

$\mathcal{B}$ PSNR

Blind Peak Signal-to-Noise Ratio

Jesús Jaime Moreno Escobar and Beatriz Adriana Jaime Fonseca

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

Abstract. The aim of this work is to define a no-referenced perceptual image quality estimator applying the perceptual concepts of the Chromatic Induction Model. The approach consists in comparing the received image, presumably degraded, against the perceptual versions (different distances) of this image degraded by means of a Model of Chromatic Induction, which uses some of the human visual system properties. Also we compare our model with a original estimator in image quality assessment, PSNR. Results are highly correlated with the ones obtained by PSNR but this proposal does not need an original image or a reference one in order to give an estimation of the quality of the degraded image.

Keywords: Human Visual System, Contrast Sensitivity Function, Perceived Images, Wavelet Transform, Peak Signal-to-Noise Ratio, No-Reference Image Quality Assessment.

1 Introduction

The early years of the 21st century have witnessed a tremendous growth in the use of digital images as a means for representing and communicating information. A significant literature describing sophisticated theories, algorithms, and applications of digital image processing and communication has evolved. A considerable percentage of this literature is devoted to methods for improving the appearance of images, or for maintaining the appearance of images that are processed. Nevertheless, the quality of digital images, processed or otherwise, is rarely perfect. Images are subject to distortions during acquisition, compression, transmission, processing, and reproduction. To maintain, control, and enhance the quality of images, it is important for image acquisition, management, communication, and processing systems to be able to identify and quantify image quality degradations. The development of effective automatic image quality assessment systems is a necessary goal for this purpose. Yet, until recently, the field of image quality assessment has remained in a nascent state, awaiting new models of human vision and of natural image structure and statistics before meaningful progress could be made.

Nowadays, 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. But some authors like Wang and Bovik in [1, 2] consider that MSE is a poor algorithm, 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:

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

and

$$PSNR = 10 \log_{10} \left(\frac{\mathcal{G}_{max}^2}{MSE} \right) \quad (2)$$

where $\mathcal{G}_{max}$ 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) $\mathcal{G}_{max} = 2^8 - 1 = 255$. 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.

An important task in image compression systems is to maximize the correlation among pixels, because the higher correlation at the preprocessing, the more efficient algorithm postprocessing. Thus, an efficient measure of image quality should take in to account the latter feature. In contrast to this, MSE does not need any positional information of 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 [3, 4] and Region of interest [5, 4]. 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.

2 Image Quality Assessment

2.1 Full Reference (FR)

Bottom-Up Approaches Psychological and physiological studies in the past century have gained us a tremendous amount of knowledge about the human visual system (HVS). Still, although much is known about the mechanisms of

early, front-end vision, much more remains to be learned of the later visual pathways and the general higher level functions of the visual cortex. While the knowledge is far from complete, current models of visual information processing mechanisms have become sufficiently sophisticated that it is of interest to explore whether it is possible to deploy them to predict the performance of simple human visual behaviors, such as image quality evaluation.

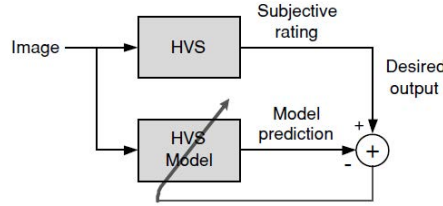
Bottom up approaches to image quality assessment are those methods that attempt to simulate well modeled functionalities of the HVS, and integrate these in the design of quality assessment algorithms that, hopefully, perform similar to the HVS in the assessment of image quality. In this chapter we begin with a brief description of relevant aspects of the anatomy and psychophysical features of the HVS. This description will focus on those HVS features that contribute to current engineering implementations of perceptual image quality measures.

Most systems that attempt to incorporate knowledge about the HVS into the design of image quality measures use an error sensitivity framework, so that the errors between the distorted image and reference image are perceptually quantized according to HVS characteristics.

Top-Down Approaches The bottom-up approaches to image quality assessment described in the last subsection (2.1) attempt to simulate the functional components in the human visual system that may be relevant to image quality assessment. The underlying goal is to build systems that work in the same way as the HVS, at least for image quality assessment tasks. By contrast, the top-down systems simulate the HVS in a different way. These systems treat the HVS as a black box, and only the input output relationship is of concern. A top-down image quality assessment system may operate in a manner quite different from that of the HVS, which is of little concern, provided that it successfully predicts the image quality assessment behavior of an average human observer.

