Asian Journal of Information Technology 4(11): 996-1000, 2005

New Perceptual Image Quality Assessment Metric

[1]Amine Samet, [1]M.A.B. Ayed, [1]Nouri Masmoudi and [2]Lazhar Khriji
[1]Laboratory of Electronics and Information Technology,
National Engineering School of Sfax, BP W 3038 Sfax, Tunisia
[2]Department of Electrical and Computer Engineering,
Sultan Qaboos University, P.O. Box 33, C.P.123, Muscat, Oman

Abstract: The goal of this study is to determine a human visible difference between two images that may be used as an excellent image quality evaluator. This study describes a new perceptual image quality metric: the New weighted Mean Square Error (NwMSE) is easy to calculate and applicable to various image processing applications. It produces clear distortion maps and quantitative error measures informed by considerations of the Human Visual System (HVS). Many experiments on various image distortion types validate that NwMSE is compliant with subjective evaluations and performs significantly better than the widely used distortion metrics the Mean Square Error (MSE), the Structural SIMilarity (SSIM) index, the Image Quality Index (IQI) and the weighted Mean Square Error (wMSE). We show that the NwMSE correlates with MOS results and with human perception of image quality.

Key words: Image quality assessment, Human Visual System (HVS), image activity measure, spatial frequency activity

INTRODUCTION

Image quality assessment is essential for most image processing applications[1]. Such metric is used; i) to monitor image quality for quality control systems, ii) to benchmark image processing systems and algorithms and iii) to be embedded into an image processing system, to optimize the algorithms and the parameter settings.

There are two ways to evaluate the image quality: objective measures and subjective measures. In the literature, the most image quality assessments are error-based methods[2]. Thus, these mathematical measures are performed by pixel based difference metrics like MSE, Root MSE (RMSE), Mean Absolute Errors (MAE), Signal to Noise Ratio (SNR) and Peak Signal to Noise Ratio (PSNR). However, these metrics do not correlate with the perceived image quality[2,3]. Indeed, they neglect the properties of the HVS and thus cannot be a reliable predictor of the perceived visual quality.

In the last decade, several image and video quality metrics have been proposed, which incorporate perceptual quality measures by considering the HVS characteristics[1,4]. In this study, we propose a new perceptual metric that associates objective and subjective measures. The New weighted Mean Square Error (NwMSE) is developed based on weighting errors according, simultaneously, to an image activity measure and to the HVS spatial frequency response.

The goal of this metric is to determine a human visible difference between two images that may be used as an excellent image quality evaluator. It must be applicable to various image processing applications; and provide meaningful comparison across different types of image distortions. Compared with others metrics, the NwMSE provides a significant correlation with the subjective assessment.

THE NEW WEIGHTED MEAN SQUARE ERROR

In multimedia application, an image quality assessment that is based on a HVS model seems to be more appropriate since the human eyes are the ultimate receivers in most image processing environment's. The system performance is to be measured in terms of perceived quality. The Mean Opinion Score (MOS) has been used for many years as a subjective quality measure. Subjective assessment tests are widely used to evaluate the image quality, but they are difficult and lengthy and results obtained might be very depending on the test condition and humans viewers[1]. Therefore, they require longer time and the results are not always repeatable.

The sensitivity of the HVS to the errors may be different for different types of errors and may also vary with visual context. This difference may not be captured

Corresponding Author: Dr. Lazhar Khriji, Department of Electrical and Computer Engineering, Sultan Qaboos University, P.O.Box 33, C.P.123, Muscat, Oman

Table 1: Quantitative results of corrupted images: Pepper, Cemetry, House and Lena

