AN ALGORITHM FOR IMAGE QUALITY ASSESSMENT

Goran Ivkovic and Ravi Sankar
Department of Electrical Engineering
University of South Florida, Tampa, Florida 33620

ABSTRACT

This paper presents a new novel algorithm for image quality assessment. First, a simple model of human visual system, consisting of a nonlinear function and a 2-D filter, processes the input images. This filter has one user-defined parameter, whose value depends on the reference image. In the next step the average value of locally computed correlation coefficients between the two processed images is found. This criterion is closely related to the way in which human observer assesses image quality. In the last step image quality measure is computed as the average value of locally computed correlation coefficients, adjusted by average correlation coefficient between the reference image and error image. This way the proposed measure differentiates between the random and signal-dependant distortion, which have different effects on human observer. Performance of the proposed quality measure is illustrated by examples involving images with different types of degradation.

1. INTRODUCTION

Image quality assessment methods can be divided into: subjective and objective. The only subjective (qualitative) measure is mean opinion score (MOS). Since human observer is the ultimate receiver of the information contained in an image, this is the best way for image quality assessment. Objective (quantitative) measures use intensities from two input images (reference and distorted image) to compute a number indicating image quality. Most widely used objective measures are mean squared error (MSE) and MSE-based measures: peak signal to noise ratio (PSNR) and signal to noise ratio (SNR). These simple measures work well when images with the same type of degradation are compared. In this case distorted image with smaller MSE will be perceived closer to the original image than the one with greater MSE. However, when images with different types of degradation are compared MSE does not produce results that correlate well with subjective quality assessment. Images with different types of degradations with the same MSE values can have very different subjective visual qualities.

In order to find a criterion, which agrees with subjective assessment, several other algorithms have been developed. These algorithms try to assess image quality by taking into account the properties of human visual system (HVS). Most of these algorithms are trying to model the following three properties of HVS: nonlinear relationship between image intensities and perceived brightness, frequency response of HVS (contrast sensitivity function) and texture masking [1-3].

An overview of image quality measures is presented in [4]. Performance of several proposed quality measures is tested in [5]. The tested measures included MSE and the measure proposed in [2]. The quality measure from [2] showed best agreement with subjective evaluation compared with other tested measures. However, none of the tested measures was able to predict subjective image quality consistently. Problems with the types of image quality measures similar to those described in [1-3] are that there are no standards for modeling the same properties of HVS. For example, models for brightness perception in [1] and [2] are totally different. Also, there are various contrast sensitivity functions estimated by different authors. Another problem is computation of perceptual threshold, which different authors compute in a totally different way. This sometimes requires computation of local contrast (as in [1] and [3]), which is difficult to define for complex images.

A different image quality measure is proposed in [6] and [7]. This measure is computed locally and it is defined as a product of three components. Most important of these components is correlation coefficient, which measures the degree of linear relationship between the corresponding blocks of pixels. Quality measure for the whole image is the average value of locally computed quality measures. Most other approaches try to modify MSE by modeling properties of HVS. This approach defines different criterion which is better related to the way in which human observer assesses image quality. No HVS model is used here.

The approach presented here will use the basic idea from [6] and [7], but it will not use image quality measure as it is defined there (only correlation coefficient will be used). In the first step reference image and distorted image are processed by simple model of HVS consisting of a nonlinear function modeling brightness perception and a 2-D filter modeling frequency response of HVS (contrast sensitivity function). The 2-D filter used in this HVS model is not fixed. It contains a user-defined parameter that can be changed depending on the content of the reference image. After this, the correlation coefficient is computed on a block-by-block basis for the processed input images. Finally, quality measure is computed as the average correlation coefficient between the reference image and distorted image, adjusted by average correlation coefficient between the reference image and error image.

Rest of the paper is organized as follows. Problems with MSE-based image quality measures are discussed in Section 2. HVS model used to process input images is presented in Section 3. Proposed image quality measure is defined next. Performance of the proposed measure will be illustrated by examples involving images with different types of distortion in Section 5, followed by the conclusion.

2. MSE AS IMAGE QUALITY MEASURE

MSE and MSE-based measures such as PSNR and SNR work when images with the same type of degradation are compared. In a sequence of images obtained by distorting an image by various amounts of the same type of degradation, the image with the smallest MSE will be perceived by a human observer to be the closest to the original image. However, when images with different types of degradation are compared, the image with the smallest MSE will not always be perceived to be the closest to the original image. It is possible to create images with the same MSE value but with totally different visual quality. It is well understood that MSE is not a very effective image quality measure due to three main reasons for disagreement between MSE-based and subjective evaluation.

