# A New Metric for Quality Assessment of Digital Images Based on Weighted-Mean Square Error[1]

Kaiwen Zhang, Shuozhong Wang, and Xinpen Zhang
*School of Communication and Information Engineering, Shanghai University*

**E-mail:** *ztszkwzr@sh163.net*   **or** *shuowang@yc.shu.edu.cn*

Phone: 86-21-5633 1435 or 86-21-56778156

Fax: 86-21-5633 1964

**Abstract:** In this paper, an image quality measure, termed pixel-based correlation weighted-mean square error (WMSE), is presented. The proposed distortion measure depends not only on the mean square error in the distorted image, but also on correlation between pixels in a predefined neighborhood. Experimental results are given to show the advantage of the described method.
**Keywords:** image quality metric, mean square error (MSE)

---

## 1. Introduction

Image quality assessment is an important but difficult issue in image processing applications such as compression coding and digital watermarking. For a long time, mean square error (MSE) and peak signal-to-noise ratio (PSNR) are widely used to measure the degree of image distortion because they can represent the overall gray-value error contained in the entire image, and are mathematically tractable as well. In many applications, it is usually straightforward to design systems that minimize MSE or PSNR.

MSE works satisfactorily when the distortion is mainly caused by contamination of additive noise. However the problem inherent in MSE and PSNR is that they do not take into account the viewing conditions and visual sensitivity with respect to image contents. With MSE or PSNR, only gray-value differences between corresponding pixels of the original and the distorted version are considered. Pixels are treated as being independent of their neighbors. Moreover, all pixels in an image are assumed to be equally important. This, of course, is far from being true. As a matter of fact, pixels at different positions in an image can have very different effects on the human visual system (HVS).

In order to make more comprehensive image quality evaluation, and to meet different requirements in a wide range of applications, various quality measures have been proposed. Wang and Bovik[8,9] showed some examples in which MSE failed to provide correct results in evaluating quality of certain distorted images. They proposed a new metric, Q, termed the universal image quality index, for evaluation of image distortion, and showed that some MSE assessments were in apparent contradiction with subjective judgments, and this could be corrected by using Q. There are many other measures that, according to the techniques used, can be classified into pixel difference-based[1-4], correlation-based[5], edge-based, spectral-based, context-based, and HVS-based[6,7], etc.

Although a great deal of efforts have been made in the past, it is still necessary to develop new techniques for objective evaluation of image quality that can easily be implemented and are truly adapted to the HVS. In this paper, a weighted-MSE metric that takes into account properties in the neighborhood of each pixel is described. Experiments show that the proposed quality metric can overcome some difficulties encountered when using some existing measures, and provide a useful means to evaluate degrees of image distortion in applications such as image restoration and digital watermarking.

## 2. Definition of Weighted MSE

As mentioned in the previous section, although MSE has its merits and is widely accepted in image processing research, it only measures gray-level difference between pixels of the ideal and the distorted images without considering correlation between the neighboring pixels. Distorted images with equal MSE or PSNR may have significantly different visual quality.

A human observer always views an image as an entirety, rather than just a collection of isolated pixels, therefore correlation between neighboring pixels plays a role in the subjective judgment of image quality. As a result, different types of signal processing procedures and noise interference can cause different perceptual effects, and the visual judgment is sometimes heavily dependent upon the degree of distortion. For example, salt-pepper noise is generally more annoying than Gaussian noise in cases of slight degradation. If the degradation becomes severe, however, the situation may be reversed. In view of this, a new quality metric that incorporates pixel correlation is introduced as follows.

Let $I$ and $\hat{I}$ be the original and the distorted images respectively, sized $M \times N$. A weighted-MSE is defined as

$$\text{WMSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \alpha_{ij} \left| I(i,j) - \hat{I}(i,j) \right|^2 \tag{1}$$

where $\alpha_{i,j}$ is a weight representing correlation of pixels within a rectangular window $W$ centered at $(i, j)$:

$$\alpha_{i,j} = 1 - \left| \frac{\sum_{k \in W}\left[I'_k(i,j) - \overline{I'_k}\right]\left[\hat{I}'_k(i,j) - \overline{\hat{I}'_k}\right]}{\sqrt{\frac{1}{N_w - 1}\left[I'_k(i,j) - \overline{I'_k}\right]^2}\sqrt{\frac{1}{N_w - 1}\left[\hat{I}'_k(i,j) - \overline{\hat{I}'_k}\right]^2}} \right| \tag{2}$$

where $N_w$ is the number of pixels within the window. It is clear that MSE is a special case of the described WMSE since WMSE approaches MSE when $\alpha_{i,j}$ tends to one. In calculating $\alpha_{i,j}$, the gray values are multiplied with a properly chosen taper function $\beta$ so that pixels within the window are given different importance:

$$I'_k(i, j) = \beta(i, j)\, I_k(i, j) \tag{3}$$

Clearly, the value of $\beta$ should decrease with radius.

