Weighted Structural Similarity Based on Edge Strength for Image Quality Assessment

D.Venkata Rao and L.Pratap Reddy

***Abstract*— In this paper a full reference objective image quality assessment technique is presented which is based on the properties of the human visual system (HVS). By integrating the notion of visual regions of interest with the measurement of structural similarity between the original image and distorted image a Weighted Structural Similarity Index (WSSI) is proposed. The method first evaluates the structural similarity indices between the original and distorted image in local regions. These local indices are then weighted based on the visual region of interest of the corresponding region, characterized by edge strength in the local region. WSSI of an image is calculated as the average of these weighted indices. A comparison with the peak-signal-to-noise ratio (PSNR) and state of the art metric, Mean Structural Similarity Index (MSSIM), shows that the proposed measure correlates better with the judgment of human observers.**

***Keywords*— Edge strength, Human visual system, Structural similarity, Visual regions of interest.**

I. INTRODUCTION

The role of images in present day communication has been steadily increasing. In this context the quality of an image plays a very important role. Different stages and multiple design choices at each stage exist in any image processing system. They have direct bearing on the quality of the resulting image. Unless we have a quantitative measure for the quality of an image, it becomes difficult to design an ideal image processing system. Though subjective quality assessment is an alternative, it is not feasible to be incorporated into real world systems. Hence, objective quality metrics play an important role in image quality assessment.

In the last two decades a lot of objective metrics have been proposed [1-6] to assess image quality. The most widely adopted statistics feature is the Mean Squared Error (MSE). However, MSE and its variants do not correlate well with subjective quality measures because human perception of image distortions and artifacts is unaccounted for. MSE is also not good because the residual image is not uncorrelated additive noise. It contains components of the original image. A detailed discussion on MSE is given by Girod [7].

A major emphasis in recent research has been given to a deeper analysis of the Human Visual System (HVS) features [1]. There are lot of HVS characteristics [8] that may influence the human visual perception on image quality. Although HVS is too complex to fully understand with present psychophysical means, the incorporation of even a simplified model into objective measures reportedly leads to a better correlation with the response of the human observers [1]. However, most of these methods are error sensitivity based approaches, explicitly or implicitly, and make a number of assumptions [9], which need to be validated. These methods suffer from the problems like the natural image complexity problem, Minkowski error pooling problem, and cognitive interaction problem [9].

Structural similarity based methods [10,11] of image quality assessment claim to account for the fact that the natural image signal samples exhibit strong dependencies amongst themselves, which is ignored by most of these methods. Structural similarity based methods replace the Minkowski error metric with different measurements that are adapted to the structures of the reference image signal, instead of attempting to develop an ideal transform that can fully decouple signal dependencies.

However, Vision models [12, 13], which treat visible distortions equally, regardless of their location in the image, may not be powerful enough to accurately predict picture quality in such cases. This is because we are known to be more sensitive to distortions in areas of the image to which we are paying attention than to errors in peripheral areas.

In this paper we present an image quality metric, which integrates the notions of structural similarity measure mimicking the overall functionality of HVS and visual regions of interest based on edge strength. We observed that the proposed index correlates effectively with subjective scores and found to posses superior performance when compared with other metrics discussed in this paper.

This paper is organized as follows. Section 2 explains the structural similarity method. Section 3 describes the computation of proposed quality index. Experimental results follow in Section 4. Finally, in Section 5, the conclusions of the paper are presented.

II. STRUCTURAL SIMILARITY

Based on the assumption that the HVS is highly adapted to extract structural information from the viewing field, a new philosophy of $SSIM$ for image quality measurement was proposed by Wang et al [11]. Let x and y be two discrete non-negative signals $x = \{x_i \mid i = 1, 2, ..., N\}$ and $y = \{y_i \mid i = 1, 2, ..., N\}$ that have been aligned with each other and let $\bar{x}$, s_x^2 and s_{xy} be the mean of x, variance of x, and the covariance of x and y respectively. $\bar{x}$, s_x^2 are the estimates of the luminance and contrast of x, and s_{xy} measures the tendency of x and y to vary together, which is an indication of structural similarity. $SSIM$ index is given as shown below where, C_1, C_2 and C_3 are small constants introduced to avoid instability when the denominator is close to zero.

$$SSIM(x, y) = \frac{(2m_x m_y + C_1)(2s_{xy} + C_2)}{(m_x^2 + m_y^2 + C_1)(s_x^2 + s_y^2 + C_2)} \quad \textbf{(1)}$$

III. Weighted structural similarity index

At first, the original and distorted images are divided into 8 x 8 non-overlapping blocks. The $SSIM$ for each block is computed using equation (1), to form a matrix S, as shown below where each element s_{ij} represents the structural similarity between corresponding blocks of the original and distorted images with coordinates (i, j), $1 \le i \le m = \lfloor H/8 \rfloor$, $1 \le j \le m = \lfloor W/8 \rfloor$, where H and W represent the height and width of the image respectively.

