A UNIVERSAL IMAGE QUALITY INDEX and SSIM comparison

Submitted by:
CHAITANYA CHUKKA
1000583191
GUIDANCE: Dr K.R Rao
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>SN.O</th>
<th>TITLE</th>
<th>PAGE NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LIST OF ACRONYMS</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>ABSTRACT</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>INTRODUCTION</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>UNIVERSAL IMAGE QUALITY INDEX( UIQI)</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>STRUCTURAL SIMILARITY INDEX</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>IMAGES WITH DISTORTIONS</td>
<td>23</td>
</tr>
<tr>
<td>7</td>
<td>RESULTS</td>
<td>31</td>
</tr>
<tr>
<td>8</td>
<td>CONCLUSIONS</td>
<td>39</td>
</tr>
<tr>
<td>9</td>
<td>REFERENCES</td>
<td>40</td>
</tr>
</tbody>
</table>
## LIST OF ACRONYMS

1. HVS  
   **Human Visual System.**

2. MPEG  
   **Moving Picture Experts Group.**

3. MSE  
   **Mean Squared Error.**

4. MSSIM  
   **Mean Structural Similarity Index Measurement.**

5. PSNR  
   **Peak to peak signal to Noise Ratio.**

6. SSIM  
   **Structural Similarity Index. Measurement**

7. UIQI  
   **Universal Image Quality Index.**

8. JPEG  
   **Joint Photographic Experts Group.**

9. MOS  
   **Mean Opinion Score.**

10. AWGN  
    **Additive White Gaussian Noise.**

11. RMSE  
    **Root Mean Square Error.**
ABSTRACT

In this project a new universal objective image quality index was studied which is applicable to various image processing applications. This index which is easy to calculate models any kind of distortion as a combination of three factors which are loss of correlation, luminance distortion and contrast distortion. The goal of this project is to determine the image quality using this index and the decision is better than the traditional error summation methods such as mean squared error (MSE), peak signal to noise ratio (PSNR), root mean square error (RMSE) and mean absolute error (MAE). However in this project we only compare our mathematically defined quality factor to the MSE of the image to define a good image. This approach does not depend on the type or size of the testing image. It is also independent on pixel size of the image and the viewing conditions. Hence the term Universal is rightly apt for this approach. The dynamic range of Q extends from 1 to -1 where positive one represents a better image compared to the one with the negative one index.

Also the structural similarity measure was studied which is based on the HVS assumption that human visual perception is highly adapted for extracting structural information from a scene. This measure is an alternative complimentary framework for quality assessment based on the degradation of structural information. The luminance of the surface of an object being observed is the product of the illumination and the reflectance, but the structures of objects in the scene are independent of the illumination. The structural information of an image is defined as that attributes the represent the structure of objects in the scene, independent of the average luminance and contrast.

The image signal which is generally stationary and space variant is distorted by a wide variety of corruptions like impulsive salt–pepper noise, additive Gaussian noise, multiplicative speckle noise, additive Gaussian noise, multiplicative speckle noise contrast stretching, blurring and JPEG compression. The MSE, PSNR , Q and
SSIM for each of them are calculated using the MATLAB code first for a Lena image and then for another test images (Couple and Goldhill). The results are shown and conclusions are drawn.

**INTRODUCTION**

**MSE - MEAN SQUARED ERROR [14]**

MSE is a signal fidelity measure. The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity/ fidelity or, conversely, the level of error/distortion between them. Usually, it is assumed that one of the signals is a pristine original, while the other is distorted or contaminated by errors.

Suppose that $x = \{ x_i | i = 1, 2, \cdots, N \}$ and $y = \{ y_i | i = 1, 2, \cdots, N \}$ are two finite-length, discrete signals (e.g., visual images), where $N$ is the number of signal samples (pixels, if the signals are images) and $x_i$ and $y_i$ are the values of the $i$th samples in $x$ and $y$, respectively. The MSE between the signals $x$ and $y$ is

$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2$$

In the MSE, we will often refer to the error signal $e_i = x_i - y_i$, which is the difference between the original and distorted signals. If one of the signals is an original signal of acceptable (or perhaps pristine) quality, and the other is a distorted version of it whose quality is being evaluated, then the MSE may also be regarded as a measure of signal quality.

A more general form is the $lp$ norm is

$$d_p(x, y) = \left( \sum_{i=1}^{N} |e_i|^p \right)^{1/p}$$

MSE is often converted into a peak-to-peak signal-to-noise ratio (PSNR) measure

$$\text{PSNR} = 10 \log_{10} \frac{L^2}{\text{MSE}}$$

where $L$ is the dynamic range of allowable image pixel intensities. For example, for images that have allocations of 8 bits/pixel of gray-scale, $L = 2^8 - 1 = 255$. The
PSNR is useful if images having different dynamic ranges are being compared, but otherwise contains no new information relative to the MSE.

