**P²SNR: Perceptual Full-Reference Image Quality Assessment for JPEG2000**

Jaime Moreno  
Superior School of Mechanical and Electrical Engineers,  
National Polytechnic Institute, Mexico City, Mexico  
jmorenoe@ipn.mx

**Abstract**—Estimation of image quality is decisive in the image compression field. This is important in order to minimize, induced error via rate allocation[1]. Traditional full-reference algorithms of image quality try to model how Human Visual System detects visual differences and extracts both information and structure of the image. In this work we propose a quality assessment, which weights the mainstream PSNR by means of a perceptual model (P²SNR). Perceptual image quality is obtained by estimating the rate of energy loss when an image is observed at monotonically increasing distances. Experimental results show that P²SNR is the best-performing algorithm, compared with another eight metrics such as MSSIM, SSIM or VIF, among others, when an image is distorted by a wavelet compression. It has been tested across TID2008 image database.

**Keywords:** Human Visual System, Wavelet Transform, Chromatic Induction Model, Image Quality Assessment

1. Introduction

Mean Squared Error (MSE) is still the most used quantitative performance metrics and several image quality measures are based on it, being Peak Signal-to-Noise Ratio (PSNR) the best example. Wang and Bovik in [2], [3] consider that MSE is a poor device to be used in quality assessment systems. Therefore it is important to know what is the MSE and what is wrong with it, in order to propose new metrics that fulfills the properties of human visual system and keeps the favorable features that the MSE has.

In this way, let \( f(i,j) \) and \( \hat{f}(i,j) \) represent two images being compared and the size of them is the number of intensity samples or pixels. Being \( f(i,j) \) the original reference image, which has to be considered with perfect quality, and \( \hat{f}(i,j) \) a distorted version of \( f(i,j) \), whose quality is being evaluated. Then, the MSE and the PSNR are, respectively, defined as:

\[
MSE = \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} [f(i,j) - \hat{f}(i,j)]^2
\]

(1)

and

\[
PSNR = 10 \log_{10} \left( \frac{G_{\text{max}}^2}{MSE} \right)
\]

(2)

where \( G_{\text{max}} \) is the maximum possible intensity value in \( f(i,j) \) (\( M \times N \) size). Thus for gray-scale images that allocate 8 bits per pixel (bpp) \( G_{\text{max}} = 2^8 - 1 = 255 \). For color images the PSNR is defined as in the Equation 2, whereas the color MSE is the mean among the individual MSE of each component. MSE does not need any positional information in the image, thus pixel arrangement is ordered as a one-dimensional vector.

Both MSE and PSNR are extensively employed in the image processing field, since these metrics have favorable properties, such as:

1) A convenient metrics for the purpose of algorithm optimization. For example in JPEG2000, MSE is used both in Optimal Rate Allocation [4], [1] and Region of interest [5], [1]. Therefore MSE can find solutions for these kind of problems, when is combined with the instruments of linear algebra, since it is differentiable.

2) By definition MSE is the difference signal between the two images being compared, giving a clear meaning of the overall error signal energy.

However, the MSE has a poor correlation with perceived image quality. An example is shown in Fig. 1, where both Baboon(a) and Splash(b) Images are distorted by means of a JPEG2000 compression with 30dB of PSNR. These noisy images present dramatically different visual qualities. Thereby either MSE or PSNR do not reflect the way that human visual system (HVS) perceives the images, since these measures represent an input image in a pixel domain.

2. perceptual Peak Signal-to-Noise Ratio

2.1 Introduction

In the referenced image quality issue, there is an original image \( f(i,j) \) and a distorted version \( \hat{f}(i,j) = \Lambda[f(i,j)] \) that is compared with \( f(i,j) \), being \( \Lambda \) a distortion model. The difference between these two images depends on the characteristics the distortion model \( \Lambda \). For example, blurring, contrast change, noise, JPEG blocking or wavelet ringing.

In Fig. 2, the images Babbon and Splash are compressed by means of JPEG2000. These two images have the same PSNR=30 dB when compared to their corresponding original image, that is, they have the same numerical degree of distortion (i.e. the same objective image quality PSNR).
But, their subjective quality is clearly different, showing the image *Baboon* a better visual quality. Thus, for this example, PSNR and perceptual image quality has a small correlation. On the image *Baboon*, high spatial frequencies are dominant. A modification of these high spatial frequencies by \( \Lambda \) induces a high distortion, resulting a lower PSNR, even if the modification of these high frequencies are not perceived by the HVS. In contrast, on image *Splash*, mid and low frequencies are dominant. Modification of mid and low spatial frequencies also introduces a high distortion, but they are less perceived by the HVS. Therefore, correlation of PSNR against the opinion of an observer is small. Fig. 2 shows the diagonal high spatial frequencies of these two images, where they are more high frequencies in image *Baboon*.