One obvious approach to building such a top-down system is to formulate it as a supervised machine learning problem, as illustrated in Fig. 1. Here the HVS is treated as a black box whose input/output relationship is to be learned. The training data can be obtained by subjective experimentation, where a large number of test images are viewed and rated by human subjects. The goal is to train the system model so that the error between the desired output (subjective rating) and the model prediction is minimized. This is generally a regression or function approximation problem. Many techniques are available to attack these kinds of problems.

Unfortunately, direct application of this method is problematic, since the dimension of the space of all images is the same as the number of pixels in the image. Furthermore, subjective testing is expensive and a typical extensive subjective experiment would be able to include only several hundred test images—hardly an adequate coverage of the image space. Assigning only a single sample at each quadrant of a ten dimensional space requires a total of 1024 samples, and the dimension of the image space is in the order of thousands to millions. An excellent example of the problem of dimensionality.

**Fig. 1.** Learning HVS.

One method that might be useful to overcome this problem is by dimension reduction. The idea is to map the entire image space onto a space of much lower dimensionality by exploiting knowledge of the statistical distribution of typical images in the image space. Since natural images have been found to exhibit strong statistical regularities, it is possible that the cluster of typical natural images may be represented by a low dimensional manifold, thus reducing the number of sample images that might be needed in the subjective experiments. However, dimension reduction is no trivial task. Indeed, no dimension reduction technique has been developed to reduce the dimension of natural images to 10 or less (otherwise, extremely efficient image compression techniques would have been proposed on the basis of such reduction). Consequently, using a dimension reduction approach for general purpose image quality assessment remains quite difficult. Nonetheless, such an approach may prove quite effective in the design of application specific quality assessment systems, where the types of distortions are fixed and known and may be described by a small number of parameters.

2.2 No-Reference

No-reference (NR) image quality assessment is, perhaps, the most difficult (yet conceptually simple) problem in the field of image analysis. By some means, an objective model must evaluate the quality of any given real world image, without referring to an original high quality image. On the surface, this seems to be a mission impossible. How can the quality of an image be quantitatively judged without having a numerical model of what a good/bad quality image is supposed to look like? Yet, amazingly, this is quite an easy task for human observers. Humans can easily identify high quality images versus low quality images, and, furthermore, they are able to point out what is right and wrong about them without seeing the original. Moreover, humans tend to agree with each other to a pretty high extent. For example, without looking at the original image, probably every reader would agree that the noisy, blurry, and JPEG2000 compressed images in Fig. 2 have lower quality than the luminance shifted and contrast stretched images.

Before developing any algorithm for image quality assessment, a fundamental question that must be answered is what source of information can be used to evaluate the quality of images. Clearly, the human eye/brain system is making use

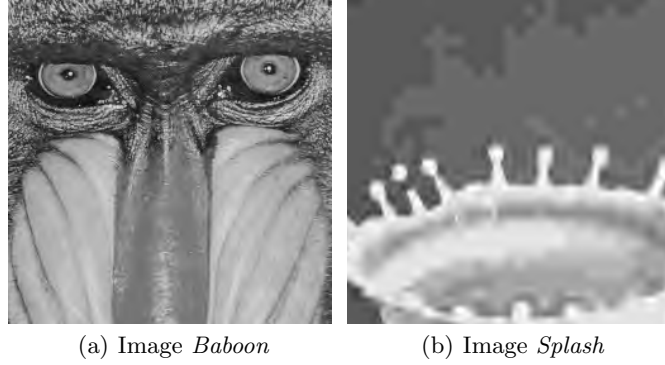


Fig. 2. 256×256 patches (cropped for visibility) of Images *Baboon* and *Splash* distorted by means of JPEG2000 compression, although both images have the same objective quality (PSNR=30dB), their visual quality is very different.

of a very substantial and effective pool of information about images in making subjective judgments of image quality.

Three types of knowledge may be employed in the design of image quality measures: knowledge about the original high quality image, knowledge about the distortion process, and knowledge about the human visual system (HVS). In FR quality assessment, the high quality original image is known a priori. In NR quality assessment, however, the original image is absent, yet one can still assume that there exists a high quality original image, of which the image being evaluated is a distorted representation. It is also reasonable to make a further assumption that such a conjectured original image belongs to the set of typical natural images.