Psycho visual studies reveal that human eye is very sensitive to the edge and contour information of the image. Edges per unit area e, was determined by detecting edges in an image, using the Canny extension of the Sobel operator [14] and then congregating the edges detected within an 8x8 block. The value of e is normalized to the range [0 1]. A block without edges will have a value of 0.

Secondly, the visual regions of interest map E as specified above is obtained for the original image as shown below. The values e_{ij} represents degree of visual importance of each block with coordinates (i, j) as defined earlier.

$$S = \begin{pmatrix} s_{11} & s_{12} & . & . & s_{1n} \\ s_{21} & s_{22} & . & . & s_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ s_{m1} & s_{m2} & . & . & s_{mn} \end{pmatrix} \qquad E = \begin{pmatrix} e_{11} & e_{12} & . & . & e_{1n} \\ e_{21} & e_{22} & . & . & e_{2n} \\ . & . & . & . & .. \\ . & . & . & . & . \\ e_{m1} & e_{m2} & . & . & e_{mn} \end{pmatrix}$$

We define Weighted Structural Similarity index $WSSI$ as the weighted average of the structural similarity indices s_{ij} in each local block with coordinates (i, j), where each s_{ij} is weighted with the corresponding visual region of interest values e_{ij}. Equation (4) gives the expression for $WSSI$

$$WSSI = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} S(i, j)E(i, j)}{\sum_{i=1}^{m}\sum_{j=1}^{n} E(i, j)} \quad \textbf{(2)}$$

IV. Experimental Results

The proposed quality index was tested using LIVE image database [15]. The database consists of twenty-nine high resolution 24-bits/pixel RGB color images (typically 768 x 512), distorted using five distortion types: JPEG2000, JPEG, White noise in the RGB components, Gaussian blur in the RGB components, and bit errors in JPEG2000 bit stream using a fast-fading Rayleigh channel model. Each image was distorted with each type, and for each type the perceptual quality covered the entire quality range. Difference Mean Opinion Score (DMOS) value for each distorted image was computed based on the perception of quality of the images by observers.

We tested the proposed method on all the images and distortions available in the LIVE database, after converting the color images to gray level images. In order to provide quantitative measures on the performance of the objective quality assessment models, different evaluation metrics were adopted in the Video Quality Experts Group (VQEG) Phase-I test [16]. We performed non-linear mapping between the objective and subjective scores, using 4-parameter logistic function of the form shown in Equation (3).

$$y = a/(1.0 + e^{-(x-b)/c}) + d \quad \textbf{(3)}$$

After the non-linear mapping, the Correlation Coefficient (CC), the Mean Absolute Error (MAE), and the Root Mean Squared Error (RMS) between the subjective and objective scores are calculated as measures of prediction accuracy. The prediction consistency is quantified using the outlier ratio (OR), which is defined as the percentage of the number of predictions outside the range of ± 2 times the standard deviation. Finally, the prediction monotonicity is measured using the Spearman rank-order-correlation coefficient (ROCC).

To evaluate the performance of the proposed metric, we considered two image quality assessment models, PSNR and $MSSIM$. Table 1 shows the evaluation results for the models being compared with that of the $WSSI$ for different types of distortions. For each of the objective evaluation criteria, $WSSI$ outperforms the other models being compared across different distortion types. Figure 1 shows the scatter plots of DMOS versus $WSSI$ for different kinds of distortions.

Table 1. Performance comparison of image quality assessment models on LIVE image database [15].

(a) JPEG2000 (b) JPEG (c) White noise (d) Gaussian blur (e) Fast fading

Model	CC	ROCC	MAE	RMS	OR%
PSNR	0.859	0.851	6.454	8.269	5.917
MSSIM	0.899	0.894	5.687	7.077	2.366
WSSI	0.931	0.925	4.773	5.929	4.142

(a)

Model	CC	ROCC	MAE	RMS	OR%
PSNR	0.842	0.828	6.636	8.622	6.285
MSSIM	0.891	0.863	5.386	7.236	5.714
WSSI	0.917	0.882	4.563	6.377	6.857

(b)

Model	CC	ROCC	MAE	RMS	OR%
PSNR	0.922	0.938	4.524	6.165	5.555
MSSIM	0.94	0.914	4.475	5.459	2.777
WSSI	0.962	0.9545	3.526	4.367	4.166

(c)

Model	CC	ROCC	MAE	RMS	OR%
PSNR	0.744	0.725	8.395	10.50	3.448
SSIM	0.947	0.940	3.992	5.027	3.448
WSSI	0.968	0.963	3.168	3.897	4.137

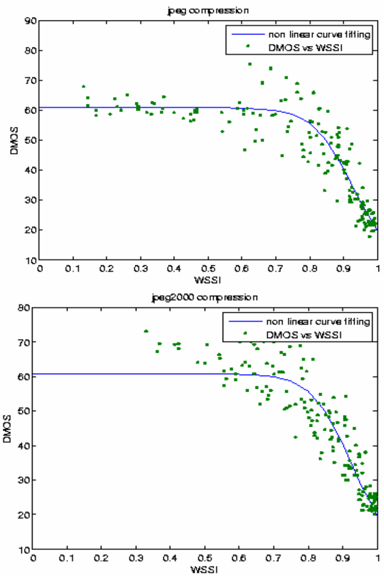
(d)