**WHY MSE [14]?**
The MSE has many attractive features:

1. It is simple. It is parameter free and inexpensive to compute, with a complexity of only one multiply and two additions per sample. It is also memoryless—the squared error can be evaluated at each sample, independent of other samples.

2. It has a clear physical meaning—it is the natural way to define the energy of the error signal. Such an energy measure is preserved after any orthogonal (or unitary) linear transformation, such as the Fourier transform (Parseval’s theorem). The energy preserving property guarantees that the energy of a signal distortion in the transform domain is the same as in the signal domain.

3. The MSE is an excellent metric in the context of optimization. Minimum-MSE (MMSE) optimization problems often have closed-form analytical solutions, and when they do not, iterative numerical optimization procedures are often easy to formulate, since the gradient and the Hessian matrix [14] of the MSE are easy to compute.

4. MSE is widely used simply because it is a convention. Historically, it has been employed extensively for optimizing and assessing a wide variety of signal processing applications, including filter design, signal compression, restoration, denoising, reconstruction, and classification. Moreover, throughout the literature, competing algorithms have most often been compared using the MSE/PSNR. It therefore provides a convenient and extensive standard against which the MSE/PSNR results of new algorithms may be compared. This saves time and effort but further propagates the use of the MSE.

**WHAT IS WRONG WITH MSE ?[14]**
It is apparent that the MSE possesses many favorable properties for application and analysis, but the reader might point out that a more fundamental issue has been missing. That is, does the MSE really measure signal fidelity? Given all of its attractive features, a signal processing practitioner might opt for the MSE if it proved to be a reasonable signal fidelity measure. But is that the case?
Unfortunately, the converse appears true when the MSE is used to predict human perception of image fidelity and quality. An illustrative example is shown in Figure 1[14], where an original Einstein image is altered by different types of distortion: a contrast stretch, mean luminance shift, contamination by additive white Gaussian noise, impulsive noise distortion, JPEG compression [18][12], blur, spatial scaling, spatial shift, and rotation.

FIG 1: Comparison of image fidelity measures for “Einstein” image altered with different types of distortion. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression [16] (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise). [2]

In figure 1, both MSE values and values of another quality index, the structural similarity (SSIM) index, are given. The SSIM index is described in detail later. Note that the MSE values [relative to the original image (a)] of several of the distorted images are nearly identical [images (b)–(g)], even though the same images present dramatically (and obviously) different visual quality. Also notice that images that undergo small geometrical modifications [images (h)–(i)] may have very large MSE values relative to the original, yet show a negligible loss of perceived quality. So a natural question is: “What is the problem with the MSE?”[14]
IMPLICIT ASSUMPTIONS WHEN USING THE MSE

1. Signal fidelity is independent of temporal or spatial relationships between the samples of the original signal. In other words, if the original and distorted signals are randomly re-ordered in the same way, then the MSE between them will be unchanged.
2. Signal fidelity is independent of any relationship between the original signal and the error signal. For a given error signal, the MSE remains unchanged, regardless of which original signal it is added to.
3. Signal fidelity is independent of the signs of the error signal samples.
4. All signal samples are equally important to signal fidelity.

Unfortunately, not one of them holds (even roughly) in the context of measuring the visual perception of image fidelity. Dramatic visual examples of the failure of the MSE with respect to the veracity of these assumptions is demonstrated in Figure 2.

SUBJECTIVE VS OBJECTIVE IMAGE QUALITY MEASURES[2]

Since human beings are the ultimate receivers I most image-processing applications, the most reliable way of assessing the quality of an image is by subjective evaluation. Indeed, the mean opinion score, a subjective quality measure requiring the services of a number of human observers, has been long regarded as the best method of image quality measurement. However, the MOS method is expensive, and it is usually too slow to be useful in real-world applications.

The goal of objective image quality assessment research is to design computational models that can predict perceived image quality accurately and automatically. We use the term predict here, since the numerical measures of quality that an algorithm provides are useless unless they correlate well with human subjectivity. In other words, the algorithm should predict the quality of an image that an average human observer will report.
Clearly, the successful development of such objective image quality measures has great potential in a wide range of application environments. First, they can be used to monitor image quality in quality control systems. For example, an image acquisition system can use a quality metric to monitor and automatically adjust itself to obtain the best quality image data. A network video server can examine the quality of the digital video transmitted on the network to control and allocate streaming resources. In light of the recent gigantic growth of internet video sources, this application is quite important.