Image		MSE	MOS	IQI	MSSIM	wMSE	NwMSE
Pepper	Fig. 1a	108.7704	2.00	0.8448	0.8564	121.7886	153.9184
	Fig. 1b	108.6308	3.86	0.6192	0.6688	121.9602	151.1266
	Fig. 1c	108.4618	4.20	0.7802	0.8406	121.7511	151.6593
Cemetry	Fig. 1d	110.9195	2.86	0.8713	0.8773	119.9114	124.6891
	Fig. 1e	110.7542	3.93	0.7492	0.7729	119.1588	120.5316
	Fig. 1f	110.5631	4.13	0.8865	0.9062	118.9745	120.6167
House	Fig. 1g	112.6284	2.86	0.7860	0.8013	133.4762	338.5193
	Fig. 1h	112.5384	2.00	0.5454	0.6031	131.0391	280.8035
	Fig. 1i	112.5570	4.33	0.7317	0.8074	131.0416	279.9724
Lena	Fig. 1j	104.4212	2.40	0.8282	0.8421	118.5718	178.5236
	Fig. 1k	104.6584	2.86	0.6427	0.6901	118.1819	167.4525
	Fig. 1l	104.7946	4.20	0.7910	0.8484	118.1822	167.2048

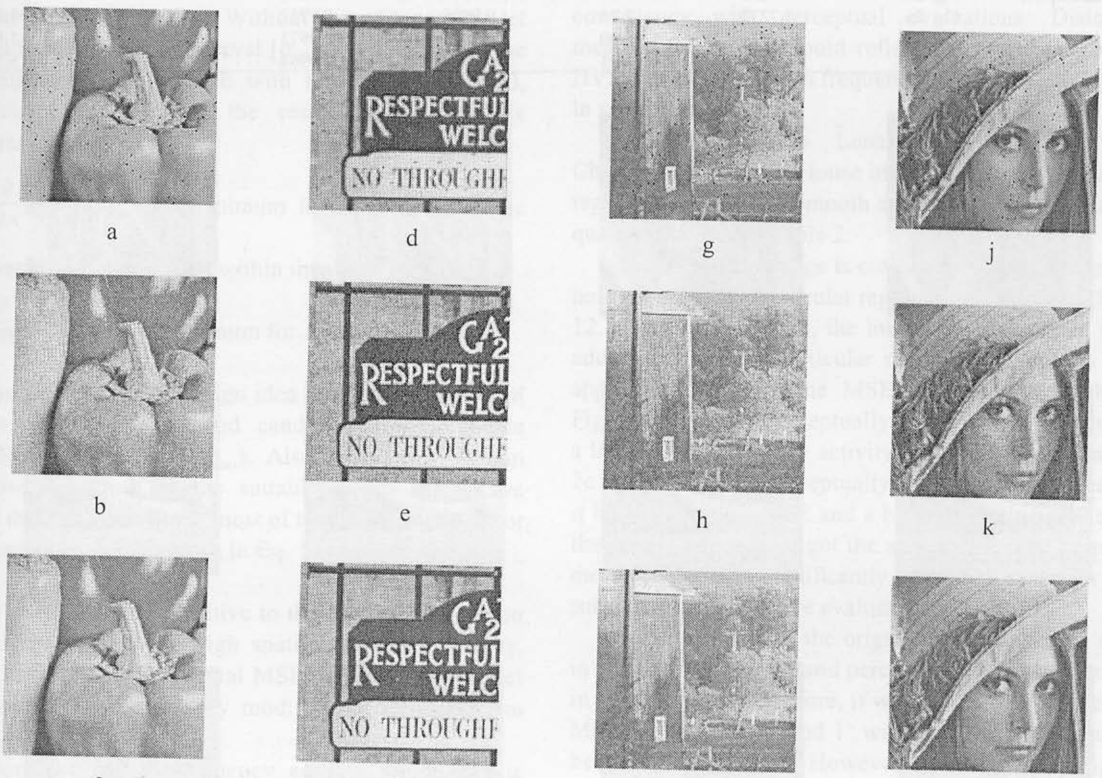

Fig. 1: Evaluation of noisy images with different types of distortions, Pepper, Cemetry, House and Lena. Images are cropped to 128x128 for visibility. (a, d, g and j) Impulsive Salt and Pepper noise; (b, e, h and k) Gaussian noise; (c, f, i and l) White Uniform noise

adequately by the MSE. In turn, the HVS takes into account the neighborhood pixels and consequently, any loss of information in the pixel neighborhoods could be a good measure of image fidelity. Present algorithm is based on the assumption that the neighborhood pixels, within the same block, are of good continuity. Such continuity is aimed to recover the discontinuity across pixels neighborhood. The NwMSE is based on that strategy. It takes into account the pixels neighborhood information in an effective way.