![Baboon](image1.png) ![Splash](image2.png)

Fig. 1

256 × 256 PATCHES OF IMAGES Baboon AND Splash DISTORTED BY MEANS OF JPEG2000 BOTH WITH PSNR=30dB, CROPPED FOR VISIBILITY.

If a set of distortions \( \hat{f}_k(i,j) = \Lambda_k[f(i,j)] \) are generated and indexed by \( k \) (for example, let \( \Lambda \) be a blurring operator), the image quality of \( \hat{f}_k(i,j) \) evolves while varying \( k \), being \( k \), for example, the degree of blurring. Hence, the evolution of \( \hat{f}_k(i,j) \) depends on the characteristics of the original \( f(i,j) \). Thus, when increasing \( k \), if \( f(i,j) \) contains many high spatial frequencies the PSNR rapidly decreases, but when low and mid frequencies predominated PSNR decreases slowly.

Similarly, the HVS is a system that induces a distortion on the observed image \( f(i,j) \), whose model is predicted by The Chromatic Induction Wavelet Model (CIWaM) [6].

CIWaM takes an input image \( I \) and decomposes it into a set of wavelet planes \( \omega_{s,o} \) of different spatial scales \( s \) (i.e., spatial frequency \( \nu \)) and spatial orientations \( o \). It is described as

\[
I = \sum_{s=1}^{n} \sum_{o=v,h,d} \omega_{s,o} + c_n,
\]

where \( n \) is the number of wavelet planes, \( c_n \) is the residual plane and \( o \) is the spatial orientation either vertical, horizontal or diagonal.

The perceptual image \( I_\rho \) is recovered by weighting these \( \omega_{s,o} \) wavelet coefficients using the extended Contrast Sensitivity Function (e-CSF).

Perceptual image \( I_\rho \) can be obtained by

\[
I_\rho = \sum_{s=1}^{n} \sum_{o=v,h,d} \alpha(\nu,r) \omega_{s,o} + c_n,
\]

where \( \alpha(\nu,r) \) is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term \( \alpha(\nu,r) \omega_{s,o} \equiv \omega_{s,o,p,d} \) can be considered the perceptual wavelet coefficients of image \( I \) when observed at distance \( d \). For details on the CIWaM and the \( \alpha(\nu,r) \) function, see [6].

Hence, CIWaM is considered a HSV particular distortion model \( \Lambda \equiv \text{CIWaM} \) that generates a perceptual image \( \hat{f}_\rho(i,j) \equiv I_\rho \) from an observed image \( f(i,j) \equiv I \), i.e \( I_\rho = \text{CIWaM}[I] \). Therefore, a set of distortions is defined as \( \Lambda_k \equiv \text{CIWaM}_d \), being \( d \) the observation distance. That is, a set of perceptual images are defined \( I_{p,d} = \text{CIWaM}_d[I] \) which are considered a set of perceptual distortions of image \( I \).

When images \( f(i,j) \) and \( \hat{f}(i,j) \) are simultaneously observed at distance \( \hat{d} \) and this distance is reduced, the differences between them are better perceived. In contrast, if \( f(i,j) \) and \( \hat{f}(i,j) \) are observed from a far distance human eyes cannot perceive their differences, in consequence, the perceptual image quality of the distorted image is always high. The distance where the observer cannot distinguish any difference between these two images is \( d = \infty \). In practice, \( \hat{d} = D \) where differences are not perceived and range some centimeters from the position of the observer. Consequently, the less distorted \( \hat{f}(i,j) \), that is, the highest the image quality of \( f(i,j) \), the shorter the distance \( D \).
2.2 Methodology

Let \( f(i,j) \) and \( \tilde{f}(i,j) = \Lambda[f(i,j)] \) be an original image and a distortion version of \( f(i,j) \), respectively. \( P^2\text{SNR} \) methodology is based on finding a distance \( D \), where there is no perpetual difference between the wavelet energies of the images \( f(i,j) \) and \( \tilde{f}(i,j) \), when an observer watch them at distance \( d \) centimeters of observation distance. So measuring the PSNR of \( f(i,j) \) at \( D \) will yield a fairer perceptual evaluation of its image quality.