First reason for this is the presence of spatial filtering in HVS. This can be demonstrated by the adding white Gaussian filtered noise to the original image for various values of filter cut-off frequencies, f_{min} and f_{max}. Noise at very low and very high frequencies is less noticeable than bandpass noise indicating bandpass character of HVS. One way to solve this problem is to use weighted MSE, where reference and distorted images are filtered by contrast sensitivity function (CSF) and then MSE is computed using filtered images [8].

Second reason for disagreement between MSE-based and subjective assessments is in the criterion itself, that is, MSE treats image intensities as a set of uncorrelated numbers [6-7]. One pixel is never seen isolated from surrounding pixels and they always create some structure in an image. This information about the structure created by a group of pixels is more important for a human observer than the intensity at some spatial location. However, this is completely ignored by MSE and its computation is not quite related to the way human observers perceive image quality. The images distorted by various noises have same MSE values but very different visual qualities. This problem cannot be solved by some modification of MSE, such as weighted MSE. It is necessary to define a different image quality measure.

Third reason for low correlation between MSE values and subjective visual quality is due to the fact that MSE does not differentiate between random and signal dependent distortion. Random noise is much less objectionable for a human viewer than signal dependant distortion. The results of these experiments are presented in Tables 1-3 in section 5.

3. HVS MODEL

Image quality measure developed here will use a simple model of HVS consisting of a nonlinear function, modeling brightness perception of HVS and a 2-D filter, modeling frequency response of HVS. This type of HVS model was used in [8-9]. However, nonlinear function and 2-D filter used here are not the same as those developed in [8]. Digital images are represented using a finite number of intensity levels. Perceived brightness is a nonlinear function of intensity and this is modeled by transforming input intensities by some nonlinear monotonically increasing function. Various functions have been suggested to model this. One frequently cited result states that brightness perceived by HVS is proportional to logarithm of intensity [4]. HVS model developed in [8] uses cube root function to model this effect. Some authors use brightness perception models developed for specific display device. An example of this can be found in [2]. Logarithmic or cube root functions are not accurate models for brightness perception by HVS especially at low intensities. The following nonlinear function will be used here instead.

$$B=\begin{cases} 0 \quad \text{if } 0 \le I \le I_{TR} \\ \dfrac{B_{MAX}}{2}\left(\dfrac{2(I-I_{TR})}{I_{MAX}-I_{TR}}\right)^2 \quad \text{if } I_{TR} \le I < (I_{MAX}+I_{TR})/2 \\ B_{MAX}-\dfrac{B_{MAX}}{2}\left(\dfrac{2(I_{MAX}-I)}{I_{MAX}-I_{TR}}\right)^2 \quad \text{if } (I_{MAX}+I_{TR})/2 < I \le I_{MAX} \end{cases} \tag{1}$$

where, I_{TR}=20 is threshold value for intensity, I_{MAX}=255 is the maximum value of intensity and B_{MAX}=100 is the maximum value of perceived brightness. This range is chosen for convenience. The role of this function is to emphasize intensities in the mid range and de-emphasize very high and very low intensities, which approximates the brightness perception by HVS.

Second part of HVS model is a 2-D filter, which models frequency response of HVS. HVS is not equally sensitive to all spatial frequencies and it is a frequency selective system. Since we see less noise at very low and very high frequencies than in the mid frequency range, HVS must be a band-pass system. This is modeled by a contrast sensitivity function (CSF), which represents the frequency response of HVS. CSF models HVS as a 2-D filter. Frequency is expressed in cycles/degree instead of cycles/cm, because images can be observed form the various distances and the same spatial frequency expressed in cycles/cm will be perceived differently for different viewing distances.

Various functions have been suggested to model this effect. Most are bandpass in nature but maximum at different frequencies. For example, it is noted in [9] that various functions suggested by different authors have maximum at frequencies ranging from 3 to 10 cycles/degree. HVS model presented in [10] uses a highpass filter. The reasoning behind this is that HVS treats near and far objects in the same way. Since there is no reliable way to tell which one of these functions represents the best model of HVS, none of them will be used here. Instead, CSF given by the following formula will be used:

$$H(f)=\begin{cases} (0.0512+0.8512f)\exp(-(0.3192f)), \\ \qquad \text{for } f \le 3 \text{ cycles/degree} \\ 1, \quad \text{for 3 cycles/degree} < f < f_0 \\ \exp(-0.1(f-f_0)^{1.1}), \quad \text{for } f > f_0 \end{cases} \tag{2}$$

where, f denotes frequency in cycles/degree and f_0 represents the frequency in cycles/degree at which CSF starts to decrease exponentially. This parameter is user-defined and depends on the reference image. If the reference image contains one large object (for example "Lena" image) than f_0 takes lower values than in the case when the reference image contains smaller objects or lot of fine details. Choice of values for this parameter will be discussed in section 5. This requires some user intervention but it yields better results. This function is shown in Fig. 1a for f_0=5 cycles/degree.