It is important to realize that the cluster of natural images occupies an extremely tiny portion in the space of all possible images. This potentially provides a strong prior knowledge about what these images should look like. Such prior knowledge could be a precious source of information for the design of image quality measures. Models of such natural scenes attempt to describe the class of high quality original images statistically. Interestingly, it has been long conjectured in computational neuroscience that the HVS is highly adapted to the natural visual environment, and that, therefore, the modeling of natural scenes and the HVS are dual problems.

Knowledge about the possible distortion processes is another important information source that can be used for the development of NR image quality measures. For example, it is known that blur and noise are often introduced in image acquisition and display systems and reasonably accurate models are sometimes available to account for these distortions. Images compressed using block based algorithms such as JPEG often exhibit highly visible and undesirable blocking artifacts. Wavelet based image compression algorithms operating at low bit rates can blur images and produce ringing artifacts near discontinuities.

Of course, all of these types of distortions are application dependent. An application specific NR image quality assessment system is one that is specifically

designed to handle a specific artifact type, and that is unlikely to be able to handle other types of distortions. The question arises, of course, whether an application specific NR system is truly reference free, since much information about the distorted image is assumed. However, nothing needs to be assumed about the original image, other than, perhaps models derived from natural scene statistics or other natural assumptions. Since the original images are otherwise unknown, we shall continue to refer to more directed problems such as these as application specific NR image quality assessment problems.

Of course, a more complex system that includes several modes of artifact handling might be constructed and that could be regarded as approaching general purpose NR image quality assessment. Before this can happen, however, the various components need to be designed. Fortunately, in many practical application environments, the distortion processes involved are known and fixed. The design of such application specific NR quality assessment systems appears to be much more approachable than the general, assumption free NR image quality assessment problem. Very little, if any, meaningful progress has been made on this latter problem.

Owing to a paucity of progress in other application specific areas, this chapter mainly focuses on NR image quality assessment methods, which are designed for assessing the quality of compressed images. In particular, attention is given to a spatial domain method and a frequency domain method for block based image compression, and a wavelet domain method for wavelet based image compression.

3 The $\mathcal{B}$ PSNR Algorithm

3.1 Chromatic Induction Wavelet Model

The *Chromatic Induction Wavelet Model* (CIWaM) [6] is a low-level perceptual model of the HVS. It estimates the image perceived by an observer at a distance d just by modeling the perceptual chromatic induction processes of the HVS. That is, given an image $\mathcal{I}$ and an observation distance d , CIWaM obtains an estimation of the perceptual image $\mathcal{I}_p$ that the observer perceives when observing $\mathcal{I}$ at distance d . CIWaM is based on just three important stimulus properties: spatial frequency, spatial orientation and surround contrast. These three properties allow to unify the chromatic assimilation and contrast phenomena, as well as some other perceptual processes such as saliency perceptual processes [7].

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

$$\mathcal{I} = \sum_{s=1}^n \sum_{o=v,h,dgl} \omega_{s,o} + c_n, \quad (3)$$

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*.

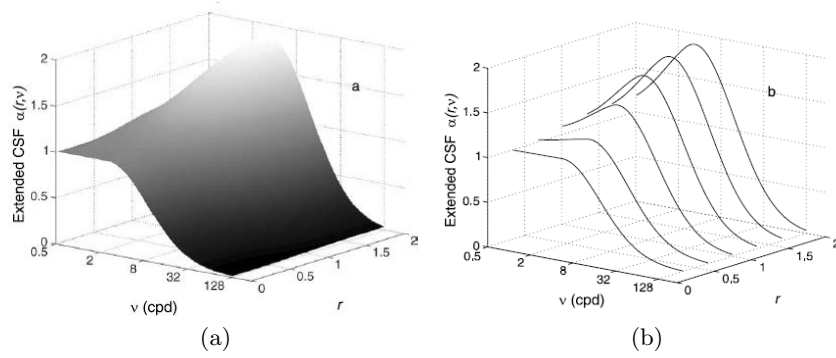


Fig. 3. (a) Graphical representation of the e-CSF ($\alpha_{s,o,i}(r, \nu)$) for the luminance channel. (b) Some profiles of the same surface along the Spatial Frequency (ν) axis for different centersurround contrast energy ratio values (r). The psychophysically measured CSF is a particular case of this family of curves (concretely for $r = 1$).