As WMSE incorporates correlation between adjacent pixels in its definition, it is likely to provide quality assessment close to human perceptive judgment. This has been verified in the experiments as will be shown in the following.

## 3. Characteristics of WMSE

A good objective quality measure should truly reflect the degree of image distortion due, for instance, to blurring, noise contamination, compression coding, inadequacy of the sensor specification, and so forth. It should possess the following attributes:

1) Conformity with subjective judgment: The metric should provide a quantitative evaluation of image distortion that, in general, conforms to subjective assessment. In other words, a visually superior image should have a quality measure representing a lower distortion. A reasonable image quality metric should also be monotonic with respect to the degree of distortion.

2) Mathematical tractability: A practically useful quality metric should be easy to calculate without taking excessive computational resources.

3) Consistency over different images: This means that an ideal objective quality metric should be able to provide correct assessment for all types of images, and not to fail for any subset of images.
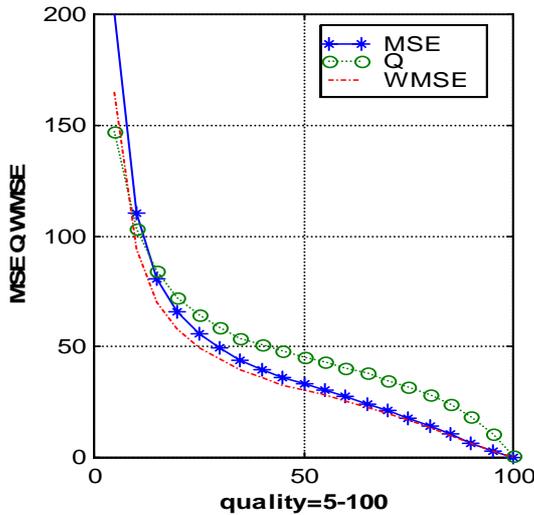


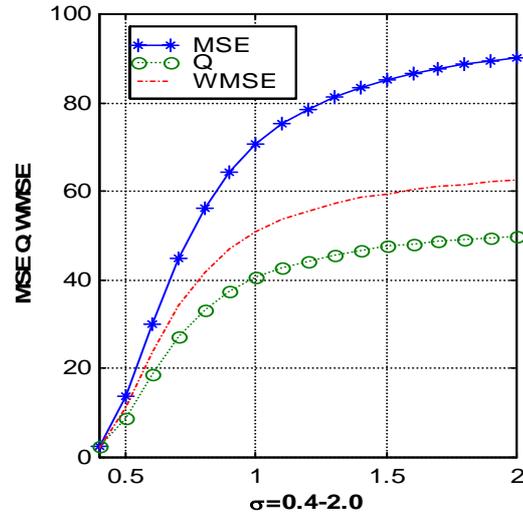Fig.1 Quality metrics vs. JPEG quality factor



Fig.2 Quality metrics vs. degree of blurring

In order to study the properties of WMSE, a test image Lena was JPEG compressed at various quality factors, and blurred with a 3×3 Gaussian kernel characterized with a standard deviation, $\sigma$. Figs.1 and 2 show curves of WMSE together with those of MSE and $P=300(1-Q)$ where Q was introduced in [8] against JPEG quality factor and $\sigma$, respectively. It is observed from these figures that WMSE satisfies conformity with visual judgment. Similar results were obtained by using other test images.

From the definition, one can see that WMSE is mathematically tractable although it requires more computation than the simple MSE. Since WMSE is a modification of MSE, and MSE can reliably measure image distortion (or error) in a monotonous order, WMSE should also be monotonous with different degrees of image distortion.

Also, since WMSE mainly depends on the difference between pixel values of the original and distorted images and, to a much less extent, on the pixel values of the original image, it is expected that, unlike $Q$ that depends on properties of particular images, WMSE can give consistent evaluation to different types of image.

