7-10-2003

An Algorithm for Image Quality Assessment

Goran Ivkovic
University of South Florida

Follow this and additional works at: <https://digitalcommons.usf.edu/etd>

Part of the [American Studies Commons](#)

Scholar Commons Citation

Ivkovic, Goran, "An Algorithm for Image Quality Assessment" (2003). *USF Tampa Graduate Theses and Dissertations*.

<https://digitalcommons.usf.edu/etd/1399>

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

An Algorithm for Image Quality Assessment

by

Goran Ivkovic

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Electrical Engineering
Department of Electrical Engineering
College of Engineering
University of South Florida

Major Professor: Ravi Sankar, Ph.D.
Wei Qian, Ph.D.
Arthur David Snider, Ph.D.

Date of Approval:
July 10, 2003

Keywords: brightness perception, contrast sensitivity function, human visual system,
image quality measure, image coding

© Copyright 2003, Goran Ivkovic

TABLE OF CONTENTS

LIST OF TABLES	ii
LIST OF FIGURES	iii
ABSTRACT	vii
1. INTRODUCTION	1
1.1. Background	1
1.2. Overview of image quality measures	1
1.3. Description and organization of thesis	5
2. HVS MODEL	7
2.1. Brightness perception	7
2.2. Contrast sensitivity function	12
2.3. Simple HVS model	22
3. IMAGE QUALITY MEASURE	23
3.1. MSE as image quality measure	23
3.2. Definition of image quality measure	29
4. RESULTS	36
4.1. “Lena” image with various types of distortion	36
4.2. “Goldhill” image with various types of distortion	37
4.3. “Couple” image with various types of distortion	42
4.4. “Tiffany”, “Lake” and “Mandrill” images with JPEG coding	48
4.5. “Woman”, “Man” and “Barbara” images with blurring	51
4.6. “Lena” image with inverted grayscale	55
4.7. “Lena” image with blurring and additive noise	57
4.8. Comparison of JPEG and JPEG2000 coding using “Lena” image	61
4.9. Performance of quality measure on two different images	70
5. CONCLUSION	71
REFERENCES	73

LIST OF TABLES

Table 1: “Lena” images with various types of distortion	36
Table 2: “Goldhill” images with various types of distortion	37
Table 3: “Couple” images with various types of distortion	43
Table 4: “Tiffany”, “Lake” and “Mandrill” images with JPEG coding	48
Table 5: “Woman”, “Man” and “Barbara” images with blurring	51
Table 6: “Lena” images with inverted grayscale	55
Table 7: “Lena” images with blurring and additive noise	58
Table 8: Comparison of JPEG and JPEG2000 coding	62
Table 9: Performance of quality measure on two different images	70

LIST OF FIGURES

Figure 1: Nonlinear function modeling brightness perception	10
Figure 2: “Lena” image with additive noise	10
Figure 3: Error image	11
Figure 4: Error image computed after original and noisy images were modified by (3)	11
Figure 5: Ideal magnitude frequency response	14
Figure 6: Magnitude frequency response of FIR filter	14
Figure 7: “Lena” image with filtered noise ($f_{\min}=0$, $f_{\max}=0.025$)	15
Figure 8: “Lena” image with filtered noise ($f_{\min}=0.05$, $f_{\max}=0.1$)	15
Figure 9: “Lena” image with filtered noise ($f_{\min}=0.1$, $f_{\max}=0.15$)	16
Figure 10: “Lena” image with filtered noise ($f_{\min}=0.15$, $f_{\max}=0.2$)	16
Figure 11: “Lena” image with filtered noise ($f_{\min}=0.3$, $f_{\max}=0.35$)	17
Figure 12: “Lena” image with filtered noise ($f_{\min}=0.35$, $f_{\max}=0.4$)	17
Figure 13: “Lena” image with filtered noise ($f_{\min}=0.45$, $f_{\max}=0.5$)	18
Figure 14: “Lena” image with filtered noise ($f_{\min}=0.5$, $f_{\max}=0.55$)	18
Figure 15: Contrast sensitivity function	19
Figure 16: Computation of vertical visual angle	20
Figure 17: Filtering of an image by CSF	22
Figure 18: Simple model of HVS	22
Figure 19: Original “Lena” image	25
Figure 20: “Lena” image with impulsive salt-pepper noise	25

Figure 21: “Lena” image with additive white noise	26
Figure 22: “Lena” image with multiplicative speckle noise	26
Figure 23: “Lena” image with mean shift	27
Figure 24: “Lena” image with contrast stretching	27
Figure 25: “Lena” image with blurring	28
Figure 26: JPEG coded “Lena” image	28
Figure 27: JPEG2000 coded “Lena” image	29
Figure 28: Function modeling the different effect of random and signal dependent noise	35
Figure 29: Original “Goldhill” image	38
Figure 30: “Goldhill” image with impulsive salt-pepper noise	38
Figure 31: “Goldhill” image with additive white noise	39
Figure 32: “Goldhill” image with multiplicative speckle noise	39
Figure 33: “Goldhill” image with mean shift	40
Figure 34: “Goldhill” image with contrast stretching	40
Figure 35: “Goldhill” image with blurring	41
Figure 36: JPEG coded “Goldhill” image	41
Figure 37: JPEG2000 coded “Goldhill” image	42
Figure 38: Original “Couple” image	43
Figure 39: “Couple” image with impulsive salt-pepper noise	44
Figure 40: “Couple” image with additive white noise	44
Figure 41: “Couple” image with multiplicative speckle noise	45
Figure 42: “Couple” image with mean shift	45
Figure 43: “Couple” image with contrast stretching	46

Figure 44: “Couple” image with blurring	46
Figure 45: JPEG coded “Couple” image	47
Figure 46: JPEG2000 coded “Couple” image	47
Figure 47: Original “Tiffany” image	48
Figure 48: JPEG coded “Tiffany” image	49
Figure 49: Original “Lake” image	49
Figure 50: JPEG coded “Lake” image	50
Figure 51: Original “Mandrill” image	50
Figure 52: JPEG coded “Mandrill” image	51
Figure 53: Original “Woman” image	52
Figure 54: “Woman” image with blurring	52
Figure 55: Original “Man” image	53
Figure 56: “Man” image with blurring	53
Figure 57: Original “Barbara” image	54
Figure 58: “Barbara” image with blurring	54
Figure 59: Inverted “Lena” image	56
Figure 60: Inverted “Lena” image with impulsive salt-pepper noise	56
Figure 61: Inverted “Lena” image with additive noise	57
Figure 62: “Lena” image with blurring	58
Figure 63: “Lena” image with blurring and additive noise of $sd=6.8$	59
Figure 64: “Lena” image with blurring and additive noise of $sd=9.4$	59
Figure 65: “Lena” image with blurring and additive noise of $sd=11.5$	60
Figure 66: “Lena” image with blurring and additive noise of $sd=13.3$	60