\( P^2\text{SNR} \) algorithm is divided in five steps, which is described as follows:

**Step 1: Wavelet Transformation**

Wavelet transform of images \( f(i,j) \) and \( \tilde{f}(i,j) \) is performed using Eq. 3, obtaining the sets \( \{\omega_{s,o}\} \) and \( \{\omega_{s,o}\} \), respectively. The employed analysis filter is the Daubechies 9-tap/7-tap filter (Table 1).

**Step 2: Distance \( D \)**

The total energy measure or the deviation signature[7] \( \bar{\varepsilon} \) is the absolute sum of the wavelet coefficient magnitudes, defined by [8]

\[
\bar{\varepsilon} = \sum_{n=1}^{N} \sum_{m=1}^{M} |x(m,n)| 
\]

where \( x(m,n) \) is the set of wavelet coefficients, whose energy is being calculated, being \( m \) and \( n \) the indexes of the coefficients. Basing on the traditional definition of a calorie, the units of \( \bar{\varepsilon} \) are wavelet calories (wCal) and can also be defined by Eq. 5, since a wCal is the energy needed to increase the absolute magnitude of a wavelet coefficient by one scale.

From wavelet coefficients \( \{\omega_{s,o}\} \) and \( \{\omega_{s,o}\} \) the corresponding perceptual wavelet coefficients \( \{\omega_{s,o,p,d}\} = \alpha(\nu,r) \cdot \omega_{s,o} \) and \( \{\omega_{s,o,p,d}\} = \alpha(\nu,r) \cdot \tilde{\omega}_{s,o} \) are obtained by applying CIWaM with an observation distance \( \tilde{d} \). Therefore Equation 6 expresses the relative wavelet energy ratio \( \varepsilon \mathcal{R}(\mathcal{d}) \), which compares how different are the energies of the reference and distorted CIWaM perceptual images, namely \( \varepsilon_{\rho} \) and \( \tilde{\varepsilon}_{\rho} \) respectively, when these images are watched from a given distance \( \tilde{d} \).

\[
\varepsilon \mathcal{R}(\mathcal{d}) = 10 \cdot \log_{10} \frac{\varepsilon_{\rho}(\mathcal{d})}{\tilde{\varepsilon}_{\rho}(\mathcal{d})}
\]

Fig. 3(a) shows that distance \( D \) is composed by the sum of two distances, \( n\mathcal{P} \) and \( \varepsilon m\mathcal{L} \). Thereby for the estimation of \( D \), Eq. 7, it is necessary to know the observation distance \( d \) besides to figure out the \( n\mathcal{P} \) and \( \varepsilon m\mathcal{L} \) distances. Furthermore Fig. 3(b) depicts a chart of \( \varepsilon \mathcal{R} \), which sketches both the behavior of the relative energy when \( \tilde{d} \) is varied from 0 to \( \infty \) centimeters and the meaning of the distances \( D \), \( n\mathcal{P} \) and \( \varepsilon m\mathcal{L} \) inside an \( \varepsilon \mathcal{R} \) chart.

\[
D = n\mathcal{P} + \varepsilon m\mathcal{L}
\]
\( \varepsilon R \,(D) \approx 0 \). This is achieved by projecting the points \((nP, \varepsilon R \,(nP))\) and \((d, \varepsilon R \,(d))\) to \((D, 0)\). Therefore \( \varepsilon mL \) is the needed length to match the energies from the point where the observer has the best evaluation of the assessed images to \(D\) and is described as follows:

\[
\varepsilon mL = \frac{\varepsilon R \,(nP)}{d \varepsilon R + \zeta} \tag{8}
\]

where \( \varepsilon R \,(nP) \) is the relative energy at \(nP\) and \(d \varepsilon R \) is the energy loss rate (wCal/cm or wCal/visual degrees) between \((nP, \varepsilon R \,(nP))\) and \((d, \varepsilon R \,(d))\), namely, the negative slope of the line joining these points, expressed as:

\[
d \varepsilon R = \frac{\varepsilon R \,(nP) - \varepsilon R \,(d)}{d - nP} \tag{9}
\]

When a lossless compression is performed, consequently \(f(i,j) = \hat{f}(i,j)\), hence \(d \varepsilon R = 0\) and \(\varepsilon mL \to \infty\). In order to numerically avoid it, parameter \(\zeta\) is introduced, which is small enough to not affect the estimation of \(\varepsilon mL\) when \(d \varepsilon R \neq 0\), in our MatLab implementation \(\zeta = \text{realmin}\).