The perceptual image $\mathcal{I}_p$ is recovered by weighting these $\omega_{s,o}$ wavelet coefficients using the *extended Contrast Sensitivity Function* (e-CSF, Fig. 3). The e-CSF is an extension of the psychophysical CSF [8] considering spatial surround information (denoted by r), visual frequency (denoted by ν , which is related to spatial frequency by observation distance) and observation distance (d). Perceptual image $\mathcal{I}_p$ can be obtained by

$$\mathcal{I}_p = \sum_{s=1}^n \sum_{o=v,h,dgl} \alpha(\nu, r) \omega_{s,o} + c_n, \quad (4)$$

where $\alpha(\nu, r)$ is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term $\alpha(\nu, r) \omega_{s,o} \equiv \omega_{s,o;\rho,d}$ can be considered the *perceptual wavelet coefficients* of image $\mathcal{I}$ when observed at distance d and is written as:

$$\alpha(\nu, r) = z_{ctr} \cdot C_d(\dot{s}) + C_{min}(\dot{s}). \quad (5)$$

This function has a shape similar to the e-CSF and the three terms that describe it are defined as:

z_{ctr} Non-linear function and estimation of the central feature contrast relative to its surround contrast, oscillating from zero to one, defined by:

$$z_{ctr} = \frac{\left[\frac{\sigma_{cen}}{\sigma_{sur}} \right]^2}{1 + \left[\frac{\sigma_{cen}}{\sigma_{sur}} \right]^2} \quad (6)$$

being σ_{cen} and σ_{sur} the standard deviation of the wavelet coefficients in two concentric rings, which represent a center-surround interaction around each coefficient.

$C_d(\dot{s})$ Weighting function that approximates to the perceptual e-CSF, emulates some perceptual properties and is defined as a piecewise Gaussian function [8], such as:

$$C_d(\dot{s}) = \begin{cases} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ e^{-\frac{\dot{s}^2}{2\sigma_2^2}}, & \dot{s} = s - s_{thr} > 0. \end{cases} \quad (7)$$

$C_{min}(\dot{s})$ Term that avoids $\alpha(\nu, r)$ function to be zero and is defined by:

$$C_{min}(\dot{s}) = \begin{cases} \frac{1}{2} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ \frac{1}{2}, & \dot{s} = s - s_{thr} > 0. \end{cases} \quad (8)$$

taking $\sigma_1 = 2$ and $\sigma_2 = 2\sigma_1$. Both $C_{min}(\dot{s})$ and $C_d(\dot{s})$ depend on the factor s_{thr} , which is the scale associated to 4 cycles per degree when an image is observed from the distance d with a pixel size l_p and one visual degree, whose expression is defined by Equation 9. Where s_{thr} value is associated to the e-CSF maximum value.

$$s_{thr} = \log_2 \left(\frac{d \tan(1^\circ)}{4 l_p} \right) \quad (9)$$

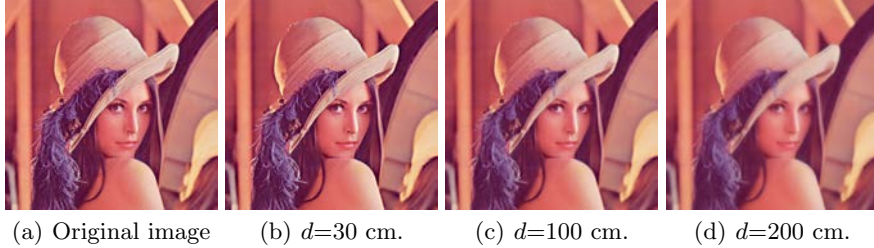


Fig. 4. (a) Original color image *Lenna*. (b)-(d) Perceptual images obtained by CIWaM at different observation distances d .

Fig. 4 shows three examples of CIWaM images of *Lenna*, calculated by Eq. 4 for a 19 inch monitor with 1280 pixels of horizontal resolution, at $d = \{30, 100, 200\}$ centimeters.

3.2 Basics

In the no-referenced image quality issue, there is only a distorted version $\hat{f}(i, j) = \Lambda[f(i, j)]$ that is compared with $f(i, j)$, being Λ a distortion model and the unknown original image $f(i, j)$ is considered a pattern $\mathcal{T}([0, 1; 1, 0])$ like a chessboard (Figs. 5) with the same size of $\hat{f}(i, j)$. The difference between these two images depends on the features of the distortion model Λ . For example, blurring, contrast change, noise, JPEG blocking or JPEG2000 wavelet ringing.