Figure 67: “Lena” image with additive noise of $sd=15$	61
Figure 68: JPEG coding at 1.1561 bits/pixel	63
Figure 69: JPEG2000 coding at 1.1561 bits/pixel	63
Figure 70: JPEG coding at 0.5501 bits/pixel	64
Figure 71: JPEG2000 coding at 0.5501 bits/pixel	64
Figure 72: JPEG coding at 0.3647 bits/pixel	65
Figure 73: JPEG2000 coding at 0.3647 bits/pixel	65
Figure 74: JPEG coding at 0.246 bits/pixel	66
Figure 75: JPEG2000 coding at 0.246 bits/pixel	66
Figure 76: JPEG coding at 0.1747 bits/pixel	67
Figure 77: JPEG2000 coding at 0.1747 bits/pixel	67
Figure 78: JPEG coding at 0.1428 bits/pixel	68
Figure 79: JPEG2000 coding at 0.1428 bits/pixel	68
Figure 80: JPEG coding at 0.1129 bits/pixel	69
Figure 81: JPEG2000 coding at 0.1129 bits/pixel	69

AN ALGORITHM FOR IMAGE QUALITY ASSESSMENT

GORAN IVKOVIC

ABSTRACT

Image quality measures are used to optimize image processing algorithms and evaluate their performances. The only reliable way to assess image quality is subjective evaluation by human observers, where the mean value of their scores is used as the quality measure. This is known as mean opinion score (MOS). In addition to this measure there are various objective (quantitative) measures. Most widely used quantitative measures are: mean squared error (MSE), peak signal to noise ratio (PSNR) and signal to noise ratio (SNR). Since these simple measures do not always produce results that are in agreement with subjective evaluation, many other quality measures have been proposed. They are mostly various modifications of MSE, which try to take into account some properties of human visual system (HVS) such as nonlinear character of brightness perception, contrast sensitivity function (CSF) and texture masking. In these approaches quality measure is computed as MSE of input image intensities or frequency domain coefficients obtained after some transform (DFT, DCT etc.), weighted by some coefficients which account for the mentioned properties of HVS. These measures have some advantages over MSE, but their ability to predict image quality is still limited.

A different approach is presented here. Quality measure proposed here uses simple model of HVS, which has one user-defined parameter, whose value depends on the reference image. This quality measure is based on the average value of locally computed

correlation coefficients. This takes into account structural similarity between original and distorted images, which cannot be measured by MSE or any kind of weighted MSE. The proposed measure also differentiates between random and signal dependant distortion, because these two have different effect on human observer. This is achieved by computing the average correlation coefficient between reference image and error image. Performance of the proposed quality measure is illustrated by examples involving images with different types of degradation.

1. INTRODUCTION

1.1. Background

Methods for image quality assessment can be divided into two groups: subjective and objective. The only subjective measure is mean opinion score (MOS) where quality of an image is judged by several individuals and then the mean value of their scores is used as the measure. Since human observer is ultimate receiver of the information contained in an image, this is the best way to assess image quality.

Objective measures try to assess image quality and express it as a number. The goal in developing objective (quantitative) measures is to find the one that correlates well with image quality as perceived by human visual system (HVS). These measures are used to assess performance of different algorithms for image enhancement and coding. Quality measures can also be used to optimize image processing algorithms. For example in image coding the goal is to represent images with as few bits as possible with the minimal loss of image quality. Quality measure is used as a criterion in a sense that the bits that contribute more significantly to the used quality measure are coded first.

1.2. Overview of image quality measures

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 we compare images with the same type of degradation.

In this case distorted image with smaller MSE will be perceived closer to the original image than the one with greater MSE. However, when we compare images with different types of degradation 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, many other algorithms have been developed. These algorithms try to assess image quality not just by using some simple formula but also by taking into account the properties of 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.

One such algorithm is described in [1]. In this approach input intensities are first modified by a nonlinear function modeling brightness perception and then filtered by a 2-D filter modeling frequency response of HVS. After this, processed images are decomposed into different subbands using a system of filters. Then perceptual threshold is computed for every spatial location in every subband. Using the computed threshold probability that the error is visible is computed again for every spatial location in each subband. Finally, probabilities at same spatial location from different subbands are combined to compute probability of visible difference at that spatial location. The output of this algorithm is map of probabilities which has the same size as input images. If the value at some spatial location is 1, then there is visible difference at that location between input images. If the value is less than 1, then the difference is not clearly visible, it is somewhere in the threshold region. This approach can identify the pixels with visible

distortion, but it cannot quantify the extent of easily visible distortion. Also, this algorithm does not produce one number that indicates image quality.

Image quality measures are sometimes developed as a part of image processing algorithms. An example of this can be found in [2], where properties of HVS are used to design an algorithm for image coding. In this approach input image is first decomposed into 16 subbands using generalized quadrature mirror filters. Each subband is downsampled by a factor of four in horizontal and vertical directions, so the total number of subband coefficients is equal to the number of pixels in the input image. Then perceptual threshold is computed for every subband coefficient at each spatial location. Computation of the perceptual thresholds takes into account the same HVS properties as in [1], but here they are modeled in a different way. If the coefficient in some subband at some location is above the computed perceptual threshold then it is visible to human observer, otherwise it is not visible. Each subband is then coded using differential pulse code modulation (DPCM), where the quantizer step for a subband is equal to the minimal perceptual threshold in that subband. Image quality measure in this case can be defined as MSE of weighted subband coefficients of the original image and distorted image, where each subband coefficient is normalized by the perceptual threshold for that subband at that spatial location.