Second, they can be employed to benchmark image-processing systems and algorithms. For instance, if a number of image denoising and restoration algorithms are available to enhance the quality of images captured using digital cameras, then a quality metric can be deployed to determine which of them provides the best quality results.

Third, they can be embedded into image-processing and transmission systems to optimize the systems and the parameter settings. For example, in the visual communication system, an image quality measure can assist in the optimal design of the pre-filtering and bit assignment algorithms at the encoder and of optimal reconstruction, error concealment, and post-filtering algorithms at the decoder.

In the design and selection of image quality assessment methods, there is often a tradeoff between accuracy and complexity, depending on the application scenario. For example, if there were an objective system that could completely simulate all relevant aspects of the human visual system, including its built-in knowledge of the environment, then it is should be able to supply precise predictions of image quality. However, our knowledge of the HVS and our models of the environment remain limited in their sophistication. As we increase our knowledge in these domains, then it is to be expected that image quality assessments systems that come very close to human performance will be developed.

However, the predictive performance of such systems to subjective human quality assessment has generally been quite poor. Indeed, while these methods for quality assessment have found considerable use as analytic metrics for theoretical algorithm design, they have been long been considered as rather weak for assessing the quality of real images, processed or otherwise.
**UNIVERSAL IMAGE QUALITY INDEX (UIQI)[1]**

Universal image quality index is easy to calculate and applicable to various image processing applications. It is a mathematically defined measure which is attractive because of two reasons. First, they are easy to calculate and usually have low computational complexity. Second, they are independent of viewing conditions and individual observers. Although it is believed that the viewing conditions play important roles in human perception of image quality, they are, in most cases not fixed and specific data is generally unavailable to the image analysis system. If there are N different viewing conditions, a viewing condition-dependent method will generate N different measurement results that are inconvenient to use. In addition, it becomes the user’s responsibilities to measure the viewing conditions and to calculate and input the condition parameters to the measurement systems. By contrast, a viewing condition-independent measure delivers a single quality value that gives a general idea of how good the image is.

The universality in the image quality index means that the approach does not depend on the imager being tested, the viewing conditions or the individual observers. More importantly, it must be applicable to various image processing and provide meaningful comparison across different types of image distortions. UIQI attempts to replace the currently and widely used PSNR and MSE techniques.
DEFINITION OF THE QUALITY INDEX

Let \( \mathbf{x} = \{x_i, i=1,2,\ldots,N\} \) and \( \mathbf{y} = \{y_i, i=1,2,\ldots,N\} \) be the original and the test image signals respectively. The proposed quality index is defined as

\[
Q = \frac{4\sigma_{xy} \bar{x} \bar{y}}{(\sigma_x^2 + \sigma_y^2) [(\bar{x})^2 + (\bar{y})^2]}
\]

Where

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i
\]

\[
\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2, \quad \sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2
\]

\[
\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}).
\]

Where \( \bar{x} \) is the mean of the original image and \( \bar{y} \) is the mean of the test image. \( \sigma_x \) and \( \sigma_y \) are the variances of the original and test images respectively. Cross variance is denoted by \( \sigma_{xy} \). The dynamic range of \( Q \) is \([-1,1]\) and the best value of 1 is achieved only if the original image is equal to the test image for all values of \( N \). The worst value of -1 occurs when the test image is twice the mean of
original image subtracted by the original image. This quality index models any distortion as a combination of three different factors: loss of correlation, luminance distortion, and contrast distortion. In order to understand this, we rewrite the definition of Q as a product of three components.

\[
Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}
\]

The first component is the correlation coefficient between the original and test image, which is the measure of linear correlation. Its range extends from 1 to -1 and the best value is obtained when the test image is equal to the original image multiplied with a positive constant. The relative distortions present after the original and test images are correlated are evaluated in the second and third components. The second component measures the mean luminance between the original and test image and its range is [0,1]. This component has the maximum value when the means of the original and test image are same. The variance of the signal can be viewed as an estimate of contrast and so the third component measures how similar the contrasts of the images are. Its range of values is also [0,1] and the best value is achieved if and if only the variances are equal.
APPLICATION TO IMAGES
An image consists of numerous pixels and signals but in practice a single overall quality value is considered as the image signals are generally non-stationary while the image quality is often space variant. That is why the statistical features are measured locally and then combined together to form an overall measure. The measurement method applied to the local regions is the sliding window approach. A sliding window of size B*B moves pixel by pixel starting from the top left corner of the image first horizontally and then vertically through all the rows and columns of the image until the bottom right corner as shown in the figure.
FIG 2: Movement of sliding window (BxB) in the horizontal position [20]
FIG 3: Movement of sliding window (B*B) in the vertical direction
FIG 4: Example of a sliding window

