Image Quality Assessment Based on Perceptual Structural Similarity

D. Venkata Rao and L. Pratap Reddy

Bapatla Engineering College, Bapatla, India
JNTU College of Engineering, Hyderabad, India
dv2002in@yahoo.co.in,pratplr@rediffmail.com

Abstract. We present a full reference objective image quality assessment technique which is based on the properties of the human visual system (HVS). It consists of two major components: 1) structural similarity measurement (SSIM) between the reference and distorted images, mimicking the overall functionality of HVS in a top down frame work. 2) A visual attention model which indicates perceptually important regions in the reference image based on the characteristics of intermediate and higher visual processes through the use of Importance Maps. Structural similarity in a region is weighted, depending on the perceptual importance of the region to arrive at Perceptual Structural Similarity Metric (PSSIM) indicative of the image quality.

Keywords: Objective image quality, HVS, structural distortion, perceptually important regions.

1 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 [1-7] 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 also not good because the residual image is not uncorrelated additive noise,it also contains components of the original image. A detailed discussion on MSE is given by Girod [8].

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 [9] that may influence the human visual perception on image quality. Although

A. Ghosh, R.K. De, and S.K. Pal (Eds.): PReMI 2007, LNCS 4815, pp. 87–94, 2007.

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 [10] which need to be validated. These methods suffer from the problems like natural image complexity problem, Minkowski error pooling problem, and cognitive interaction problem [10].

Structural similarity based methods [11, 12] 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[13, 14, 15, 16] 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 perceptually important regions based on the characteristics of intermediate and higher visual processes. 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 explains the perceptual importance map. Section 4 describes the computation of proposed quality index. Experimental results follow in Section 5. Finally, in Section 6, the conclusions of the paper are presented.

2 Structural SIMIlarity(SSIM)

Based on the assumption that the HVS is highly adapted to extract structural information from the viewing field, a new philosophy for image quality measurement*SSIM* was proposed by Wang et al [12]. Let $x = \{x_i | i = 1, 2, 3, ...N\}$and $y = \{y_i | i = 1, 2, 3, ...N\}$ be two discrete non-negative signals been aligned with each other and let $\overline{x}$,σ_x^2,and σ_{xy} be the mean of x,variance of x,and the covariance of x and y respectively. $\overline{x}$,σ_x^2 are the estimates of luminance and contrast of x and σ_{xy} measures the tendency of x and y to vary together, which is an indication of structural similarity.$SSIM$ index is given in equation(1) where C_1, C_2 and C_3 are small constants introduced to avoid instability when the denominator is close to zero.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

3 Perceptual Importance Map

Visual attention is a process that locates features in an image that have high information content so that limited computational resources can be directed toward them. Cave [17] writes that attention ”only allows a small part of the incoming sensory information to reach short-term memory and visual awareness, allowing us to break down the problem of scene understanding into rapid series of computationally less demanding, localized visual analysis problems”.

Psycho visual studies reveal that human eye is very sensitive to the edge and contour information of the image. Texture is one of the important characteristics used in identifying objects or regions of interest in an image. Texture contains important information about the structural arrangement of surfaces. The HVS converts luminance into contrast at an early stage of processing. Regions which have a high contrast with their surrounds attract our attention and are likely to be of greater visual importance [18, 19]. The following explains the computation of perceptual importance map based on these three parameters.

1. Edges per unit area E is determined by detecting edges in an image using Canny extension of the Sobel operator[20] and then congregating the edges detected within an 8x8 block [21].The value of E is normalized to the range [0 1]. A block without edges will have a value of 0.
2. The texture content$N$, in a local region is quantified by computing the entropy [22]. From basic information theory, entropy is defined as in equation (2), where z_i is a random variable indicating intensity, $p(z_i)$ is the histogram of the intensity levels in a region,L is the number of possible intensity levels (0 to 255 for gray scale images).

$$N = -\sum_{i=0}^{L-1} p(z_i) \lg_2 p(z_i) \tag{2}$$

   The value of N is normalized to fit in the range [0 1]. Image blocks with rich texture will have a value of 1.
3. Michaelson contrast C [23]is most useful in identifying high contrast regions, generally considered to be an important choice feature for human vision. Michaelson contrast is calculated as in equation (3) where l_m is the mean luminance within an 8 x 8 block and L_M is the overall mean luminance of the image. C is scaled to the range [0 1], 1 indicating highest contrast block and 0 indicating lowest contrast block.

$$C = \|(l_m - L_M)/(l_m + L_M)\| \tag{3}$$

For each block of the original image, the importance value is calculated as the normalized value of sum of the squares of the respective parameters, which forms the perceptual importance map P of the image.

4 Perceptual Structural Similarity

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),$1 \leq i \leq m = \lfloor H/8 \rfloor$, $1 \leq j \leq n = \lfloor W/8 \rfloor$, where H and W represent the height and width of the image respectively.