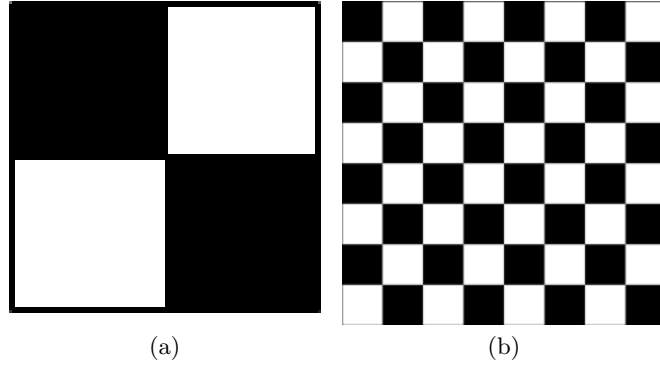


Fig. 5. (a) Pattern $[0,1;1,0]$ or $\mathcal{T}$. (b) Pattern $\mathcal{T}$ repeated sixteen times.

In Fig. 2, the images *Baboon* 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 A 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. 6 shows the diagonal high spatial frequencies of these two images, where there are more high frequencies in image *Baboon*.

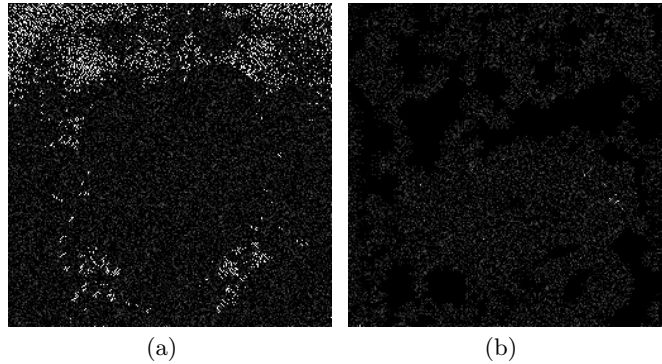


Fig. 6. Diagonal spatial orientation of the first wavelet plane of Images (a) *Baboon* and (b) *Splash* distorted by JPEG2000 with PSNR=30dB.

If a set of distortions $\hat{f}_k(i, j) = \Lambda_k[f(i, j)]$ is generated and indexed by k (for example, let Λ 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 slowly decreases.

Similarly, the HVS is a system that induces a distortion on the observed image $\hat{f}(i, j)$, whose model is predicted by CIWaM. Hence, CIWaM is considered a HSV particular distortion model $\Lambda \equiv \text{CIWaM}$ that generates a perceptual image $\hat{f}_\rho(i, j) \equiv \mathcal{I}_\rho$ from an observed image $f(i, j) \equiv \mathcal{I}$, i.e. $\mathcal{I}_\rho = \text{CIWaM}[\mathcal{I}]$. 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 is defined $\mathcal{I}_{\rho,d} = \text{CIWaM}_d[\mathcal{I}]$ which is considered a set of perceptual distortions of the hypothetical image $\mathcal{I}$.

When image $\hat{f}(i, j)$ is observed at distance $\bar{d}$ and this distance is reduced, the artifacts, if this possesses, are better perceived. In contrast, $\hat{f}(i, j)$ is observed from a far distance human eyes cannot perceive their artifacts, in consequence, the perceptual image quality of the distorted image is always high. The distance where the observer can perceive the best image quality of image $\hat{f}(i, j)$ is considered as the distance D .

3.3 Methodology

Let $f(i, j)$ and $\hat{f}(i, j) = \Lambda[f(i, j)]$ be an pattern image and a distorted image, respectively. $\mathcal{BPSNR}$ 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 $\hat{f}(i, j)$, when an observer observe them at d centimeters of observation distance. So measuring the PSNR of $\hat{f}(i, j)$ at D will yield a fairer and blind perceptual evaluation of its image quality.

$\mathcal{BPSNR}$ algorithm is divided in five steps, which is summarized by the Figure 7 and described as follows:

Step 1: Wavelet Transformation Forward wavelet transform of images $f(i, j)$ and $\hat{f}(i, j)$ is performed using Eq. 3, obtaining the sets $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$, respectively. The employed analysis filter is the Daubechies 9-tap/7-tap filter (Table 1).