## 4. Experimental Results and Discussion

In the experiments, a two dimensional Gaussian function ($\sigma = 0.1717$) is used as the taper function $\beta$. A number of images including Lena, Baboon and Couple were tested. These test images were degraded in a variety of ways such as impulsive salt-pepper noise interference, additive Gaussian noise, blurring, and JPEG compression. Fig.3 shows some of the test samples.
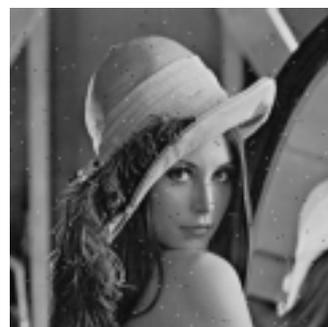
It is confirmed from Fig.3 that different signal manipulations and contamination by different noise types have different effects upon the HVS. At low noise levels, salt-pepper noise is considered more harmful than Gaussian noise, whereas at high noise levels, the reverse is true. The block effects due to JPEG coding is usually less annoying than blurring processing, etc.
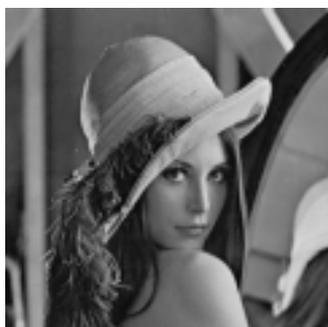


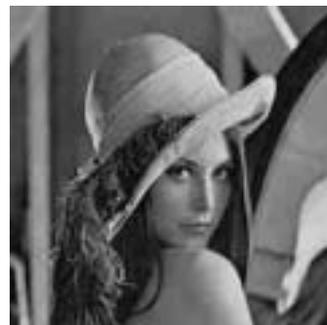(a) the original image      (b) MSE=80 Gaussian noise      (c) MSE=80 salt-pepper noise

(d) MSE=225 Gaussian noise      (e) MSE=225 salt-pepper noise      (f) MSE=81 JPEG

(g) MSE=82 blurring  (h) Q=0.9106 Gaussian noise  (i) Q=0.911 salt-pepper noise

Fig. 3

In order to make comparison between WMSE and the other two metrics, we produced two groups of degraded sample images using impulsive salt-pepper noise and additive Gaussian noise, referred to as S and G, respectively. MSE, $Q$, and WMSE of each pair of sample images were calculated, denoted $MSE(G_i)$, $MSE(S_i)$, $Q(G_i)$, $Q(S_i)$, $WMSE(G_i)$, and $WMSE(S_i)$, where $i$ refers to different sample images. The levels of salt-pepper and Gaussian noise were chosen such that each pair of sample images had identical MSE values with a tolerance of 1.

Define the following quantities:

$$M(i) = \frac{1}{2}[MSE(G_i) + MSE(S_i)] \tag{4}$$

$$\Delta_Q(i) = Q(G_i) - Q(S_i) \tag{5}$$

$$\Delta_{WMSE}(i) = WMSE(G_i) - WMSE(S_i) \tag{6}$$

In Fig.4, $\Delta_Q$ is plotted against $M$. Since all $\Delta_Q$ values are negative in a wide range of different degrees of distortion, the metric $Q$ always gives the judgment that Gaussian noise is less annoying. This, of course, does not conform to the experiences of most observers.
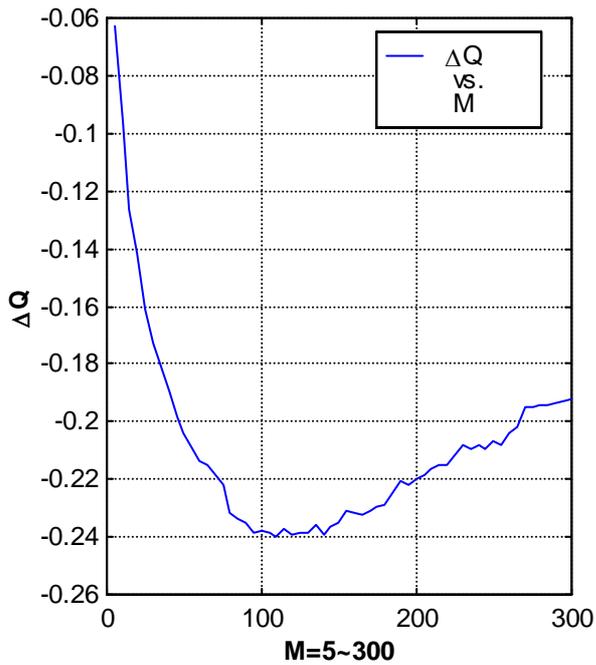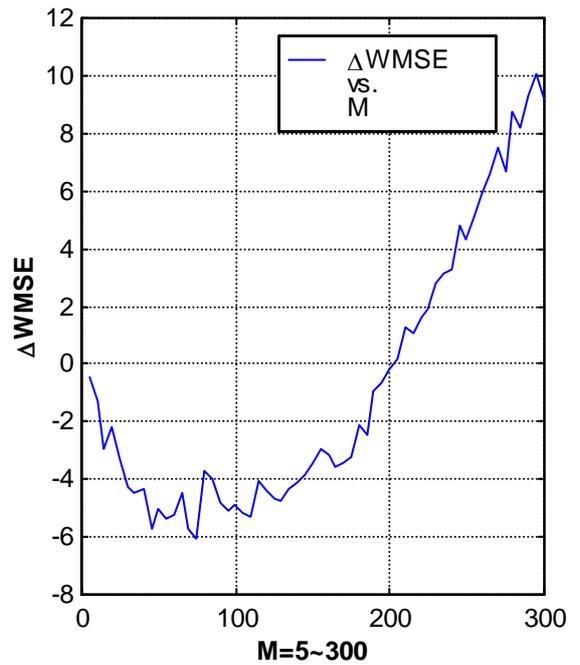