Secondly, the perceptual importance map P specified in section 4 is obtained for the original image as shown below. p_{ij} represents perceptual importance of each block with coordinates (i, j) as defined earlier. We define Perceptual Structural Similarity index $PSSIM$ as the weighted average of the structural similarity indices s_{ij} with coordinates (i, j), where each $s_{i,j}$ is weighted with the corresponding perceptual importance $p_{i,j}$. Eqaution(4) gives the expression for $PSSIM$.

$$S{=}\begin{pmatrix} s_{11} & s_{12} & . & . & s_{1n} \\ s_{21} & s_{22} & . & . & s_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ s_{m1} & s_{m2} & . & . & s_{mn} \end{pmatrix} \quad P{=}\begin{pmatrix} p_{11} & p_{12} & . & . & p_{1n} \\ p_{21} & p_{22} & . & . & p_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ p_{m1} & p_{m2} & . & . & p_{mn} \end{pmatrix}$$

$$PSSIM = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}[S][P]}{\sum_{i=1}^{m}\sum_{j=1}^{n}[P]} \tag{4}$$

5 Experimental Results

The proposed quality index was tested using LIVE image database [24]. 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 [27]. We performed non-linear mapping between the objective and subjective scores, using 4-parameter logistic function of the form shown in Equation (5).

$$y = a/(1.0 + e^{-(x-b)/c}) + d \tag{5}$$

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 $SSIM$. Table 1 shows the evaluation results for the models being compared with that of the $PSSIM$ for different types of distortions.For each of the objective evaluation criteria, $PSSIM$ outperforms the other models being compared across different distortion types. Fig. 1 shows the scatter plots of DMOS versus $PSSIM$ for different kinds of distortions.

Table 1. Performance comparison of image quality assessment models on LIVE image database [18]. CC: non-linear regression correlation coefficient; ROCC: Spearman rank-order correlation coefficient; MAE: mean absolute error; RMS: root mean square error; OR: outlier ratio. (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
SSIM	0.899	0.894	5.687	7.077	2.366
PSSIM	0.941	0.935	4.426	5.442	2.958

(a)

Model	CC	ROCC	MAE	RMS	OR
PSNR	0.842	0.828	6.636	8.622	6.285
SSIM	0.891	0.863	5.386	7.236	5.714
PSSIM	0.930	0.893	4.262	5.871	6.285

(b)

Model	CC	ROCC	MAE	RMS	OR
PSNR	0.922	0.938	4.524	6.165	5.555
SSIM	0.94	0.914	4.475	5.459	2.777
PSSIM	0.964	0.952	3.514	4.247	3.472

(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
PSSIM	0.969	0.964	3.240	3.871	2.758

(d)

Model	CC	ROCC	MAE	RMS	OR
PSNR	0.857	0.859	6.383	8.476	6.896
SSIM	0.956	0.945	3.806	4.799	5.517
PSSIM	0.967	0.959	3.328	4.189	4.827

(e)

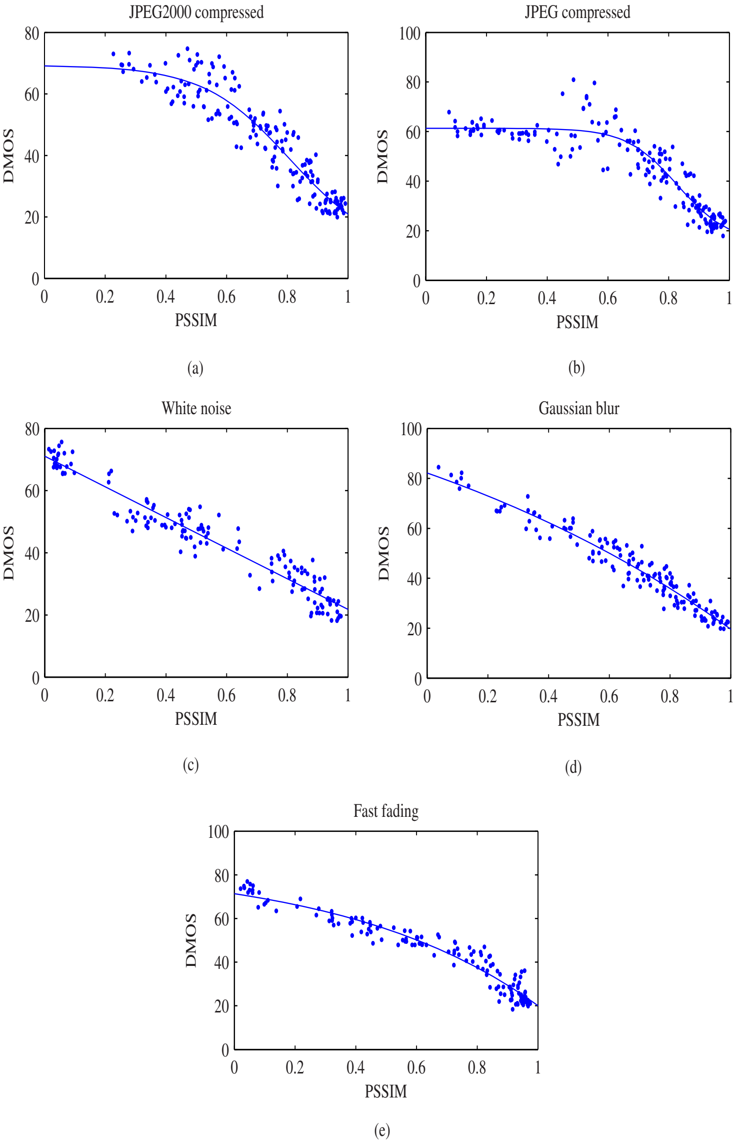

Fig. 1. Scatter plots for DMOS versus model prediction for (a) JPEG2000 (b) JPEG (c) White noise (d) Gaussian blur (e) Fast fading distorted images

6 Conclusions

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 attention model. The results prove the fact that human eye is sensitive to important image features like edges, texture, and contrast. The results also justify the visual attention model built on these three parameters. 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.