Table 1. 9/7 Analysis Filter.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	0.6029490182363579	1.115087052456994
± 1	0.2668641184428723	-0.5912717631142470
± 2	-0.07822326652898785	-0.05754352622849957
± 3	-0.01686411844287495	0.09127176311424948
± 4	0.02674875741080976	

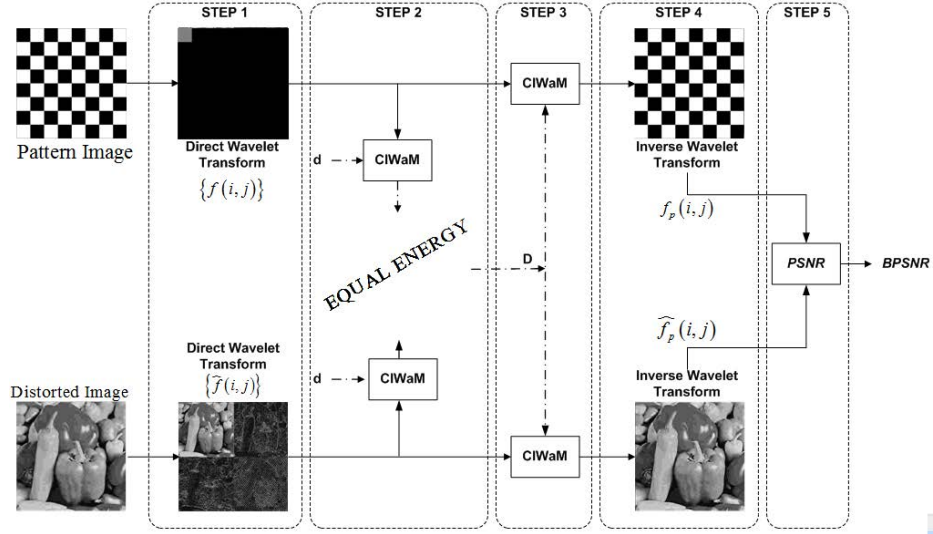


Fig. 7. Methodology for No-Reference PSNR weighting by means of CIWaM. Both Pattern and Distorted images are wavelet transformed. The distance D where the energy of perceptual images obtained by CIWaM are equal is found. Then, PSNR of perceptual images at D is calculated, obtaining the $\mathcal{B}$ PSNR metrics.

Step 2: Distance D The total energy measure or the *deviation signature*[9] $\bar{\varepsilon}$ is the absolute sum of the wavelet coefficient magnitudes, defined by [10]

$$\bar{\varepsilon} = \sum_{n=1}^N \sum_{m=1}^M |x(m, n)| \quad (10)$$

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. 10, since one wCal is the energy needed to increase the absolute magnitude of a wavelet coefficient by one scale.

From wavelet coefficients $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$ the corresponding perceptual wavelet coefficients $\{\omega_{s,o;\rho,\tilde{d}}\} = \alpha(\nu, r) \cdot \omega_{s,o}$ and $\{\hat{\omega}_{s,o;\rho,\tilde{d}}\} = \alpha(\nu, r) \cdot \hat{\omega}_{s,o}$ are obtained by applying CIWaM with an observation distance $\tilde{d}$. Therefore, Equation 11 expresses the relative wavelet energy ratio $\varepsilon\mathcal{R}(\tilde{d})$, which compares how different are the energies of the reference and distorted CIWaM perceptual images, namely ε_ρ and $\hat{\varepsilon}_\rho$ respectively, when these images are watched from a given distance $\tilde{d}$.

$$\varepsilon\mathcal{R}(\tilde{d}) = 10 \cdot \left| \log_{10} \frac{\varepsilon_\rho(\tilde{d})}{\hat{\varepsilon}_\rho(\tilde{d})} \right| \quad (11)$$

Thus, the main goal of this step is to find $\varepsilon\mathcal{R}(D)$, namely, at D ε_ρ is equal to $\widehat{\varepsilon}_\rho$, where the energy of the distorted images are the same than the energy of the pattern.

Step 3: Perceptual Images Getting the perceptual images $\{f_p(i, j)\}$ and $\{\hat{f}_p(i, j)\}$ from the $\{f(i, j)\}$ and $\{\hat{f}(i, j)\}$ images watched at D centimeters, using Equation 4.

Step 4: Inverse Wavelet Transformation Perform the Inverse Wavelet Transform of $\{\omega_{s,o;\rho,D}\}$ and $\{\hat{\omega}_{s,o;\rho,D}\}$, 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 2. 9/7 Synthesis Filter.

Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1.115087052456994	0.6029490182363579
± 1	0.5912717631142470	-0.2668641184428723
± 2	-0.05754352622849957	-0.07822326652898785
± 3	-0.09127176311424948	0.01686411844287495
± 4		0.02674875741080976

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 $\mathcal{B}$ PSNR.