This property of HVS is taken into account by filtering input image by CSF. The input image f(m,n) is filtered by 2-D filter $H(f_1,f_2)$, which results in processed image x(m,n). A simple HVS system comprising of a nonlinear function followed by a 2-D filter models the brightness perception and frequency response of HVS. Nonlinear function is given by eqn. (1) and 2-D filter is given by $H(f_1,f_2)=H(\sqrt{f_1^2+f_2^2})$ (assumed radial symmetry), where H(f) is given by eqn. (2). The output signal in this case depends on the viewing distance, the width and the height and the number of pixels of the input image. This is reasonable because image perception by HVS also depends on these parameters.

4. DEFINITION OF IMAGE QUALITY MEASURE

Original or reference image is denoted by f(m,n) and distorted image is denoted by p(m,n). Both images have MxN pixels. In the first step, both images are transformed by the HVS system, which models the brightness perception and frequency selectivity of HVS. The result of the filtering in the HVS model depends on the viewing distance, the width and height and the number of pixels in input images, as it was described in the previous section, so these parameters must also be given. The processing of the input images f(m,n) and p(m,n) by this model will produce images x(m,n) and y(m,n) respectively, which will be used in the computation of the quality measure.

Processed images x(m,n) and y(m,n) can be used to compute weighted MSE (WMSE) as was done in [8]. Instead of this, a different procedure, similar to what was done in [6] and [7], will be used here. Images x(m,n) and y(m,n) obtained after processing the reference image f(m,n) and the distorted image p(m,n) are partitioned into 8x8 pixel blocks. For each of these blocks, the correlation coefficient between the corresponding samples from images x(m,n) and y(m,n) is computed. If we denote the value of the correlation coefficient between x(m,n) and y(m,n) in the i-th block of data by $?_{xy}(i)$ then we can find the average value of the correlation coefficient between the x(m,n) and y(m,n) as:

$$\rho_{xy\_avg} = \frac{1}{L}\sum_{i}^{L} \rho_{xy}(i) \tag{3}$$

where L is the total number of data blocks and the correlation coefficient between two vectors $X = [x_1\ x_2\ \ldots\ x_N]$ and $Y = [y_1\ y_2\ \ldots\ y_N]$ is defined as $\rho_{XY} = \sigma_{XY} / \sigma_X \sigma_Y$ where σ_{XY} is the covariance and σ_x, σ_y are the standard deviations of X and Y.

By computing correlation coefficient locally we force it to concentrate more on the difference in details, which is essential in image quality assessment. This is also closely related to the way in which human observer assesses image quality by scanning an image piece by piece and determining the difference between corresponding blocks of pixels.

Average correlation coefficient is a good indicator of similarity between the reference image and the distorted image. But it does not differentiate between random and signal dependent distortions. Therefore, it will overestimate image degradation in the case of random noise and underestimate it in the case of signal dependent noise. This problem can be eliminated by computing average correlation coefficient between the reference image x(m,n) and the error image e(m,n), which is given by:

$$e(m,n) = x(m,n) - sign\{\rho_{xy\_avg}\}\, y(m,n) \tag{4}$$

Then local correlation coefficients are computed block by block between images x(m,n) and e(m,n). If the local correlation coefficient for the i-th block of data is denoted by $?_{xe}(i)$, the average correlation coefficient, ρ_{xe_avg} between images x(m,n) and e(m,n) can be computed the same way as in eqn. (3). This coefficient will have very small positive or negative value close to zero, when there is little or no correlation between the images x(m,n) and e(m,n) and this will indicate random noise. On the other hand if there is a significant correlation between x(m,n) and e(m,n) the average correlation coefficient will have higher absolute value, indicating signal dependent noise. Finally, image quality measure denoted by Q is defined as:

$$Q = sign\{\rho_{xy\_avg}\}\,|\rho_{xy\_avg}|^{f(\rho_{xe\_avg})}$$

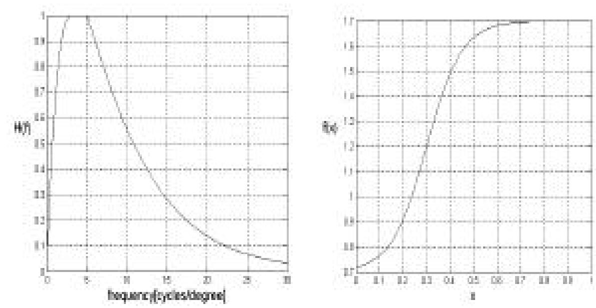
where

$$f(\rho_{xe\_avg}) = 1.2 + 0.5\tanh\left(\frac{|\rho_{xe\_avg}| - 0.3}{0.15}\right) \tag{5}$$

Quality measure (Q) given in eqn. (5) will have the same sign as ρ_{xy_avg}, but its magnitude will be modified value of the magnitude of the average correlation coefficient. The way in which this value is modified depends on the average correlation coefficient, ρ_{xe_avg}. This modification is obtained using function f(ρ_{xe_avg}) and is shown in Fig. 1b. If the magnitude of ρ_{xe_avg} is small (close to zero), then it indicates random noise. If the value of this coefficient is close to one, then it indicates signal dependant noise. When the value of this coefficient is somewhere in between that indicates transition between random and signal dependant noise. If an image is corrupted by random noise the magnitude of the quality measure will be higher than the magnitude of ρ_{xy_avg} and lower if the image is corrupted by signal dependent noise, with transition period in between. This is in accordance with the previous discussion on effects of random and signal dependent noise. The values of the various constants that appear in function f(x) can be used to give different weight to random and signal dependent distortion. Their values can be chosen according to the personal preferences of user.

Finally, the described algorithm can be summarized in the following four steps:

1. Given the reference image f(m,n), distorted image p(m,n), width, height and number of pixels of the input images and viewing distance, compute images x(m,n) and y(m,n) using the described model of HVS.
2. Compute ρ_{xy_avg} as the average value of locally computed correlation coefficients between images x(m,n) and y(m,n).
3. Compute ρ_{xe_avg} as the average value of the locally computed correlation coefficients between images x(m,n) and e(m,n) given by equation (4).
4. Find the image quality measure using equation (5).



(a) (b)

Figs. 1 (a) Contrast sensitivity function, (b) Function to model effects of random and signal dependent noise

5. PERFORMANCE RESULTS

Performance of the algorithm will be illustrated by several examples using the following parameter values: equal image height and width, viewing distance equal to four times image height and image size of 512x512 pixels. Parameter f_0 in eqn. (2) will depend on the original image. First example is the sequence of "Lena" images with various types of filtered noise as discussed in section 2. For this image, we set f_0=5. All three images in the sequence have the same MSE value of 100, but their visual qualities are very different as shown in Table 1.

Table 1. "Lena" images with various types of filtered noise

Min. and max. frequency	MSE	Quality measure (Q)
f_{min}=0 f_{max}=0.03	100	0.9120
f_{min}=0.03 f_{max}=0.15	100	0.6561
f_{min}=0.5 f_{max}=0.707	100	0.9292

Second example is the sequence of "Lena" images with various types of distortions. All images in the sequence have similar MSE values, but their visual qualities are very different. The results are shown in Table 2.

Table 2. "Lena" images with various types of distortion

Type of distortion	MSE	Quality measure (Q)
Contrast stretching	225	0.9766

Additive white Gaussian noise	225	0.6999
Blurring	224	0.2627
JPEG2000 coding	225	0.1898

Third example illustrates the performance of the proposed quality measure on the sequence of "Lena" images distorted by two sources of degradation: blurring caused by low-pass filtering and additive white Gaussian noise (zero-mean and standard deviation of s). First image contains only blurring, second image contains some blurring and some additive Gaussian noise and third image contains only noise. All images in the sequence have the same MSE. The results are shown in Table 3.

Table 3. "Lena" images with blurring and additive noise

Image	MSE	Quality measure (Q)
blurring only	420	0.1395
blurring and additive noise (s = 17.15)	420	0.2668
additive noise (s = 20.55)	420	0.6273

The proposed quality measure has another interesting property, that is, it can take positive and negative values (range of -1 to 1). It takes negative values when ρ_{xy_avg} is negative. This happens when one of the input images is inverted, which means the bright area becomes dark and vice versa. In this example the reference image is the original "Lena" image and distorted images are inverted original "Lena" image and inverted "Lena" image with additive noise. If the original image is f(m,n), then the corresponding inverted image can be computed as 255-f(m,n). The results are given in Table 4.