Another approach is presented in [3]. It is assumed here that input image is degraded by two sources of degradation: linear distortion (for example as a result of filtering) and additive noise. In the first step linear distortion and additive noise are separated and then two quality measures are computed: distortion measure and noise quality measure. Distortion measure is computed using discrete Fourier transform (DFT) coefficients of

the original image and original image distorted by linear distortion only. Noise quality measure is computed using the original image degraded by linear distortion and noise and the original image degraded by linear distortion only. These two images are decomposed into subbands using a system of filters and perceptual threshold is computed for every subband and every spatial location. If the difference between the two images is below the perceptual threshold it is not considered visible and that difference is ignored, otherwise the difference is taken into account. Finally, noise quality measure is compute as SNR of the modified input images, which contain only visible difference between the two images. This approach requires as input not only two images but also the algorithm that causes linear distortion and produces two numbers, which assess image quality.

An overview of image quality measures is presented in [4]. Performance of several proposed quality measures is tested in [5]. Distorted images are created using four different image coding algorithms, which produce different types of degradation. Then tested quality measures are computed for each distorted image and these values are compared with subjective assessment. 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.

In [6] and [7] previously proposed quality measures are discussed and it is noted that these measures frequently use some problematic assumptions. Also it is pointed out that MSE or some modifications of MSE (such as the one described in [2]) completely fail to measure structural similarity between two images, which is very important for a human observer. In other words MSE computes only error between the corresponding pixels but

does not take into account correlation between that pixel and surrounding pixels in any of the two input images. It is possible to introduce a very large MSE by shifting the mean value of image intensities or by contrast stretching and preserve structural similarity between the two images. On the other hand if the same amount of MSE is introduced as a result of low-pass filtering or image coding, the resulting image can be severely distorted. Algorithms described in [1], [2] and [3] do not measure structural similarity at all. Based on this discussion a new 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. This approach does not use HVS model at all.

1.3. Description and organization of the thesis

The approach presented here will use some ideas 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 which consists 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 one user-defined parameter which 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.

The thesis is organized as follows. Chapter two describes HVS model used to process input images. Chapter three discusses problems with MSE and presents the proposed algorithm. Performance of the proposed algorithm will be illustrated by examples involving images with different types of distortion in Chapter four. Most of these images (but not all) are the same ones used to test quality measure developed in [6] and [7]. Chapter five concludes the thesis.

2. HVS MODEL

2.1. Brightness perception

Digital images are represented using a finite number of intensity levels. For example, gray scale images with 8 bits/pixel are represented by 256 intensity levels, where 0 corresponds to the darkest level and 255 corresponds to the brightest level. Intensities between these two extremes represent various shades of gray level. Brightness perceived by HVS is not a linear function of intensity represented by integers from 0 to 255. This can be shown by the following experiment. First we create an image which consists of uniform background of intensity I and a small square in the center of intensity $I + \Delta I$, and then we repeat this for different values of background intensity I while keeping ΔI constant. When the background intensity is very low (for example $I=10$) or very high ($I=240$) the difference in perceived brightness between the background and the center square is smaller than in the case when the background intensity is in the mid range ($I=125$). Since the intensity difference between the background and the center square is the same in each case, perceived brightness must be a nonlinear function of intensity. In other words equal increases in intensity do not correspond to equal increases in perceived brightness. This phenomenon is well known but it is very difficult to model it because we do not know how to measure perceived brightness. Usually, this is modeled by transforming input intensities by some nonlinear function, such as logarithmic or cube

root function. If we denote perceived brightness by B and input intensity by I , then by using logarithmic function they are related as:

$$B = K \log I \quad (1)$$

where K is a constant. This is known as Weber's law and is cited in many books on image processing (for example [4] and [8]). The second formula, which is used in [9], is obtained using cube root function:

$$B = K(I)^{0.33} \quad (2)$$

where K is again some constant. None of these two formulas models the perception of brightness accurately, especially at low intensities. According to them the sensitivity to intensity increases is greatest at low intensities and decreases at mid and high intensities. This is clearly wrong, because an increase in intensity result in a very small increase in perceived brightness, when the intensity is low. Therefore, the following nonlinear function is used here to model brightness perception:

$$B = \begin{cases} \frac{B_{MAX}}{2} \left(\frac{2I}{I_{MAX}} \right)^{2.5} & \text{if } 0 < I < I_{MAX}/2 \\ B_{MAX} - \frac{B_{MAX}}{2} \left(\frac{2(I_{MAX}-I)}{I_{MAX}} \right)^{2.5} & \text{if } I_{MAX}/2 < I < I_{MAX} \end{cases} \quad (3)$$

where $I_{MAX}=255$ is the maximum value of intensity and $B_{MAX}=100$ is the maximum value of perceived brightness. In other words when the intensity goes from 0 to 255, the perceived brightness goes from 0 to 100 as shown in Figure 1. 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. The biggest problem with brightness perception models is how to test them and tell whether they produce reasonable results or not. In this case “Lena” image corrupted by additive white Gaussian noise shown in Figure 2 is used as a test. Then the first residual image is computed as original image minus corrupted image, which is shown in Figure 3. The second residual image shown in Figure 4 is computed as the difference between the same images, but this time both images were modified by the function given by equation (3). In the first residual image we see uniform amount of noise across the image. However, in the second residual image there is less noise in the areas corresponding to the very dark or very bright areas in the original image and there is more noise in the areas where intensity is in the mid range. This is approximately what we see in the noisy image: noise is less visible in very bright or very dark areas than in other areas of the image. According to the models given by equations (1) and (2) the noise would be most visible in the very dark areas of the image, which is clearly not the case. That is why the model given by equation (3) is used instead of the models given by equations (1) and (2). It is not possible to give general formula that relates perceived brightness and intensity, because this depends on the display device, so the model of brightness perception must be determined separately for each display device.