Let $X = \{x_{i,j} \mid i = 1, \ldots, M;\ j = 1, \ldots, N\}$ and $Y = \{y_{i,j} \mid i = 1, \ldots, M;\ j = 1, \ldots, N\}$ be the original image and the test image, respectively. The NwMSE is given as:

$$NwMSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\left(x_{i,j} - y_{i,j}\right)^2 .\left(1 + A.e^{-B\sigma_{i,j}}\right)\right) \quad (1)$$

Where, $\sigma^2_{i,j}$ is the local test image variance of a block centered on a given pixel $y_{i,j}$ with size kxl pixels having an

Table 2: Quantitative results of corrupted images over particular regions: Lena, Boat, Couple, Man, Churchandcapitol and House

	Image	MSE	MOS	IQI	MSSIM	wMSE	NwMSE
Lena	Fig. 2a	12.3440	3.26	0.9723	0.9743	15.4963	35.4849
	Fig. 2b	12.3202	3.20	0.9761	0.9792	15.3839	33.0899
	Fig. 2c	12.3103	5.00	0.9850	0.9855	13.2190	12.3340
	Fig. 2d	12.3190	5.00	0.9900	0.9905	13.2297	12.3482
Boat	Fig. 2e	10.1771	3.26	0.9884	0.9888	11.8981	14.0161
	Fig. 2f	10.1444	2.60	0.9899	0.9907	11.8505	13.8843
	Fig. 2g	10.1009	5.00	0.9946	0.9947	10.5919	10.5645
	Fig. 2h	10.1493	5.00	0.9967	0.9969	10.6400	10.6138
Man	Fig. 2i	8.4445	3.33	0.9897	0.9900	10.0833	318.9722
	Fig. 2j	8.4316	2.40	0.9908	0.9916	10.0678	318.4466
	Fig. 2k	8.4048	5.00	0.9952	0.9953	8.8500	32.9899
	Fig. 2l	8.4509	4.93	0.9973	0.9974	9.0028	38.3432
Churchandcapitol	Fig. 2m	5.7233	3.40	0.9906	0.9912	10.1000	230.3059
	Fig. 2n	5.7233	2.46	0.9906	0.9912	10.1000	230.3059
	Fig. 2o	5.7771	5.00	0.9972	0.9973	5.9871	5.8146
	Fig. 2p	5.7416	4.93	0.9984	0.9984	5.9504	5.7852