After the sliding window covers the whole image the overall quality index is calculated. The sliding window size is selected to be $B=8$ as default. The quality index at $j$-th step is computed as $Q_j$ and the overall quality index is given by
One recently proposed approach to image fidelity measurement, which may also prove highly effective for measuring the fidelity of other signals, is the SSIM index. The principal philosophy underlying the original SSIM approach is that the human visual system [3] is highly adapted to extract structural information from visual scenes. Therefore, at least for image fidelity measurement, the retention of signal structure should be an important ingredient. Equivalently, an algorithm may seek to measure structural distortion to achieve image fidelity measurement. Figure 5 10] helps illustrate the distinction between structural and nonstructural distortions. In the figure, the nonstructural distortions (a change of luminance or brightness, a change of contrast, Gamma distortion, and a spatial shift) are caused by ambient environmental or instrumental conditions occurring during image acquisition and display. These distortions do not change the structures of images of the objects in the visual scene. However, other distortions (additive noise and blur and lossy compression) significantly distort the structures of images of the objects. If we view the human visual system as an ideal information extractor that seeks to identify and recognize objects in the visual scene, then it must be highly sensitive to the structural distortions and automatically compensates for the nonstructural distortions. Consequently, an effective objective signal fidelity measure should simulate this functionality.
FIG 5: Examples of structural versus nonstructural distortions.[14]

FIG 6: Block diagram of structural similarity measurement system [10]
The system diagram of structural similarity measurement system is shown in Fig. 4. Suppose that \( x = \{ x_i | i = 1, 2, \cdots, N \} \) and \( y = \{ y_i | i = 1, 2, \cdots, N \} \) are two finite-length image signals, which have been aligned with each other (e.g., spatial patches extracted from each image), where \( N \) is the number of signal samples (pixels, if the signals are images) and \( x_i \) and \( y_i \) are the values of the \( i \)th samples in \( x \) and \( y \), respectively.

If we consider one of the signals to have perfect quality, then the similarity measure can serve as a quantitative measurement of the quality of the second signal. The system separates the task of similarity measurement into three comparisons: luminance, contrast and structure. First, the luminance of each signal is compared. Assuming discrete signals, this is estimated as the mean intensity

\[
\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{1}
\]

The luminance comparison function \( l(x, y) \) is then a function of \( \mu_x \) and \( \mu_y \).

Second, we remove the mean intensity from the signal. In discrete form, the resulting signal \( (x - \mu_x) \) corresponds to the projection of vector onto the hyperplane defined by

\[
\sum_{i=1}^{N} X_i = 0 \tag{2}
\]

where \( X = \{ X_i | i = 1, 2, \cdots, N \} \) is a finite-length image signal.

We use the standard deviation (the square root of variance) as an estimate of the signal contrast. An unbiased estimate in discrete form is given by

\[
\sigma_x = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2} \tag{3}
\]

The contrast comparison \( c(x, y) \) is then the comparison of \( \sigma_x \) and \( \sigma_y \).

Third, the signal is normalized (divided) by its own standard deviation, so that the two signals being compared have unit standard deviation. The structure comparison \( s(x, y) \) is implemented on these normalized signals \( [(x - \mu_x)/\sigma_x] \) and \( [(y - \mu_y)/\sigma_y] \).

Finally, the three components are combined to yield an overall similarity measure

\[
S(x, y) = f(l(x, y), c(x, y), s(x, y)) \tag{4}
\]
where \[ l(x,y) = f(\mu_x, \mu_y) \]
\[ c(x,y) = f(\sigma_x, \sigma_y) \]
\[ s(x,y) = f([(x - \mu_x) / \sigma_x], [(y - \mu_y) / \sigma_y]) \]

An important point is that the three components are relatively independent. For example, the change of luminance and/or contrast will not affect the structures of images. In order to complete the definition of the similarity measure in (4), we need to define the three functions \( l(x,y) \), \( c(x,y) \), and \( s(x,y) \) as well as the combination function \( f(.) \). We also would like the similarity measure to satisfy the following properties.

1) Symmetry: \( S(x,y) = S(y,x) \).

2) Boundedness: \( S(x,y) \leq 1 \).

3) Unique maximum: \( S(x,y) = 1 \) if and only if \( x = y \) (in discrete representations, \( x_i = y_i \) for all \( i = 1, 2, \cdots, N \)).