**Step 3: Perceptual Images**

Obtain the perceptual wavelet coefficients \(\{\omega_{s,0,p,D}\} = \alpha(\nu, r) \cdot \omega_{s,0}\) and \(\{\hat{\omega}_{s,0,p,D}\} = \alpha(\nu, r) \cdot \hat{\omega}_{s,0}\) at distance \(D\), using Equation 4.

**Step 4: Inverse Wavelet Transformation**

Perform the Inverse Wavelet Transform of \(\{\omega_{s,0,p,D}\}\) and \(\{\hat{\omega}_{s,0,p,D}\}\), obtaining the perceptual images \(f_{\rho(i,j),D}\) and \(\hat{f}_{\rho(i,j),D}\), respectively. The synthesis filter in Table 2 is an inverse Daubechies 9-tap/7-tap filter.

<table>
<thead>
<tr>
<th>Synthesis Filter</th>
<th>Low-Pass Filter (h_{\ell},(i))</th>
<th>High-Pass Filter (h_{h},(i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.1158070523456994</td>
<td>0.00249018263579</td>
</tr>
<tr>
<td>±1</td>
<td>-0.5912717631142470</td>
<td>-0.266841184428723</td>
</tr>
<tr>
<td>±2</td>
<td>-0.05754352622849957</td>
<td>-0.078223652652898785</td>
</tr>
<tr>
<td>±3</td>
<td>-0.0912717631142494</td>
<td>0.01686411844287495</td>
</tr>
<tr>
<td>±4</td>
<td>0.02674875741080976</td>
<td></td>
</tr>
</tbody>
</table>

**Step 5: PSNR between perceptual images**

Calculate the PSNR between perceptual images \(f_{\rho(i,j),D}\) and \(\hat{f}_{\rho(i,j),D}\) using Eq. 2 in order to obtain the CIWaM weighted PSNR i.e. the \(P^2\)SNR.

### 2.3 Discussion

Figures 4(c) *Sailboat on Lake* and 5(b) *Splash* \(D_2 = D_1 = 129\,cm\), but subjective quality of *Splash* is clearly better than the one of *Sailboat on Lake*. Thus, even when CIWaM versions of *Splash* and *Sailboat on Lake* are calculated at 129cm, the resultant perceptual images have different objective quality. Hence \(P^2\)SNR predicts that the error in Figure 5(b) is twice less (\(\sim 3dB\)) than in Figure 4(c).

That is why overall \(P^2\)SNR algorithm is the estimation of the objective quality taking into account the set of the interactions of parameters \(nP, d\) and \(D\). Figures 5 and 6 show examples when perceptual quality is equal and their respective points \((nP, \varepsilon R \,(nP))\) do not correspond. In Figure 5, there is a difference of 6cm between \(D_1\) and \(D_2\), while in Figure 6, there is no difference between distances \(D_1\) and \(D_2\).

![Relative Energy Chart of Images Tiffany and Sailboat on Lake](image)

*Fig. 4*

Relative energy chart of images Tiffany and Sailboat on Lake (a), which are distorted by means of JPEG2000 PSNR=31dB and observation distance \(d=120\,cm\). Perceptual quality \(P^2\)SNR is equal to 34.82dB for (b) and 36.77dB for (c).

### 3. Experimental Results

In this section, \(P^2\)SNR performance is assessed by comparing the statistical significance with the psychophysical results obtained by human observers when judging the visual quality of an specific image. These results are expressed in Mean Opinion Scores. In this way, perceived image quality predicted by \(P^2\)SNR is tested for JPEG2000 distortion...
across the Tampere Image Database (TID2008) of the Tampere University of Technology, presented by Ponomarenko et.al in [9], [10].

TID2008 Database contains 25 original images (Figure 7), which are distorted by 17 different types of distortions, each distortion has 4 degrees of intensity, that is, 68 versions of each source image. TID2008 also supplies subjective ratings by comparing original and distorted images by 654 observers from Italy, Finland and Ukraine. Thus, for JPEG and JPEG2000 compression distortions, there are 200 (25 images × 2 distortions × 4 distortion degrees) images in the database. MOS is presented as the global rating.