Figure 1: Nonlinear function modeling brightness perception

Figure 2: "Lena" image with additive noise

Figure 3: Error image

Figure 4: Error image computed after original and noisy images were modified by the function (3)

2.3. Contrast sensitivity function

Contrast sensitivity function (CSF) describes the frequency response of HVS. HVS is not equally sensitive to all spatial frequencies, which can be shown by the following experiment. First we design a 2-D filter, which will approximately have the following magnitude frequency response:

$$H(f_1, f_2) = \begin{cases} 1, & \text{for } f_{\min} < \sqrt{f_1^2 + f_2^2} < f_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

In the last formula f_1 and f_2 are normalized frequencies, taking values between -0.5 and 0.5. Therefore, the values for $f_{\min}$ and $f_{\max}$ must be between 0 and 0.707. This filter cannot be realized exactly but it can be numerically approximated by creating a finite impulse response (FIR) filter using frequency sampling method. FIR filter obtained by this method will have the same magnitude frequency response as the ideal filter at given points in the frequency plane. Magnitude frequency response of the ideal filter and the obtained FIR filter are shown in Figures 5-6 for $f_{\min}=0.35$ and $f_{\max}=0.4$. Then we generate white Gaussian noise and filter it by the FIR filter obtained using the described procedure. This filtered noise is added to the original Lena image for various values of $f_{\min}$ and $f_{\max}$. Each time the filtered noise is scaled so that its standard deviation is 10. We get the sequence of images shown in Figures 7-14. All images in the sequence are corrupted by the noise with the same variance, but the noise is not equally visible in each image. The noise is quite noticeable at low frequencies and then it becomes even more noticeable with increase in frequency up to some point. After that we see gradual decrease in noise visibility with increase in frequency. At very high frequencies we can hardly see any noise at all even though these images are corrupted by the noise with the

same variance as the rest of the sequence. This example shows that HVS is a frequency selective system. If HVS were not a frequency selective system, we would see the same amount of noise in each image in the sequence. 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 and it is given in the normalized units. It is equal to one at the frequency where the sensitivity is maximal and everywhere else is less than one. Also frequency is expressed in cycles/degree. This may look strange because it seems natural to express it in cycles/cm. The reason for this is the following. Images can be observed from the various distances and the same spatial frequency expressed in cycles/cm will be perceived differently for different viewing distances. If the viewing distance is greater signal of the certain spatial frequency will appear to have greater frequency than that same signal viewed from the smaller distance. That means that HVS perceives spatial frequency in cycles per unit visual angle instead of frequency in cycles per unit length.

Various functions have been suggested to model this effect. Most of them have a band-pass character but they have maximum at different frequencies. For example CSF shown in [4] has maximum at 3 cycles/degree, while CSF used in [9] has maximum at 8 cycles/degree. It is also noted in [10] that various CSF suggested by different authors have maximum at frequencies ranging from 3 to 10 cycles/degree. 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:

Figure 5: Ideal magnitude frequency response

Figure 6: Magnitude frequency response of FIR filter

Figure 7: "Lena" image with filtered noise ($f_{\min}=0$, $f_{\max}=0.025$)

Figure 8: "Lena" image with filtered noise ($f_{\min}=0.05$, $f_{\max}=0.1$)

Figure 9: "Lena" image with filtered noise ($f_{\min}=0.1$, $f_{\max}=0.15$)

Figure 10: "Lena" image with filtered noise ($f_{\min}=0.15$, $f_{\max}=0.2$)

Figure 11: "Lena" image with filtered noise ($f_{\min}=0.3$, $f_{\max}=0.35$)

Figure 12: "Lena" image with filtered noise ($f_{\min}=0.35$, $f_{\max}=0.4$)

Figure 13: "Lena" image with filtered noise ($f_{\min}=0.45$, $f_{\max}=0.5$)

Figure 14: "Lena" image with filtered noise ($f_{\min}=0.5$, $f_{\max}=0.55$)

$$H(f) = \begin{cases} S(f), & \text{for } f \leq 3.5 \text{ cycles/degree} \\ [S(f)]^\alpha, & \text{for } f > 3.5 \text{ cycles/degree} \end{cases}$$

where

$$S(f) = (0.0509 + 0.695f) \exp(-(0.2622f)^{1.1}) \quad (5)$$

In the last formula, f denotes frequency in cycles/degree and α represents the parameter which controls the rate at which CSF decreases after it reaches its maximum at 3.5 cycles/degree. This function is shown in Figure 15 for $\alpha=0.75$.

Figure 15: Contrast sensitivity function

Although the previous experiment with the sequence of “Lena” images shows that HVS is frequency selective, we must be cautious with this interpretation. This example shows that noise at high frequencies is less visible than noise at low frequencies. However, when high frequency components represent information contained in an image

(for example some details or small objects) instead of noise, those components become very important for a human observer. Because of this effect, HVS is not modeled here as a filter with fixed parameters. Instead, a parameter α , which controls attenuation at frequencies higher than 3.5 cycles/degree, is introduced. This parameter is user-defined and depends on the reference image. If the reference image contains one large object (for example “Lena” image) than we use higher values for α than in the case when the reference image contains smaller objects or lot of fine details. This requires some user intervention but it yields better results.

This property of HVS can be taken into account by filtering input image by CSF. The first step in this process is to compute vertical and horizontal viewing angles. The vertical visual angle can be computed from Figure 16, where the height of the image is denoted by h , viewing distance is denoted by d and eye of a human observer is at the point O

Figure 16: Computation of vertical visual angle

Then the vertical visual angle can be found from the following formula:

$$\tan \frac{\theta_y}{2} = \frac{h}{2d} \quad (6)$$

If we express this angle in degrees we get:

$$\theta_y = \frac{180}{\pi} 2 \tan^{-1} \left(\frac{h}{2d} \right) \quad (7)$$

If we denote width of the image by w, the horizontal viewing angle can be found as:

$$\theta_x = \frac{180}{\pi} 2 \tan^{-1} \left(\frac{w}{2d} \right) \quad (8)$$

For h=15 cm, w=15 cm and d=60 cm, the values for the horizontal and vertical viewing angles are: $\theta_x=14.25^\circ$ and $\theta_y=14.25^\circ$ degrees. If this image is 512x512 pixels, the sampling frequencies in horizontal and vertical directions in samples/degree are found as:

$$\begin{aligned} f_{xs} &= \frac{512}{\theta_x} = 35.93 \text{ samples/degree} \\ f_{ys} &= \frac{512}{\theta_y} = 35.93 \text{ samples/degree} \end{aligned} \quad (9)$$

Then we compute the DFT of the input image denoted by $f(m,n)$, whose size is M by N pixels:

$$F(k,l) = DFT(f(m,n)) \quad (10)$$

If we denote CSF by $H(f)$, the DFT of the output image denoted by $X(k,l)$ is:

$$X(k,l) = H(f(k,l))F(k,l) \quad (11)$$

where

$$f(k,l) = \sqrt{f_x(l)^2 + f_y(k)^2} \quad (12)$$

$$f_x(l) = \begin{cases} \frac{l}{N} f_{xs}, & \text{for } 0 \leq l < N/2 \\ \frac{N-l}{N} f_{xs}, & \text{for } N/2 \leq l \leq N-1 \end{cases} \quad (13)$$

$$f_y(k) = \begin{cases} \frac{k}{M} f_{ys}, & \text{for } 0 \leq k < M/2 \\ \frac{M-k}{M} f_{ys}, & \text{for } M/2 \leq k \leq M-1 \end{cases} \quad (14)$$