Table 4. "Lena" images with inverted grayscale

Image	MSE	Quality measure (Q)
Inverted "Lena" image	9258	-0.9955
Inverted "Lena" image with additive noise	9485	-0.6976

In the last example, the algorithm will be applied to the sequence of distorted "Couple" images. All distorted images have same or very close MSE values but their visual qualities are different. However, the differences in visual qualities are smaller than in the previous example involving "Lena" images. For this image, parameter f_0 is set to 12 cycles/degree because the image has a lot of small objects that attract viewer's attention, which means that information contained in high frequency components is important. If we used f_0=5 (as in the case of "Lena" image) for this image, then it would eliminate this information and distortion would be underestimated. On the other hand, if we used f_0=12 for "Lena" image, then it would overestimate high frequency distortion in that case. That is why this parameter is not fixed. Instead its value is set depending on the content of each image. The rule for choice of parameter f_0 is as follows: if reference image contains large object(s) then f_0 takes lower values; if reference image contains smaller object(s) then f_0 takes higher values. Three different images "Tiffany", "Man" and "Couple" and a sequence of "Couple" images were used to illustrate this point and results are given in Table 5.

In these examples proposed image quality measure produces results that are in good agreement with subjective visual quality of corresponding images. In contrast MSE values are the same for images with very different visual quality. (Note: Since any size reduction of an image to fit into one column will result in smoothing (distortion), their subjective visual quality cannot be judged fairly and hence the figures for the results are omitted from the paper. Unless the images are viewed in their normal size and can display 256 gray levels, they are not very useful for displaying visual quality).

Table 5: "Couple" images with various types of distortion

Type of distortion	MSE	Quality measure (Q)
Contrast stretching	81	0.9893
Additive white noise	81	0.8424
Blurring	81	0.7750
JPEG2000 coding	82	0.5419

6. CONCLUSION

An algorithm for image quality assessment has been presented. First, reasons for disagreement between MSE-based and subjective visual quality evaluation have been identified. Then a new quality measure has been defined. The proposed measure takes into account two HVS properties: nonlinear relationship between intensity and perceived brightness and presence of spatial filtering in HVS. This measure is based on average value of locally computed correlation coefficients, which is more closely related to the way in which human observer determines quality of an image than MSE. Finally, this value is modified by average value of locally computed correlation coefficients between original image and error image. This way the proposed measure differentiates between random and signal-dependant distortion, which have different effects on human observer.

Proposed image quality measure performs reasonably well. The examples presented here demonstrate that this measure ranks images according to their visual quality in cases when MSE-based measures fail to do that. However, subjective evaluation is still the best way for image quality assessment. HVS is more sophisticated than any mathematically defined image quality measure.

REFERENCES

[1] S. Daly, "The visible difference predictor: an algorithm for assessment of image fidelity," in Digital images and human vision, A. B. Watson (ed.), MIT Press, Cambridge, MA, pp. 179-206, 1993.

[2] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependant quantization and post-quantization data compression," Proc. ICASSP, Vol. 3, pp. 1945-1948, 1989.

[3] N. Damera-Venkata et al., "Image quality assessment based on a degradation model," IEEE Trans. on Image Processing, Vol. 9, No. 4, pp. 630-650, April 2000.

[4] T. N. Pappas and R. J. Safranek, "Perceptual criteria for image quality evaluation," in Handbook of image and video processing, Academic Press, May 2000.

[5] T. N. Pappas et al., "Supra-threshold perceptual image coding," Proc. ICIP, Vol. I, pp. 237-240, 1996.

[6] Z. Wang et al., "Why is image quality assessment so difficult?," in Proc. ICASSP, Vol. 4, pp. 3313-3316, 2002.

[7] Z. Wang and A. C. Bovik, "Universal image quality index," IEEE Signal Processing Letters, Vol. 9, pp. 81-84, March 2002.

[8] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," IEEE Trans. on Information Theory, Vol. IT-20, No. 4, pp. 525-536, July 1974.

[9] T. N. Pappas et al., "Least-squares model based halftoning," IEEE Trans. on Image Processing, pp. 1102-1116, Aug. 1999.

[10] Z. Xie et al., "Toward the unification of three visual laws and two visual models in brightness perception" IEEE Trans. on Systems, Man and Cybernetics, pp. 379–387, Mar-Apr 1989.