Model	CC	ROCC	MAE	RMS	OR%
PSNR	0.857	0.859	6.383	8.476	6.896
MSSIM	0.956	0.945	3.806	4.799	5.517
WSSI	0.962	0.960	3.656	4.467	2.758

(e)

V. Conclusion

In this paper we present an image quality assessment technique, which is based on the properties of the human visual system (HVS). It combines the notions of structural similarity with visual regions of interest. The results prove the fact that human eye is sensitive to edges that are present in the image, quantified by edge strength in the local block. The results also justify the visual regions of interest built on this. Statistical indices of performance as set by VQEG for the proposed quality index indicate that the index matches well with the Human Visual System obviating the need for subjective tests and proves to be a better choice than other indices mentioned in the paper. The index is found to have good sensitivity across all the distortion types mentioned.



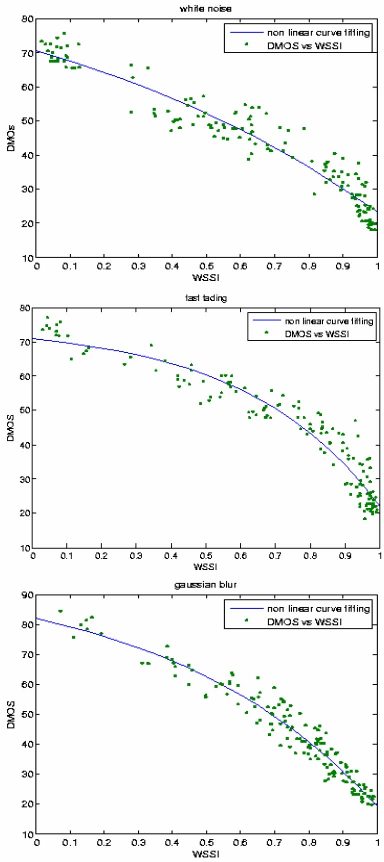


Fig. 1. Scatter plots for DMOS versus model prediction for JPEG, JPEG2000, White noise, Fast fading, Gaussian blur distorted images

References

[1] A. M. Eskicioglu and P. S. Fisher, "Image quality *measures* and their performance," IEEE Transactions on Communications, vol. 43, no. 12, pp. 2959-2965, Dec. 1995.
[2] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," IEEE Transactions on Image Processing, vol. 4, no. 6, pp. 713-724, June 1995.
[3] J. A. Saghri, "Image quality measure based on a human visual system model," Optical Engineering, vol. 28, no. 7, pp. 813-818, July 1989.
[4] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," Digital Images and Human Vision, A. B. Watson Ed., Chapter 14, pp. 179-206, the MIT press, 1993.
[5] J. Lubin, "A visual discrimination model for imaging system design and evaluation," Vision Models for Target Detection and Recognition, E. Peli ed., Chapter 10, pp. 245-283, World Scientific Publishing Co. Pte. Ltd., 1995.
[6] A. B. Watson, "DCT quantization matrices visually optimize for individual images," Human Vision, Visual Processing and Digital Display IV, Proc. SPIE, vol. 1913, pp. 202-216, 1993.
[7] B. Girod, "What's wrong with mean-squared error," in Digital Images and Human Vision, A. B. Watson, ed., pp. 207-220, MIT Press, 1993.
[8] B. A. Wandell, Foundations of Vision, Sinauer Associates, Inc., 1995.
[9] Z.Wang, H.R. Sheikh and A.C. Bovik, “Objective video quality assessment,” in The Handbook of Video Databases: Design and Applications (B.Furht and O. Marques, eds.), pp. 1041-1078, CRC press, Sept 2003.
[10] Z.Wang L. Lu and A .C. Bovik, “Video quality assessment based on structural distortion measurement,” Signal Processing: Image Communication, special issue on objective video quality metrics, vol. 19, Jan. 2004.

[11] Z.Wang A.C. Bovik, H. R. Sheikh, and E .P. Simocelli, “Image quality assessment: From error measurement to structural similarity,” IEEE Trans. Image Processing, vol. 13, no.4, pp.600-612, Apr .2004.
[12] G. A. Geri and Y. Y. Zeevi. Visual assessment of variable-resolution imagery. Journal of the Optical Society of America, 12(10):2367-2375, October 1995.
[13] P. Kortum and W. Geisler.Implementation of a foveated image coding system for image bandwidth reduction. In SPIE - Human Vision and ElectronicImaging, volume 2657, pages 350-360, February 1996.
[14] J. Canny, “A Computational Approach to Edge Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 8, no. 6,pp. 679-698, 1986.
[15] H. R. Sheikh, A. C. Bovik, L. Cormack, and Z. Wang, “LIVE Image Quality Assessment Database,”2004, Available On line at http://live.ece.utexas.edu/research/quality
[16] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” http://www.vqeg.org/, Mar. 20004.