Finally, the output signal denoted by $x(m,n)$ is found as inverse DFT of $X(k,l)$:

$$x(m,n) = IDFT(X(k,l)) \quad (15)$$

Figure 17: Filtering of an image by CSF

The described procedure can be represented as a simple 2-D filtering as shown in Figure 17, where it is assumed that 2-D filter has a radial symmetry:

$$H(f_1, f_2) = H(\sqrt{f_1^2 + f_2^2}) \quad (16)$$

2.3. Simple HVS model

Based on the discussion in this chapter we can create a simple model of HVS shown in Figure 18, which models brightness perception and frequency response of HVS. 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

Figure 18: Simple model of HVS

3. IMAGE QUALITY MEASURE

3.1. MSE as image quality measure

The objective of this research is to develop an image quality measure, which will produce results that are in agreement with subjective assessment. But before this, it is reasonable to ask why simple measures such as MSE or SNR are not working well. These simple measures work when we have images with the same type of degradation. For example, if an image is distorted by additive white Gaussian noise the image with the smallest MSE or SNR will be perceived to be the closest to the original image. This will also be the case for all others types of distortion such as blurring or coding. If we have a sequence of images obtained by distorting an image by various amounts of the same type of degradation, the image with the smallest MSE or SNR will be perceived by a human observer to be the closest to the original image. However, when we compare images with different types of degradation, the image with the smallest MSE or SNR will not always be perceived to be the closest to the original image. This can be easily seen by looking at the sequence of “Lena” images distorted by various types of degradation: impulsive salt-pepper noise, additive white noise, multiplicative speckle noise, mean shift, contrast stretching, blurring (which is a result of low-pass filtering), JPEG and JPEG2000 coding. The original image and distorted images are shown in Figures 19-27. All distorted images have the same MSE (except for the JPEG coded image) but their visual quality is very different. Since images with the same MSE have very different visual quality, it can

be concluded that MSE is not the criterion used by a human observer in judging how close the distorted image is to the original image. If MSE were the criterion then the images with the same MSE would have the same visual quality. Another example for this is the sequence of “Lena” images shown in Figures 7-14. All images in the sequence have the same MSE (because the variance of added noise is always the same), but the visual quality of those images is very different.

This is a well-known fact and this is why many other algorithms for image quality assessment have been developed. These algorithms are using various models of HVS, which are used to transform input intensities from the original and distorted images. Quality measure is then computed using transformed intensities as MSE or some kind of weighted MSE. The basic idea in these algorithms is to take into account the properties of HVS by modeling the process of visual perception. However our knowledge of this process is very limited and these algorithms frequently use very questionable assumptions. In some cases these algorithms are very complicated and they do not perform very well.

Main problem with MSE or MSE-based image quality measures is the fact that they treat an image as a set of uncorrelated numbers. When MSE is computed correlation between the neighboring pixels is completely ignored. This information is very important because human observer is more concerned with correlation between the neighboring pixels, which determines what type of structure they form in an image, than with the actual intensity at a given location. In other words the intensity at a given point is not so important, as the structure which that pixel creates with surrounding pixels.

Figure 19: Original "Lena" image

Figure 20: "Lena" image with impulsive salt-pepper noise (MSE=225, Q=0.8543)

Figure 21: "Lena" image with additive white noise ($MSE=225$, $Q=0.7115$)

Figure 22: "Lena" image with multiplicative speckle noise ($MSE=225$, $Q=0.7224$)

Figure 23: "Lena" image with mean shift (MSE=225, $Q=0.9954$)

Figure 24: "Lena" image with contrast stretching (MSE=225, $Q=0.9839$)

Figure 25: "Lena" image with blurring (MSE=224, $Q=0.2611$)

Figure 26: JPEG coded "Lena" image (MSE=215, $Q=0.2216$)

Figure 27: JPEG2000 coded "Lena" image (MSE=225, Q=0.1837)

3.2. 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 $M \times N$ pixels. In the first step, both images will be transformed by the simple model of HVS, which models the brightness perception and frequency selectivity (CSF). This model is presented in Figure 18 in the previous Chapter. The model is based on the properties of HVS discussed in Chapter two. The nonlinear function models brightness perception and it is given by equation (3). The 2-D filter is given by equation (16), where $H(f)$ is given by equation (5).

The result of this filtering depends on the viewing distance, the width and height and the number of pixels in input images, as it was described in Chapter 2, 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) or weighted SNR (WSNR). This is better than regular MSE or SNR, because it takes into account some properties of HVS, but it still does not produce satisfactory results. For example these measures do not explain why we perceive contrast stretched image as very similar to the original image, despite very large error. The possible answer to this question is that HVS must use some criterion other than MSE when determining similarity of the distorted image to the reference image. By looking at the sequence of the distorted “Lena” images we see that the images with mean shift and contrast stretching are much closer to the original image than rest of the distorted images. The main difference between these two images and the rest of them is that in these two cases there is a high degree of linear relationship between these two images and the original image, while in all other cases it is much lower. This leads to an important conclusion: HVS perceives two images to be similar when there is a linear relationship between them. The degree of linear relationship between the two vectors can be measured by correlation coefficient. If we have two vectors $X = [x_1 \ x_2 \ \dots \ x_N]$ and $Y = [y_1 \ y_2 \ \dots \ y_N]$, then the correlation coefficient between them is defined as:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

where

$$\begin{aligned} \sigma_{XY} &= \frac{1}{N} \sum_{i=1}^N (x_i - m_X)(y_i - m_Y) \\ \sigma_X &= \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - m_X)^2} , \quad \sigma_Y = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - m_Y)^2} \\ m_X &= \frac{1}{N} \sum_{i=1}^N x_i , \quad m_Y = \frac{1}{N} \sum_{i=1}^N y_i \end{aligned} \quad (17)$$

Correlation coefficient defined by equation (17) takes values between -1 and 1. It takes the value of 1 only when there is a linear relationship between X and Y and in all other cases is smaller than one. The closer it is to zero, the lower degree of linear relationship exists between X and Y.