For luminance comparison, we define

\[
l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}
\]

where the constant \( C_1 \) is included to avoid instability when \( \mu_x^2 + \mu_y^2 \) is very close to zero. Specifically, we choose

\[
C_1 = L^2K_1^2
\]

where \( L \) is the dynamic range of the pixel values (255 for 8-bit grayscale images), and \( K_1 \ll 1 \) is a small constant. Similar considerations also apply to contrast comparison and structure comparison described later. Equation (4) is easily seen to obey the three properties of SSIM. Equation (4) is also qualitatively consistent with Weber’s law, which has been widely used to model light adaptation (also called luminance masking) in the HVS. According to Weber’s law [18], the magnitude of a just-noticeable luminance change \( \Delta I \) is approximately proportional to the background luminance for a wide range of luminance values. In other words, the HVS is sensitive to the relative luminance change, and not the absolute luminance change. Letting \( R \) represent the size of luminance change relative to background luminance, we can write \( R \) as

\[
R = \frac{\mu_y}{\mu_x} - 1
\]

Substituting \( R \) into (4) gives
\[ l(x, y) = \frac{2(1 + R)}{1 + (1 + R)^2 + \frac{C_1}{\mu_x^2}} \]  
\[ (8) \]

If we assume \( C_1 \) is small enough (relative to \( \mu_x^2 \)) to be ignored, then \( l(x, y) \) is a function only of \( R \), qualitatively consistent with Weber’s law. The contrast comparison function takes a similar form

\[ c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \]
\[ (9) \]

where \( C_2 = (K_2L^2) \), and \( K_2 \ll 1 \).

This definition again satisfies the three properties. An important feature of this function is that with the same amount of contrast change \( \Delta \sigma = \sigma_y - \sigma_x \), this measure is less sensitive to the case of high base contrast than low base contrast. This is consistent with the contrast-masking feature of the HVS [10].

Structure comparison is conducted after luminance subtraction and variance normalization. Specifically, we associate the two unit vectors \( \frac{(x - \mu_x)}{\sigma_x} \) and \( \frac{(y - \mu_y)}{\sigma_y} \), each lying in the hyperplane defined by (4), with the structure of the two images. The correlation (inner product) between these is a simple and effective measure to quantify the structural similarity. Notice that the correlation between \( \frac{(x - \mu_x)}{\sigma_x} \) and \( \frac{(y - \mu_y)}{\sigma_y} \) is equivalent to the correlation coefficient between \( x \) and \( y \). Thus, we define the structure comparison function as follows:

\[ s(x, y) = \frac{\sigma_{xy}^2 + C_3}{\sigma_x \sigma_y + C_3} \]
\[ (10) \]

As in the luminance and contrast measures, we have introduced a small constant \( C_3 \) in both denominator and numerator. In discrete form \( \sigma_{xy} \), can be estimated as

\[ \sigma_{xy} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)} \]
\[ (11) \]

Note also that \( s(x, y) \) can take on negative values. Finally, we combine the three comparisons of (4), (7) and (8) and name the resulting similarity measure the SSIM index between signals and

\[ SSIM(x, y) = [l(x, y)]^p [c(x, y)]^q [s(x, y)]^r \]
\[ (12) \]
where $\alpha > 1$, $\beta > 1$, and $\gamma > 1$ are parameters used to adjust the relative importance of the three components. It is easy to verify that this definition satisfies the three conditions given above. In order to simplify the expression, we set $\alpha = \beta = \gamma = 1$ and $C_3 = C_2 / 2$. This results in a specific form of the SSIM index

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

(13)

In practice, one usually requires a single overall quality measure of the entire image. We use a mean SSIM (MSSIM) index to evaluate the overall image quality:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^{M} SSIM(x_j, y_j)$$

(14)

where $X$ and $Y$ are the reference and the distorted images, respectively; $x_j$ and $y_j$ are the image contents at the $j^{th}$ local window; and $M$ is the number of local windows of the image.

To apply the SSIM index for image quality measurement it is preferable to apply it locally (on image blocks or patches) than globally (over the entire image). SSIM index is most commonly computed within the local window which moves pixel by pixel across the entire image. Such a sliding window approach is shown in figure 7.
FIG 7: SLIDING WINDOW FOR SSIM[16]

**IMAGES WITH DISTORTIONS[20]**

Here we have applied different types of distortion to the image “Lena.gif” and the Mean Squared Error (MSE) and Quality index (Q) for the various pair of images is calculated. The traditional error measuring techniques are mainly MSE and Peak Signal to Noise Ratio (PSNR). These are widely used because they are simple to calculate and are independent of viewing conditions and individual observers. Quality index on the other hand is designed by modeling any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. It performs significantly better than the widely used distortion metric mean square error.
1). Salt and pepper noise: It represents itself as randomly occurring white and black pixels. An effective noise reduction method for this type of noise involves the usage of a median filter. Salt and pepper noise creeps into images in situations where quick transients, such as faulty switching, take place. The image after distortion from salt and pepper noise looks like the image below.
2). **Multiplicative Speckle noise**: Speckle noise is a granular noise that inherently exists in and degrades the quality of images. Speckle noise is a multiplicative noise, i.e. it is in direct proportion to the local grey level in any area. The signal and the noise are statistically independent of each other. The sample mean and variance of a single pixel are equal to the mean and variance of the local area that is centered on that pixel.
3). **Image blurring**: Blurring an image usually makes the image unfocused. In signal processing, blurring is generally obtained by convolving the image with a low pass filter. In this particular example we use Gaussian blurring.
4). **Contrast stretching**: Low-contrast images occur often due to poor or non-uniform lighting conditions or due to non-linearity or small dynamic range of the imaging sensor. It can be expressed as:

\[
V = \begin{cases} 
\alpha u, & 0 \leq u < a \\
\beta (u-a) + v_a, & a \leq u < b \\
\gamma (u-b) + v_b, & b \leq u < L 
\end{cases}
\]

The slope of the transformation is chosen greater than unity in the region of stretch. The parameters \(a\) and \(b\) can be obtained by examining the histogram of the image. For example the gray scale intervals where the pixels occur most frequently would be stretched most to improve the overall visibility of a scene. Here we have considered \(\alpha\) to be greater than 1.
Gaussian noise: Gaussian noise is statistical noise that has a probability density function of the normal distribution (also known as Gaussian distribution). In other words, the values that the noise can take on are Gaussian-distributed. It is most commonly used as additive white noise to yield additive white Gaussian noise (AWGN).
6) **Mean shift Algorithm**: The Mean Shift algorithm clusters an n-dimensional data set (i.e., each data point is described by a feature vector of n values) by associating each point with a peak of the data set’s probability density. For each point, Mean Shift computes its associated peak by first defining a spherical window at the data point of radius r and computing the mean of the points that lie within the window. The algorithm then shifts the window to the mean and repeats until convergence, i.e., until the shift is less than a threshold (e.g., 0.01). At each iteration the window will shift to a more densely populated portion of the data set until a peak is reached, where the data is equally distributed in the window.
7). **Jpeg Compression**: The original image is compressed in size and the MSE is calculated for the image. The compression ratio in this case is 4.8574. It is very difficult to compress using JPEG and to maintain at a certain MSE. The first dimension solutions of JPEG were more dominant although the visual inspection of these images resulted in three possible attributes such as blockiness, blur and ringing. For JPEG compression the original images were converted into monochrome images and a public-domain software package for JPEG encoding and decoding (independent software Group[4] ) was used to code the images
# RESULTS

## TABULATED RESULTS FOR LENA IMAGE

<table>
<thead>
<tr>
<th>S.no</th>
<th>Image</th>
<th>Mean square error</th>
<th>PSNR (dB)</th>
<th>Q</th>
<th>SSIM</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Original image</td>
<td>0</td>
<td>Inf</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>Mean shift</td>
<td>224.9993</td>
<td>24.6090</td>
<td>0.9894</td>
<td>0.9890</td>
<td>0.9894</td>
</tr>
<tr>
<td>3.</td>
<td>Contrast stretching</td>
<td>225.0932</td>
<td>24.6072</td>
<td>0.9372</td>
<td>0.9494</td>
<td>0.9470</td>
</tr>
<tr>
<td>4.</td>
<td>Impulsive salt-pepper noise</td>
<td>225.3684</td>
<td>24.6019</td>
<td>0.6494</td>
<td>0.7227</td>
<td>0.6723</td>
</tr>
<tr>
<td>5.</td>
<td>Multiplicative speckle Noise</td>
<td>224.7482</td>
<td>24.6138</td>
<td>0.4408</td>
<td>0.5009</td>
<td>0.4883</td>
</tr>
<tr>
<td>6.</td>
<td>Additive Gaussian noise</td>
<td>225.1804</td>
<td>24.6055</td>
<td>0.3891</td>
<td>0.4508</td>
<td>0.4390</td>
</tr>
<tr>
<td>7.</td>
<td>Blurring</td>
<td>224.1397</td>
<td>24.6256</td>
<td>0.3461</td>
<td><strong>0.6880</strong></td>
<td><strong>0.6007</strong></td>
</tr>
<tr>
<td>8.</td>
<td>JPEG Compression</td>
<td>215.1139</td>
<td>24.8041</td>
<td>0.2876</td>
<td><strong>0.6709</strong></td>
<td><strong>0.5572</strong></td>
</tr>
</tbody>
</table>
**LENA IMAGE WITH DISTORTION**

- **Original image**, $\text{MSE}=0, Q=1, \text{PSNR}=\infty \text{ dB}$
  - $\text{SSIM}=1, \text{MSSIM}=1$