3.1 Performance Measures

Strength of Relationship (SR) is measured by a correlation coefficient. SR means how strong is the tendency of two variables to move in the same (opposite) direction. Pearson Correlation Coefficient (PCC) is the most common measure for predicting SR, when parametric data are used. But in the case of the correlation of non-parametric data the most common indicator is Spearman Rank-Order Correlation Coefficient (SROCC). Results of image quality metrics have no lineal relationship, which is why, it is not convenient to employ PCC, since even PSNR and MSE are the same metrics, PCC calculates different values.

Hence SROCC is a better choice for measuring SR between the opinion of observers and the results of a given metrics. However SROCC is appropriate for testing a null hypothesis, but when this null hypothesis is rejected is difficult to interpret[11]. In the other hand, Kendall Rank-Order Correlation Coefficient (KROCC) corrects this problem by reflecting SR between compared variables. Furthermore KROCC estimates how similar are two rank-sets against a same object set. Thus, KROCC is interpreted as the probability to rank in the same order taking into account the number of inversions of pairs of objects for transforming one rank into the other[12]. Which is why, $P^2\text{SNR}$ and the rest of metrics are evaluated using KROCC. One of Limitation of KROCC is located in complexity of the algorithm, which takes more computing time than PCC and SROCC, but
KROCC can show us an accurate Strength of Relationship between a metric and the opinion of a human observer. MSE[13], PSNR[13], SSIM[14], MSSIM[15], VIF[16], VIFP[14], IFC[17] and WSNR[18] are compared against the performance of $P^2$SNR for JPEG2000 compression distortion. I chose for evaluating these assessments the implementation provided in [19], since it is based on the parameters proposed by the author of each indicator.

$P^2$SNR is implemented assuming the following features:

- Observation Distance, $d=8H$, where $H$ is the height of a $512 \times 512$ image.
- 19” LCD monitor with horizontal resolution of 1280 pixels and 1024 pixels of vertical resolution.
- Gamma correction, $\gamma = 2.2$
- Wavelet Transform, set of wavelet planes $\omega$ with $n = 3$, Eq. 3.

### 3.2 Overall Performance

Table 3 shows the performance of $P^2$SNR and the other eight image quality assessments across the set of images from TID2008 image database employing KROCC for testing the distortion produced by a JPEG2000 compression.

Thus, for JPEG2000 compression distortion, $P^2$SNR is also the best metrics for each database. $P^2$SNR gets its better results when correlation is 0.8718 for a corpus of 100 images of the TID2008 database. For this distortion, MSSIM is the second best indicator. Furthermore $P^2$SNR improves 0.2336 the perceptual functioning of PSNR when this metrics compares perceptual images in a dynamic way.

### 4. Conclusions

In this work, I presented a new metrics of full-reference image quality based on perceptual weighting of PSNR by means of a perceptual model. The $P^2$SNR metrics is based on the measurement the objective quality of perceptual images predicted by CIWaM at $D$.

$P^2$SNR was tested in TID2008 image database, over viewing distances proposed in this image database. Results show that $P^2$SNR significantly increases the correlation of PSNR with perceived image quality, maintaining its advantageous features, in addition to be the best-ranked image quality gauge in overall performance, in comparison to a diversity of existing metrics. Accuracy of Second best-performing algorithm, MSSIM, is 1.5% lower than $P^2$SNR forJPEG2000 distortion. While when CIWaM weights PSNR correlation of predicting subjective ratings either of PSNR or MSE improves the results by 23.36% for the same kind of distortions, on the average.

$P^2$SNR is mainly developed for optimizing the perceptual error under the constraint of a limited bit-budget, but it contains another properties that can be used for quantizing, since the CIWaM algorithm calculates one value of an extended contrast sensitivity function by pixel. In this way, it is possible to quantize a particular pixel while an algorithm of bit allocation is working, incorporating into embedded compression schemes such as EZW[20], SPIHT[21], JPEG2000[22] or Hi-SET[23].

### Acknowledgment

This work is supported by the Ph.D grants 207950 and SeAca/COTEPABE/PL-255/10 of the Mexican Science and Technology National Council (CONACyT) and the National Polytechnic Institute of Mexico (IPN), respectively.
References