Correlation coefficient will be used to create the image quality measure. Images $x(m,n)$ and $y(m,n)$ obtained after processing the reference image $f(m,n)$ and the distorted image 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 $\rho_{xy}(i)$ then we can find the average value of the correlation coefficient between the $x(m,n)$ and $y(m,n)$ as:

$$\rho_{xy_avg} = \frac{1}{L} \sum_i^L \rho_{xy}(i) \quad (18)$$

where L is the total number of data blocks. Correlation coefficient between two images can be computed in one step, by applying equation (17) on all the pixels of the two

images at once. This method would result in relatively high values for correlation coefficient regardless of the distortion, because correlation coefficient computed this way would overlook difference in details between the two images. By computing correlation coefficient locally we force it to concentrate more on the difference in details, which are 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. That is why correlation coefficient is computed as the average value of locally computed correlation coefficients instead of computing it in one step.

Average correlation coefficient given by equation (18) is a good indicator of similarity between the reference image and the distorted image, but it has one downside: it does not differentiate between random and signal dependent distortion. It is important for any image quality measure to differentiate between these two types of distortion, because they have very different visual effect on a human observer. Random distortion has much smaller effect on image quality than the same amount of signal dependent distortion. This can be easily seen by comparing the first three distorted “Lena” images (images with impulsive noise, additive noise and multiplicative noise) with the last three “Lena” images (images with blurring, JPEG and JPEG2000 coding). The first three images have random distortion and higher visual quality compared to last three images, which have signal dependant distortion and lower visual quality.

Since average correlation coefficient from equation (18) does not differentiate between random and signal dependant distortion, it will treat any loss of correlation equally. 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 image $x(m,n)$ and the error image $e(m,n)$, which is given by:

$$e(m, n) = x(m, n) - \text{sign} \{ \rho_{xy_avg} \} y(m, n) \quad (19)$$

Then we compute local correlation coefficients 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 $\rho_{xe}(i)$, the average correlation coefficient between images $x(m,n)$ and $e(m,n)$ is:

$$\rho_{xe_avg} = \frac{1}{L} \sum_i^L \rho_{xe}(i) \quad (20)$$

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 given by equation (20) will have higher absolute value, indicating signal dependent noise. Finally, image quality measure denoted by Q is defined as:

$$Q = \text{sign} \{ \rho_{xy_avg} \} | \rho_{xy_avg} |^{f(\rho_{xe_avg})}$$

where

$$f(\rho_{xe_avg}) = 1.2 + 0.5 \frac{\exp(\frac{|\rho_{xe_avg}| - 0.3}{0.15}) - \exp(-\frac{|\rho_{xe_avg}| - 0.3}{0.15})}{\exp(\frac{|\rho_{xe_avg}| - 0.3}{0.15}) + \exp(-\frac{|\rho_{xe_avg}| - 0.3}{0.15})} \quad (21)$$

Equation (21) needs some explanation. Quality measure (Q) will have the same sign as average correlation coefficient given by equation (18), 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 between the images $x(m,n)$ and $e(m,n)$ given by equation (20). This modification is obtained using function $f(\rho_{xe_avg})$ which appears in equation (21) and is shown in Figure 28. If the magnitude of the average correlation coefficient given by equation (20) 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 the average correlation coefficient given by equation (18) 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 equation (21) are determined by trial and error experiments using various images with random and signal dependant distortion.

Figure 28: Function modeling the different effect of random and signal dependent noise

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 the average correlation coefficient given by equation (18) as the average value of locally computed correlation coefficients between images $x(m,n)$ and $y(m,n)$.
3. Compute the average correlation coefficient given by equation (20) as the average value of the locally computed correlation coefficients between images $x(m,n)$ and $e(m,n)$, which is given by equation (19).
4. Find the image quality measure using equation (21).

4. RESULTS

4.1. “Lena” image with various types of distortion

Performance of the algorithm will be illustrated by several examples. In all examples in this chapter the following parameter values will be used: image height and width of 15 cm, viewing distance of 60 cm and image size of 512x512 pixels. Parameter α from the equation (6) will depend on the original image. First example will be the sequence of “Lena” images with various types of distortion shown in Figures 19-27. For this image, we set $\alpha=0.75$. All images in the sequence have the same MSE value of 225 (except for JPEG coded image, which has MSE of 215), but their visual quality is very different. The results are shown in Table 1.

Table 1: “Lena” images with various types of distortion

Type of distortion	MSE	Quality measure (Q)
Impulsive salt-pepper noise	225	0.8543
Additive white Gaussian noise	225	0.7115
Multiplicative speckle noise	225	0.7224
Mean shift	225	0.9954
Contrast stretching	225	0.9839
Blurring	224	0.2611
JPEG coding	215	0.2216
JPEG2000 coding	225	0.1837

These results are in agreement with visual quality of the corresponding images. Although we cannot define some clear criterion for subjective quality assessment, human observer can intuitively feel when distorted image is more or less close to the reference

image. Most human observers would agree with the rankings given by the proposed quality measure and in that sense these results are very reasonable.

4.2. “Goldhill” image with various types of distortion

In this example, the proposed algorithm is applied to the sequence of distorted “Goldhill” images. The types of distortion are the same as in the previous example with distorted “Lena” images, but in this case MSE values are smaller and differences in subjective visual quality between distorted images are smaller. In this case, we set $\alpha=0.25$ because this image contains smaller objects compared to “Lena” image. Again the distorted images have same or very close MSE values. These images are shown in the Figures 29-37 and results are shown in Table 2. The results are in agreement with subjective visual quality of the distorted images.

Table 2: “Goldhill” images with various types of distortion

Type of distortion	MSE	Quality measure (Q)
Impulsive salt-pepper noise	120	0.9353
Additive white Gaussian noise	121	0.8258
Multiplicative speckle noise	121	0.8469
Mean shift	121	0.9975
Contrast stretching	121	0.9889
Blurring	122	0.5206
JPEG coding	117	0.4076
JPEG2000 coding	122	0.3880

Figure 29: Original "Goldhill" image

Figure 30: "Goldhill" image with impulsive salt-pepper noise (MSE=120, $Q=0.9353$)

Figure 31: "Goldhill" image with additive white noise ($MSE=121$, $Q=0.8258$)

Figure 32: "Goldhill" image with multiplicative speckle noise ($MSE=121$, $Q=0.8469$)

Figure 33: "Goldhill" image with mean shift (MSE=121, $Q=0.9975$)

Figure 34: "Goldhill" image with contrast stretching (MSE=121, $Q=0.9889$)

Figure 35: "Goldhill" image with blurring (MSE=122, $Q=0.5206$)

Figure 36: JPEG coded "Goldhill" image (MSE=117, $Q=0.4076$)

Figure 37: JPEG2000 coded "Goldhill" image (MSE=122, Q=0.3880)

4.3. "Couple" image with various types of distortion

In this 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 two examples. For this image, parameter α is set to 0.15 because the image has a lot of small objects which attract viewer's attention, which means that information contained in high frequency components is important. If we used $\alpha=0.75$ (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 $\alpha=0.15$ 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. Sequence of “Couple” images is shown in Figures 38-46 and results are given in Table 3.