References

1. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Transactions on Communications 43(12), 2959–2965 (1995)
2. Karunasekera, S.A., Kingsbury, N.G.: A distortion measure for blocking artifacts in images based on human visual sensitivity. IEEE Transactions on Image Processing 4(6), 713–724 (1995)
3. Mill, N.B.: A visual model weighted cosine transform for image compression and quality assessment. IEEE Transactions on Communications 33(6), 551–557 (1985)
4. Saghri, J.A.: Image quality measure based on a human visual system model. Optical Engineering 28(7), 813–818 (1989)
5. Daly, S.: The visible differences predictor: an algorithm for the assessment of image fidelity. In: Watson, A.B. (ed.) Digital Images and Human Vision, Ch. 14, pp. 179–206. MIT press, Cambridge (1993)
6. Lubin, J.: A visual discrimination model for imaging system design and evaluation. In: Peli, E. (ed.) Vision Models for Target Detection and Recognition, Ch.10, pp. 245–283. World Scientific Publishing (1995)
7. Watson, A.B.: DCT quantization matrices visually optimize for individual images. In: Human Vision, Visual Processing and Digital Display IV, Proc. SPIE, vol. 1913, pp. 202–216 (1993)
8. Girod, B.: What's wrong with mean-squared error. In: Watson, A.B. (ed.) Digital Images and Human Vision, pp. 207–220. MIT Press, Cambridge (1993)
9. Wandell, B.A.: Foundations of Vision, Sinauer Associates, Inc. (1995)
10. Wang, Z., Sheikh, H.R., Bovik, A.C.: Objective video quality assessment. In: Furht, B., Marques, O. (eds.) The Handbook of Video Databases: Design and Applications, pp. 1041–1078. CRC press (September 2003)
11. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement, Signal Processing: Image Communication. special issue on objective video quality metrics 19 (January 2004)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simocelli, E.P.: Image quality assessment: From error measurement to structural similarity. IEEE Trans. Image Processing 13(4), 600–612 (2004)
13. Geri, G.A., Zeevi, Y.Y.: Visual assessment of variable-resolution imagery. Journal of the Optical Society of America 12(10), 2367–2375 (1995)
14. Kortum, P., Geisler, W.: Implementation of a foveated image coding system for image bandwidth reduction. In: SPIE - Human Vision and Electronic Imaging, vol. 2657, pp. 350–360 (February 1996)
15. Stelmach, L.B., Tam, W.J., Hearty, P.J.: Static and dynamic spatial resolution in image coding: An investigation of eye movements. In: Proceedings of the SPIE, San Jose, vol. 1453, pp. 147–152 (1991)
16. Yarbus, A.L.: Eye Movements and Vision Press, New York (1967)

17. Cave, R.: The feature Gate model of visual selection. Psychological research 62, 182–194 (1999)
18. Findlay, J.: The visual stimulus for saccadic eye movement in human observers. Perception 9, 7–21 (1980)
19. Senders, J.: Distribution of attention in static and dynamic scenes. In: Proceedings SPIE, San Jose, vol. 3016, pp. 186–194 (February 1997)
20. Canny, J.: A Computational Approach to Edge Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
21. Richards, W., Kaufman, L.: Centre-of-Gravity Tendencies for Fixations and Flow Patterns. Perception and Psychology 5, 81–84 (1969)
22. Gonzalez, Woods.: Digital Image Processing. Prentice Hall, Englewood Cliffs (2002)
23. Mannan, S.K., Ruddock, K.H., Wooding, D.S.: The Relationship between the Locations of Spatial Features and Those of Fixations Made during Visual Examination of Briefly Presented Images. Spatial Vision 10(3), 165–188 (1996)
24. Sheikh, H.R., Bovik, A.C., Cormack, L., Wang, Z.: LIVE Image Quality Assessment Database (2004), http://live.ece.utexas.edu/research/quality
25. Corriveau, P., et al.: Video quality experts group: Current results and future directions. In: presented at the SPIE Visual Communication and Image Processing, vol. 4067 (June 2000)
26. Van Dijk, A.M., Martens, J.B., Watson, A.B.: Quality assessment of coded images using numerical category scaling. In: Proc. SPIE, vol. 2451, pp. 90–101 (March 1995)
27. VQEG: Final report from the video quality experts group on the validation of objective models of video quality assessment (March 2004), http://www.vqeg.org/