of blocking artifacts in images," *Proceedings of IEEE 2000 International Conferencing on Image Processing*, Vancouver, BC, Canada, September 10-13, 2000.
3. Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," *Proceedings of IEEE 2002 International Conferencing on Image Processing*, Rochester, NY, September 22-25, 2002.
4. L. Meesters and J.-B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Processing*, Vol. 82, pp. 369-387, 2002.
5. P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," *IEEE 2002 International Conference on Image Processing*, Rochester, NY, September 22-25, 2002.
6. E.-P. Ong, W. Lin, Lu, Z. Yang, S. Yao, F. Pan, L. Jiang and F. Moschetti, "A no-reference quality metric for measuring image blur," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
7. M. Carnec, P. Le Callet and D. Barba, "An image quality assessment method based on perception of structural information," *2003 International Conference on Image Processing*, Barcelona, Spain, September 14-17, 2003.
8. Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," Human Vision and Electronic Imaging X Conference, San Jose, CA, January 17-20, 2005, Proceedings of SPIE, Vol. 5666.
9. Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp. 81-84, March 2002.
10. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error measurement to structural similarity," *IEEE Transactions on Image Processing*, Vol. 13, No. 4, April 2004.
11. D. Van der Weken, M. Nachtegael and E. E. Kerre, "A new similarity measure for image processing," *Journal of Computational Methods in Sciences and Engineering*, Vol. 3, No. 2, pp. 209-222, 2003.
12. A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
13. A. M. Rohaly, J. Libert, P. Corriveau, and A. Webster (editors), "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," March 2000.