- **Mean shift**, $\text{MSE}=224.9933, Q=0.9894, \text{PSNR}=24.6090 \text{ dB}$
  - $\text{SSIM}=0.9890, \text{MSSIM}=0.9894$

- **Contrast stretching**, $\text{MSE}=225.0932, Q=0.9372$
  - $\text{PSNR}=24.6072 \text{ dB}, \text{SSIM}=0.9494, \text{MSSIM}=0.9470$

- **Impulsive salt and pepper Noise**, $\text{MSE}=225.0932, Q=0.6494, \text{PSNR}=24.6019 \text{ dB}$
  - $\text{SSIM}=0.7227, \text{MSSIM}=0.6723$

- **Multiplicative speckle Noise**, $\text{MSE}=224.7482, Q=0.448, \text{PSNR}=24.61 \text{ dB}$
  - $\text{SSIM}=0.5009, \text{MSSIM}=0.4883$

- **Additive gaussian**, $\text{MSE}=225.1804, Q=0.3891$
  - $\text{PSNR}=24.6055 \text{ dB}, \text{SSIM}=0.4508, \text{MSSIM}=0.4390$
Blurring, MSE = 224.1397, Q = 0.3461, PSNR = 24.6256 dB
SSIM = 0.6880, MSSIM = 0.6

JPEG compression, MSE = 215.1139, Q = 0.2876, PSNR = 24.8041 dB
SSIM = 0.6709, MSSIM = 0.5572
## TABULATED RESULTS FOR GOLDHILL IMAGE

<table>
<thead>
<tr>
<th>S.no</th>
<th>Image</th>
<th>Mean square error</th>
<th>PSNR (dB)</th>
<th>Q</th>
<th>SSIM</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Original image</td>
<td>0</td>
<td>Inf</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Mean shift</td>
<td>121</td>
<td>27.3029</td>
<td>0.9928</td>
<td>0.9927</td>
<td>0.9929</td>
</tr>
<tr>
<td>3</td>
<td>Contrast stretching</td>
<td>120.9002</td>
<td>27.3065</td>
<td>0.9498</td>
<td>0.9698</td>
<td>0.9672</td>
</tr>
<tr>
<td>4</td>
<td>Impulsive salt-pepper noise</td>
<td>120.2122</td>
<td>27.3313</td>
<td>0.8290</td>
<td>0.8643</td>
<td>0.8402</td>
</tr>
<tr>
<td>5</td>
<td>Multiplicative speckle Noise</td>
<td>121.4297</td>
<td>27.2876</td>
<td>0.6758</td>
<td>0.7032</td>
<td>0.7067</td>
</tr>
<tr>
<td>6</td>
<td>Additive Guassian noise</td>
<td>121.1260</td>
<td>27.2984</td>
<td>0.6151</td>
<td>0.6556</td>
<td>0.6553</td>
</tr>
<tr>
<td>7</td>
<td>Blurring</td>
<td>121.9371</td>
<td>27.2694</td>
<td>0.5080</td>
<td><strong>0.6671</strong></td>
<td><strong>0.6372</strong></td>
</tr>
<tr>
<td>8</td>
<td>JPEG Compression</td>
<td>117.4739</td>
<td>27.4314</td>
<td>0.4963</td>
<td><strong>0.6824</strong></td>
<td><strong>0.6385</strong></td>
</tr>
</tbody>
</table>
GOLDFIILL IMAGE WITH DISTORTIONS

ORIGINAL IMAGE, MSE=0, PSNR=INF, Q=1, SSIM=1, MSSIM=1

MEAN SHIFT, MSE=121, PSNR 27.30dB, Q=0.9928, SSIM=0.9927, MSSIM=0.9929

CONTRAST STRICHING, MSE=120.9002, PSNR=27.3065, Q=0.9498, SSIM=0.9698, MSSIM=0.9672

IMPULSIVE SALT-PEPPER NOISE MSE=120.2122, PSNR=27.331MB Q=0.8290, SSIM=0.8645, MSSIM=0.8402

MULTIPlicative SPECKLE NOISE MSE=121.4297, PSNR=27.2876dB, Q=0.6758, SSIM=0.7032, MSSIM=0.7067

ADDITIVE GAUSSIAN NOISE MSE=121.1260, PSNR=27.2984dB, Q=0.6151, SSIM=0.6556, MSSIM=0.6553

BLURRING MSE=121.9371, PSNR=27.2694dB Q=0.5080, SSIM=0.6671, MSSIM=0.6372

JPEG COMPRESSION MSE=117.4739, PSNR=27.4314dB, Q=0.4963, SSIM=0.6824, MSSIM=0.6385
**TABULATED RESULTS FOR COUPLE IMAGE**