4 Experimental Results

It is important to mention that $\mathcal{B}$ PSNR estimates the degradation, thus, the smaller the better. In this section, $\mathcal{B}$ PSNR performance is assessed by comparing the statistical significance of the images *Lenna* and *Baboon*, in addition to the Pearson correlation between $\mathcal{B}$ PSNR and PSNR data.

Figure 8 depicts three JPEG2000 distorted versions of the image *Lenna* with 0.05(8(a)), 0.50 (8(b)) and 1.00 (8(c)) bits per pixel. PSNR estimates 23.41, 32.74 and 34.96 dB, respectively. While $\mathcal{B}$ PSNR computes 48.42, 36.56 and 35.95 dB, respectively. Thus, both PSNR and $\mathcal{B}$ PSNR estimate that image at 1.00 bpp has lower distortion.

When this experiment is extended computing the JPEG2000 distorted versions from 0.05 bpp to 3.00bpp (increments of 0.05 bpp, depicted at Figure 9), we found that the correlation between PSNR and $\mathcal{B}$ PSNR is 99.32 %, namely for image *Lenna* for every 10,000 estimation $\mathcal{B}$ PSNR misses only in 68 assessments.

Figure 10 depicts three JPEG2000 distorted versions of the image *Baboon* with 0.05(10(a)), 0.50 (10(b)) and 1.00 (10(c)) bits per pixel. PSNR estimates 18.55, 23.05 and 25.11 dB, respectively. While $\mathcal{B}$ PSNR computes 43.49, 30.07 and 28.71 dB, respectively. Thus, both PSNR and $\mathcal{B}$ PSNR estimate that image at 0.05 bpp has higher distortion.



Fig. 8. JPEG2000 Distorted versions of color image *Lenna* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) Medium Distortion and (c) Low Distortion.

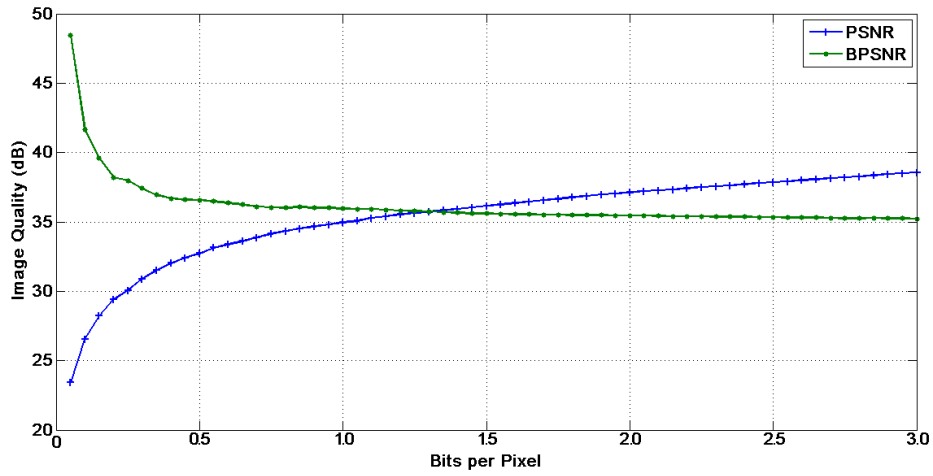


Fig. 9. Comparison of PSNR and $\mathcal{B}$ PSNR for the JPEG2000 distorted versions of image *Lenna*.

When this experiment is extended computing the JPEG2000 distorted versions from 0.05 bpp to 3.00bpp (increments of 0.05 bpp, depicted at Figure 11), we found that the correlation between PSNR and $\mathcal{B}$ PSNR is 96.95 %, namely for image *Baboon* for every 10,000 estimation $\mathcal{B}$ PSNR misses only in 305 assessments.

5 Conclusions

$\mathcal{B}$ PSNR is a new metric for no-reference or blind image quality based on perceptual weighting of PSNR by using a perceptual low-level model of the Human Visual System (CIWaM model). The proposed $\mathcal{B}$ PSNR metrics is based on five steps.