Table 3: “Couple” images with various types of distortion

Type of distortion	MSE	Quality measure (Q)
Impulsive salt-pepper noise	81	0.9553
Additive white Gaussian noise	81	0.8393
Multiplicative speckle noise	81	0.8436
Mean shift	81	0.9990
Contrast stretching	81	0.9924
Blurring	81	0.7832
JPEG coding	82	0.6046
JPEG2000 coding	82	0.5535

Figure 38: Original "Couple" image

Figure 39: "Couple" image with impulsive salt-pepper noise ($MSE=81$, $Q=0.9553$)

Figure 40: "Couple" image with additive white noise ($MSE=81$, $Q=0.8393$)

Figure 41: "Couple" image with multiplicative speckle noise (MSE=81, $Q=0.8436$)

Figure 42: "Couple" image with mean shift (MSE=81, $Q=0.9990$)

Figure 43: "Couple" image with contrast stretching ($MSE=81$, $Q=0.9924$)

Figure 44: "Couple" image with blurring ($MSE=82$, $Q=0.7832$)

Figure 45: JPEG coded "Couple" image (MSE=82, Q=0.6046)

Figure 46: JPEG2000 coded "Couple" image (MSE=82, Q=0.5535)

4.4. “Tiffany”, “Lake” and “Mandrill” images with JPEG coding

In the next example, the algorithm is applied on three different JPEG coded images: “Tiffany”, “Lake” and “Mandrill”. All three images have similar MSE values but very different visual quality. Values for parameter α are: 0.75 for “Tiffany” and “Mandrill” images and 0.25 for “Lake” image. Images are shown in Figures 47-52 and results are presented in Table 4.

Table 4: “Tiffany”, “Lake” and “Mandrill” images with JPEG coding

Image	MSE	Quality measure (Q)
Tiffany	165	0.2331
Lake	168	0.4323
Mandrill	163	0.9356

Figure 47: Original "Tiffany" image

Figure 48: JPEG coded "Tiffany" image (MSE=165, Q=0.2331)

Figure 49: Original "Lake" image

Figure 50: JPEG coded "Lake" image (MSE=168, Q=0.4323)

Figure 51: Original "Mandrill" image

Figure 52: JPEG coded "Mandrill" image (MSE=163, Q=0.9356)

4.5. "Woman", "Man" and "Barbara" images with blurring

Next example shows performance of the algorithm on three different blurred images: "Woman", "Man" and "Barbara". Again, all three images have identical MSE values, but very different visual quality. Values of parameter α are: 0.75 for "Woman" image and 0.5 for "Man" and "Barbara" images. These images are shown in Figures 53-58 and results are presented in Table 5.

Table 5: "Woman", "Man" and "Barbara" images with blurring

Image	MSE	Quality measure (Q)
Woman	200	0.1456
Man	200	0.4074
Barbara	200	0.7855

Figure 53: Original "Woman" image

Figure 54: "Woman" image with blurring (MSE=200, Q=0.1456)

Figure 55: Original "Man" image

Figure 56: "Man" image with blurring (MSE=200, Q=0.4074)

Figure 57: Original "Barbara" image

Figure 58: "Barbara" image with blurring (MSE=200, $Q=0.7855$)

In all examples presented here the proposed quality measure ranks images according to their visual quality. This is in contrast to MSE values which can be identical even when the difference in the extent of distortion is very large.

4.6. “Lena” image with inverted gray scale

The proposed quality measure has another interesting property. As can be seen from equation (21), this quality measure can take values that range from -1 to 1. In other words it can take positive and negative values. It takes negative values when the average correlation coefficient given by equation (18) 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” images with impulsive and additive noise. If the original image is $f(m,n)$, then the corresponding inverted image can be computed as $255-f(m,n)$. These images are shown in Figures 59-61 and the results are given in Table 6.

Table 6: “Lena” images with inverted grayscale

Image	MSE	Quality measure (Q)
Inverted “Lena” image	9258	-1
Inverted “Lena” image with impulsive noise	9369	-0.8543
Inverted “Lena” image with additive noise	9485	-0.7115

Figure 59: Inverted "Lena" image (MSE=9258, $Q=-1$)

Figure 60: Inverted "Lena" image with impulsive salt-pepper noise (MSE=9369, $Q=-0.8543$)

Figure 61: Inverted "Lena" image with additive white noise (MSE=9485, Q=-0.7115)

Quality measure values for inverted images with impulsive and additive noise have the same magnitude as the values obtained for the corresponding non-inverted images, but they have negative sign. This is in a sense related to what we see: images with the same structure but inverted grayscale. MSE produces only very large values, which is another illustration that MSE is not related to the way we perceive image similarity.

4.7. "Lena" image with blurring and additive noise

Next 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 σ). Images contain some amount of noise and some blurring in different proportions.

First image contains only blurring, second image contains less blurring and some noise, third image contains less blurring and more noise than the second etc. Last image contains only noise. All images in the sequence have the same MSE. These images are shown in Figures 62-67 and results are shown in Table 7.