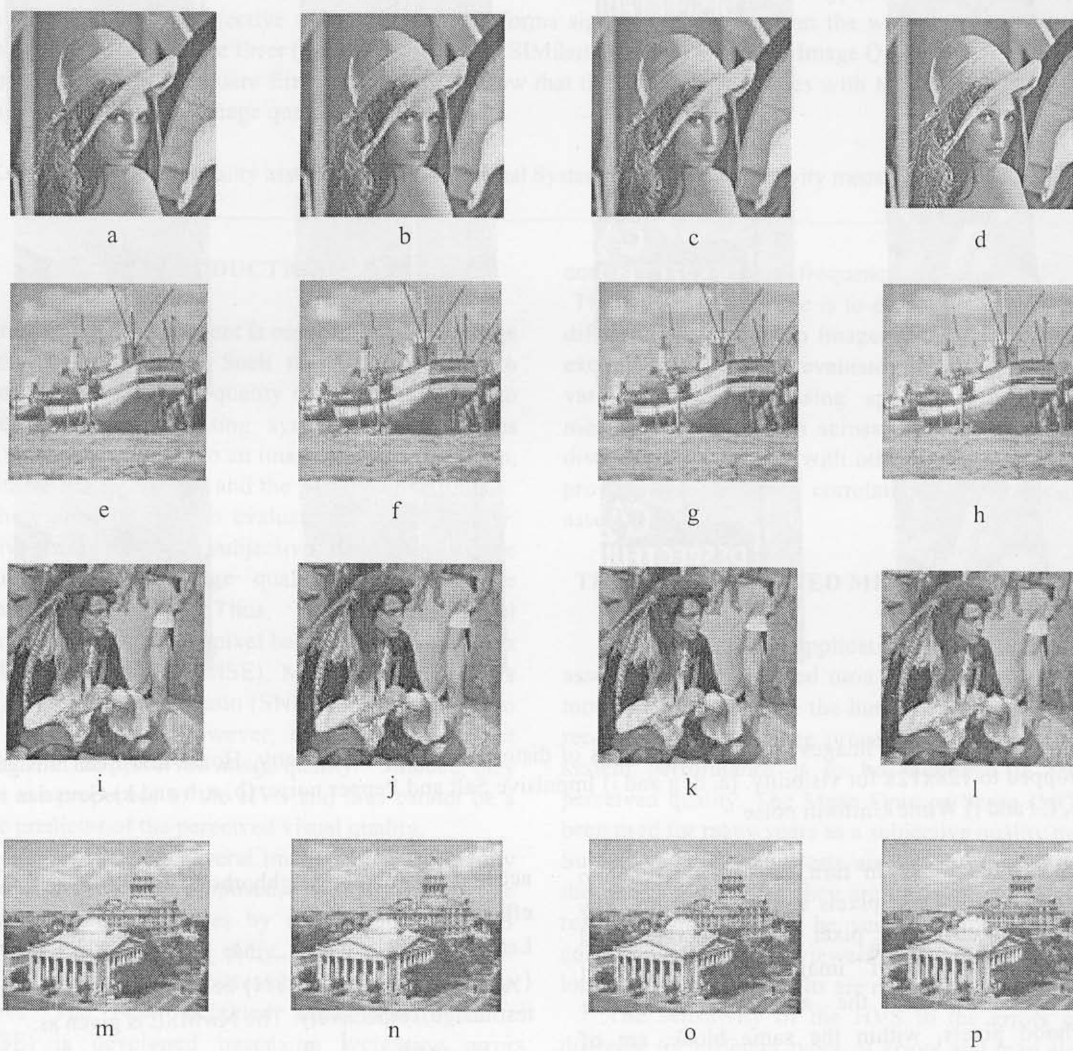

Fig. 2: Evaluation of distorted images over particular regions, Lena, Boat, Man and Churchandcapitol. Over smooth region: (a, e, I and m) Gaussian noise and (b, f, j and n) White Uniform noise. Over texture region: (c, g, k and o) Gaussian noise and (d, h, l, p) White Uniform noise

average value equals μ. A, B are constants to be determined to match this new metric with perceptual and objective tests.

In addition, experimental studies were performed to determine the spatial distribution of the local variance for several test images. In the most of the cases the distribution has an exponential shape, which confirms the announced NwMSE.

Moreover, we take into account the constraint that all parameters should be connected to the local characteristics of the image such as the energy, the mean and the standard deviation. Without loss of generality, let A and B belong to the interval $[\sigma_{min}, \sigma_{max}]$. By taking the derivatives of the NwMSE with respect to A and B, respectively and setting the result to the possible minimum.

- $\frac{\partial}{\partial A}(\text{NwMSE})$ is minimum for $B = \sigma_{min}$, since the variables are bounded within the range $[\sigma_{min}\ \sigma_{max}]$.
- $\frac{\partial}{\partial B}(\text{NwMSE})$ is minimum for $A=\sigma_{min}$.

So, as long as the design idea is to rely on the local image characteristics, good candidates for the couple (A, B) would be $(\sigma_{min}, \sigma_{max})$. Also, by considering an adaptive algorithm for the suitable values of (k,l), we found that (k = l =9) in the most of the time. The behavior of present new metric given in Eq. 1 is explained by:

- The HVS is less sensitive to the contrast in a region of the image with high spatial frequency activity. Consequently, the partial MSE for a particular pixel will not be completely modified. Indeed, the term $\left(1 + A.e^{-B\sigma_{i,j}}\right)$ is close to 1.
- For low spatial frequency activity region, $\sigma_{i,j}$ is relatively small. Consequently, the term $\left(1 + A.e^{-B\sigma_{i,j}}\right)$ is greater than 1. Therefore, the partial MSE will be amplified and contributed to the increase of the NwMSE value.