The $\mathcal{B}$ PSNR assessment was tested in two well-known images, such as *Lenna* and *Baboon*. It is a well-correlated image quality method in these images for

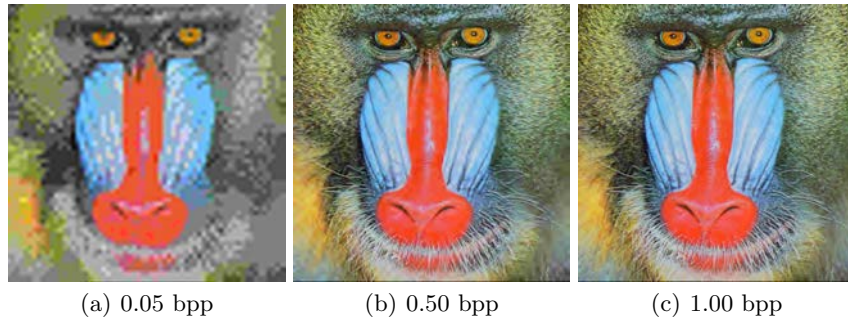


Fig. 10. JPEG2000 Distorted versions of color image *Baboon* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) Medium Distortion and (c) Low Distortion.

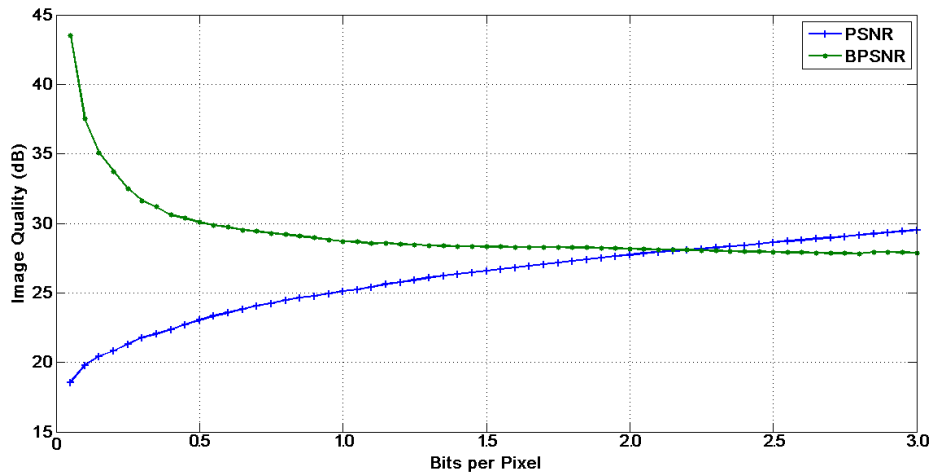


Fig. 11. Comparison of PSNR and β PSNR for the JPEG2000 distorted versions of image *Baboon*.

JPEG2000 distortions when compared to PSNR. Concretely, β PSNR correlates with PSNR, on the average in 98.13%.

Acknowledgment

This work is supported by The National Polytechnic Institute of Mexico by means of a granted fund by the Committee of Operation and Promotion of Academic Activities (COFAA).

References

- [1] Wang, Z., Bovik, A.: Mean squared error: Love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE* **26**(1) (jan. 2009) 98 –117

- [2] Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. 1 edn. Morgan & Claypool Publishers: Synthesis Lectures on Image, Video, & Multimedia Processing (February 2006)
- [3] Auli-Llinas, F., Serra-Sagrista, J.: Low complexity JPEG2000 rate control through reverse subband scanning order and coding passes concatenation. *IEEE Signal Processing Letters* **14**(4) (april 2007) 251 –254
- [4] Taubman, D.S., Marcellin, M.W.: JPEG2000: Image Compression Fundamentals, Standards and Practice. ISBN: 0-7923-7519-X. Kluwer Academic Publishers (2002)
- [5] Bartrina-Rapesta, J., Auli-Llinas, F., Serra-Sagrista, J., Monteagudo-Pereira, J.: JPEG2000 Arbitrary ROI coding through rate-distortion optimization techniques. In: Data Compression Conference. (25-27 2008) 292 –301
- [6] Otazu, X., Párraga, C., Vanrell, M.: Toward a unified chromatic induction model. *Journal of Vision* **10**(12)(6) (2010)
- [7] Murray, N., Vanrell, M., Otazu, X., Parraga, A.: Saliency estimation using a non-parametric low-level vision model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011). (2010) 433 –440
- [8] Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology* **359** (February 1985) 381–400
- [9] van de Wouwer, G., Scheunders, P., van Dyck, D.: Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing* **8**(4) (April 1999) 592 –598
- [10] Wilson, B.A., Bayoumi, M.A.: A computational kernel for fast and efficient compressed-domain calculations of wavelet subband energies. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* **50**(7) (July 2003) 389 – 392