Fig.4 $\Delta_Q$ vs. $M$



Fig.5 $\Delta_{WMSE}$ vs. $M$

Fig.5 shows the relationship between $\Delta_{\text{WMSE}}$ and $M$, in which one can see that, at low noise levels, salt-pepper noise is worse than Gaussian noise, while at high noise level, Gaussian noise is worse. This indicates the advantage of WMSE over $Q$ and the simple MSE. The transition point, which is $M=200$ in the plot, depends on the choice of parameters such as the size of window and the shape of $\beta$, and the particular image tested.

Experiments on other types of degradation such as blurring and JPEG coding produced similar results.

## 5. Conclusions

A weighted-MSE quality metric for assessment of distortion in digital image has been proposed. It is not just based on gray-value differences of individual pixels like the simple metric MSE, but also on correlation between neighboring pixels. This leads to a better performance than MSE as verified in the experiments carried out on images degraded by impulsive salt-pepper noise and additive Gaussian noise.

Satisfactory performance can be obtained by choosing an appropriate window size and a proper function $\beta$. The optimal choice may depend on specific images.

However, WMSE is not meant to replace other image quality metrics, but to provide a useful alternative method. It is believed that, since the HVS is too complex to be matched with any existing objective metric, no single quantity at present can serve as a truly universal measure that can give accurate assessment to all images in whatever applications. In this area, further studies are needed.

## References:

[1]  H. de Ridder, "Minkowsky Metrics as a Combination Rule for Digital Image Coding Impairments," in *Human Vision, Visual Processing, and Digital Display III, Proc. SPIE*, 1666, 17-27, 1992.

[2]  International Commission of Illumination (CIE), Recommendations on Uniform Color Spaces, Color Difference Equations, Psychometric Color Terms, Publication CIE 15 1971.

[3]  V. V. Staravoitov, C. Köse, B. Sankur, "Generalized Distance Based matching of Nonbinary Images," International Conference on Image processing, Chicago, 1998.

[4]  V. Digesu, V. V. Staravoitov, "Distance-based Functions for Image Comparison," *Pattern Recognition Letters*, 20(2), 207-213, 1999.

[5]  A. M. Eskicioglu, P. S. Fisher, "Image Quality Measures and Their Performance," *IEEE Trans. Commun.*, 43(12), 2959-2965, 1995.

[6]  Marcus J. Nadenau, Stefan Winkler, David Alleysson and Murat Kunt, "Human Vision Models for Perceptually Optimized Image," http://......

[7]  T. Frese, C. A. Bouman, and J. P. Allebach, "A Methodology for Designing Image Similarity Metrics Based on Human Visual System Models," *Tech. Rep. TR-ECE 97-2*, Purdue University, West Lafayette.

[8]  Z. Wang and A. C. Bovik, "A Universal Image Quality Index," *IEEE Signal Processing Letters*, Vol. 9, No. 3, March 2002. http://anchovy.ece.utexas.edu/~zwang/papers/quality_2c.pdf

[9]  http://anchovy.Ece.Utexas.edu/~zwang/research/quality_index/