<table>
<thead>
<tr>
<th>S.no</th>
<th>Image</th>
<th>Mean square error</th>
<th>PSNR (dB)</th>
<th>Q</th>
<th>SSIM</th>
<th>MSSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Original image</td>
<td>0</td>
<td>Inf</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.</td>
<td>Mean shift</td>
<td>80.9952</td>
<td>29.0462</td>
<td>0.9947</td>
<td>0.9940</td>
<td>0.9947</td>
</tr>
<tr>
<td>3.</td>
<td>Contrast stretching</td>
<td>80.9063</td>
<td>29.0510</td>
<td>0.9621</td>
<td>0.9631</td>
<td>0.9657</td>
</tr>
<tr>
<td>4.</td>
<td>Impulsive salt-pepper noise</td>
<td>80.8358</td>
<td>29.0548</td>
<td>0.8806</td>
<td>0.9057</td>
<td>0.8874</td>
</tr>
<tr>
<td>5.</td>
<td>Multiplicative speckle Noise</td>
<td>81.2837</td>
<td>29.0308</td>
<td>0.7183</td>
<td>0.7613</td>
<td>0.7553</td>
</tr>
<tr>
<td>6.</td>
<td>Additive Guassian noise</td>
<td>80.6841</td>
<td>29.0629</td>
<td>0.7039</td>
<td>0.7511</td>
<td>0.7444</td>
</tr>
<tr>
<td>7.</td>
<td>Blurring</td>
<td>81.2747</td>
<td>29.0313</td>
<td>0.7500</td>
<td><strong>0.8329</strong></td>
<td><strong>0.8238</strong></td>
</tr>
<tr>
<td>8.</td>
<td>JPEG Compression</td>
<td>81.9302</td>
<td>28.994</td>
<td>0.6761</td>
<td><strong>0.8013</strong></td>
<td><strong>0.7771</strong></td>
</tr>
</tbody>
</table>
COUPLE IMAGE WITH DISTORTIONS

ORIGINAL IMAGE
MSE=0, PSNR=INF dB,
Q=I, SSIM=1, MSSIM=1

MEAN SHIFT
MSE=80.9952, PSNR=29.0462 dB,
Q=0.9917, SSIM=0.9941,
MSSIM=0.9947

CONTRAST STRECHING
MSE=80.9063, PSNR=29.0510 dB,
Q=0.9621, SSIM=0.9631,
MSSIM=0.9657

IMPULSIVE SALT-PEPPER NOISE
MSE=80.8358, PSNR=29.0548 dB,
Q=0.8806, SSIM=0.9057

MULTIPLICATIVE SPECKLE NOISE
MSE=81.2387, PSNR=29.0308 dB,
Q=0.7183, SSIM=0.7613,
MSSIM=0.7555

ADDITIVE GAUSSIAN NOISE
MSE=80.6841, PSNR=29.0629 dB,
Q=0.7039, SSIM=0.7511,
MSSIM=0.7444

BLURRING
MSE=81.2747, PSNR=29.0313 dB,
Q=0.7500, SSIM=0.8329,
MSSIM=0.8238

JPEG COMPRESSION
MSE=81.9302, PSNR=28.9944 dB,
Q=0.6761, SSIM=0.8013,
MSSIM=0.7771
CONCLUSIONS

Q which is a simple and mathematical model seems to be a better metric in image quality than compared to traditional MSE and PSNR. The success is due to the strong ability of Q to measure the structural distortion occurred during the image degradation processes when compared to MSE which is sensitive to energy of errors. There is no doubt that precise modeling of HVS is always better but a well defined mathematical framework of the model can ease in successful quality metric.

The SSIM index is a particular implementation of the philosophy of structural similarity from an image formation point of view. The key success of SSIM is the concept of structural information and structural distortion. SSIM index exhibit input-dependent behavior in measuring signal distortions. The MSSIM is a better metric than UIQI in the case of distortions like blurring and JPEG compression due to the inclusion of the constants C1,C2 in SSIM which avoid the instability. Even though the MSE of different distortions is same, SSIM and MSSIM truly represent the visual (perceptual) qualities.
REFERENCES


5. VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment.”


8. Z. Wang’s website


11. JPEG reference website [http://ee.uta.edu/Dip/Courses/EE5359/JPEG_1.pdf](http://ee.uta.edu/Dip/Courses/EE5359/JPEG_1.pdf)

12. JPEG reference website [http://ee.uta.edu/Dip/Courses/EE5359/JPEG_2.pdf](http://ee.uta.edu/Dip/Courses/EE5359/JPEG_2.pdf)


20. Digital image processing assignment website: [http://www-ee.uta.edu/dip/Courses/EE5356/ee_5356.htm](http://www-ee.uta.edu/dip/Courses/EE5356/ee_5356.htm).