RESULTS AND DISCUSSION

To evaluate the performance of the proposed metric the LIVE database[5]. Six image quality assessments are being compared, including MSE, MOS, IQI, Mean SSIM (MSSIM)[6], wMSE and the proposed NwMSE. The MOS of each image is obtained by averaging subjective scores given by a group of 15 human viewers. The viewers were asked to score an image. The ranking was done on an integer scale from 1 to 5, i.e. 1) bad, 2) poor, 3) fair, 4) good and 5) excellent. Corrupted versions with different scenarios of pepper, cemetry, house and lena images were generated (Fig. 1). Each set of images has, nearly, identical MSE. Objective image quality measures are calculated for all images and presented in Table 1. From Fig. 1c and f, it is alarming to note that a higher MSE does not always mean worse image quality and vice versa. The NwMSE values of the images in Fig. 1 g and i are 338.5193, 280.8035 and 279.9724, respectively, which appear to have better consistency with perceived image quality than MSE. The images with similar MSE have significantly different visual quality and NwMSE delivers better consistency with perceptual evaluations. Distortion measure for image should reflect the limitations of the HVS. For example, high frequency noise is not noticeable in smooth images.

Figure 2 shows Lena, Boat, Couple, Man, Churchandcapitol and House images corrupted over some regions of texture and smooth areas. A quantitative Image quality is shown in Table 2.

In Fig. 2a, the image is created by adding a Gaussian noise to a smooth particular region, producing an MSE of 12.3440 and in Fig. 2c, the image has a Gaussian noise added to a texture particular region. Both images have approximately the same MSE. It is clear to note that in Fig. 2a the noise is perceptually obvious since it occurs in a low spatial frequency activity region. However, in Fig. 2c the noise is not perceptually identified. Consequently, it has a higher NwMSE and a higher IQI despite the fact that both images have got the same MSE. The proposed metric outperforms significantly MSE in all cases, for both subjective and objective evaluations.

With reference to the original Man image, the noise in Fig. 2j is more vivid and perceptually obvious than that in Fig. 2k and l. Therefore, if we have to find a match for Man image, Fig. 2k and l will represent, perceptually, better the original one. However, using the MSE, Fig. 2 j will result in the smallest MSE and will be chosen as the best match for the original image.

The images in Fig. 2i and k present a higher difference in terms of NwMSE (318.9722 for the first and 32.9899 for the second). But, these images do not show any significant difference in terms of MSSIM (0.99 for the first and 0.9953 for the second). Therefore, the NwMSE is more significant and competent then MSSIM(Table 2).

Table 1 and 2 show the image quality assessments for corrupted test images over the whole images and over particular regions, respectively. An image is ranked as acceptable if it maintains satisfactory diagnostic value i.e. MOS=4. It is seen that, for the most of the images, the NwMSE has similar results with MOS prediction method. Thus, this new metric is a better indicator of perceived image quality than its counterpart.

CONCLUSIONS

A new perceptual metric for image quality evaluation is introduced. The NwMSE is derived based on statistical studies in addition to the use of objective and subjective characteristics in an effective way. Simulation results and subjective evaluation demonstrate that the proposed metric outperform all other metrics used in this study. Furthermore, the NwMSE is closer to perception than the MSE and IQI. Therefore, in some situations the newly objective metric may successfully replace the use of subjective tests. Future studies will consider the use of NwMSE as an image quality assessment for color images and for real time applications.

REFERENCES

1. Pappas, T.N. and R.J. Safranek, 2000. Perceptual Criteria for Image Quality Evaluation. Handbook of Image and Video Proc. (Bovik, A., Ed.), Academic Press.
2. Wang, Z. and A.C. Bovik, 2002. A Universal Image Quality Index. IEEE Signal Processing Letters, 9: 81-84.
3. Samet, A., M.A. Ben Ayed, M. Loulou and N. Masmoudi, 2004. Perceptual Evaluation of JPEG2000. J. Eur. Trans. Telecomm. (ETT)., 15: 135-143.
4. Grgic, S., M. Grgic and M. Mark, 2004. Reliability of objective picture quality measures. J. Elect. Eng., 55: 3-10.
5. Sheikh, H.R., Z. Wang, A.C. Bovik and L.K. Cormack, Image and video quality assessment research at LIVE. http://live.ece.utexas.edu/research/.
6. Wang, Z., A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, 2004. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Processing, Vol: 13.