Table 7: "Lena" images with blurring and additive noise

Image	MSE	Quality measure (Q)
blurring only	225	0.2031
blurring and additive noise ($\sigma = 6.8$)	225	0.2625
blurring and additive noise ($\sigma = 9.4$)	225	0.3549
blurring and additive noise ($\sigma = 11.5$)	225	0.5411
blurring and additive noise ($\sigma = 13.3$)	225	0.6637
additive noise ($\sigma = 15$)	225	0.7108

Figure 62: "Lena" image with blurring (MSE=225, Q=0.2031)

Figure 63: "Lena" image with blurring and additive noise ($\sigma=6.8$), (MSE=225, Q=0.2625)

Figure 64: "Lena" image with blurring and additive noise ($\sigma=9.4$), (MSE=225, Q=0.3547)

Figure 65: "Lena" image with blurring and additive noise ($\sigma = 11.5$), (MSE=225, Q=0.5411)

Figure 66: "Lena" image with blurring and additive noise ($\sigma = 13.3$), (MSE=225, Q=0.6637)

Figure 67: "Lena" image with additive noise ($\sigma = 15$), (MSE=225, Q=0.7108)

4.8. Comparison of JPEG and JPEG2000 coding using "Lena" image

In the next example, the proposed quality measure will be used to compare two algorithms for image coding: JPEG and JPEG2000. In both of these algorithms image intensities are transformed into some coefficients and those coefficients are coded and transmitted. JPEG uses discrete cosine transform and JPEG2000 uses wavelet transform, so they produce different types of degradation. The performance of these two algorithms is compared at several coding rates and quality measure and MSE are computed for each case. These images are shown in Figures 68-81 and results are shown in Table 8.

MSE values for JPEG2000 coded images are always lower than the corresponding values for JPEG coded images at the same bit rate. However, quality measure values for

bit rates greater than 0.5 bits/pixel are slightly higher for JPEG coded images than those for JPEG2000 images. By careful inspection of these images it can be seen that JPEG images at these bit rates are slightly sharper compared to JPEG2000 coded image which contain some blurring. When the bit rate is lower than 0.5 bits/pixel, quality measure decreases rapidly for JPEG coding, but very slowly for JPEG2000 coding. This is in agreement with MSE values and visual quality which deteriorates more quickly for JPEG coding than for JPEG2000 coding.

Table 8: Comparison of JPEG and JPEG2000 coding

Rate[bits/pixel]	JPEG		JPEG2000	
	MSE	Q	MSE	Q
1.1561	9.11	0.9628	6.45	0.9255
0.5501	19.97	0.8339	13.64	0.8104
0.3647	32.85	0.6667	20.75	0.7289
0.2460	59.17	0.4806	31.27	0.6319
0.1747	120.11	0.3283	44.85	0.5868
0.1428	218.60	0.2290	54.51	0.5153
0.1129	1030.40	0.0441	67.56	0.4873

Figure 68: JPEG coding at 1.1561 bits/pixel (MSE=9.11, Q=0.9628)

Figure 69: JPEG2000 coding at 1.1561 bits/pixel (MSE=6.45, Q=0.9255)

Figure 70: JPEG coding at 0.5501 bits/pixel (MSE=19.97, Q=0.8339)

Figure 71: JPEG2000 coding at 0.5501 bits/pixel (MSE=13.64, Q=0.8104)

Figure 72: JPEG coding at 0.3647 bits/pixel (MSE=32.85, Q=0.6667)

Figure 73: JPEG2000 coding at 0.3647 bits/pixel (MSE=20.75, Q=0.7289)

Figure 74: JPEG coding at 0.246 bits/pixel (MSE=59.17, Q=0.4806)

Figure 75: JPEG2000 coding at 0.246 bits/pixel (MSE=31.27, Q=0.6319)

Figure 76: JPEG coding at 0.1747 bits/pixel (MSE=120.11, $Q=0.3283$)

Figure 77: JPEG coding at 0.1747 bits/pixel (MSE=44.85, $Q=0.5868$)

Figure 78: JPEG coding at 0.1428 bits/pixel (MSE=218.6, $Q=0.2290$)

Figure 79: JPEG2000 coding at 0.1428 bits/pixel (MSE=54.51, $Q=0.5153$)

Figure 80: JPEG coding at 0.1129 bits/pixel (MSE=1030.4, $Q=0.0441$)

Figure 81: JPEG coding at 0.1129 bits/pixel (MSE=67.56, $Q=0.4873$)

4.9. Performance of quality measure on two different images

In the last example in this chapter, the algorithm is applied on two totally different images, where one of them is the reference image and the other is the distorted image. In this case there is no similarity at all between the two images and the quality measure takes values very close to zero. These values are much smaller than any of the values obtained when the distorted image is a degraded version of the reference image. The results are given in Table 9.

Table 9: Performance of quality measure on two different images

Reference image	Distorted image	Quality measure (Q)
“Lena”	“Man”	-0.00002
“Woman”	“Barbara”	-0.00007
“Goldhill”	“Couple”	-0.0009
“Tiffany”	“Lena”	0.0005
“Lake”	“Mandrill”	0.0004

5. CONCLUSION

An algorithm for image quality assessment has been described. The algorithm uses a simple HVS model, which is used to process input images. CSF used in this model is not fixed; it has one user-defined parameter, which controls attenuation at high frequencies. This way it is possible to get better results than in the case when CSF with fixed parameters is used. This is due to the fact that HVS treats very differently high frequency components present in the original image than those of noise. Two processed images are used to compute average correlation coefficient, which measures the degree of linear relationship between two images. This way we take into account structural similarity between two images, which is ignored by MSE-based measures. Finally, image quality measure is computed as the average correlation coefficient between two input images modified by the average correlation coefficient between original image and error image. This way we differentiate between random and signal dependant distortion, which have different impact on a human observer.

Image quality measure described here depends on viewing distance and width, height and number of pixels of input images. This is reasonable because image perception by HVS also depends on these same parameters. Details of an image viewed from a close distance may not be visible when the same image is viewed from a greater viewing distance. This is modeled by CSF. As the viewing distance increases sampling

frequencies in samples/degree given by equation (8) also increase which moves image spectrum into higher frequencies. Since CSF has greater attenuation at higher frequencies this will suppress spectral components at high frequencies which are responsible for image details. The results presented here are obtained for the following parameters: viewing distance 60 cm, width of images 15 cm, height of images 15 cm and all images are 512x512 pixels.

Images in this document are not shown in their normal size. Their size was reduced so that two images can fit into one page, which significantly reduced the number of pages. Therefore, their subjective visual quality cannot be judged based on images shown here, it can only be judged when images are viewed in their normal size and when display device can display 256 different gray levels. If the display device cannot display 256 gray levels some information from the input images will be discarded and some quantization noise will be introduced. This is particularly significant if the difference between the input images is small, because then this difference can be masked by the quantization noise introduced by the display device.

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. No algorithm, no matter how sophisticated, can model something as complex as HVS.

Future research in this area should focus on finding better models for brightness perception, which will include characteristics of display device. Another possible improvement is creating a CSF, which models HVS more accurately.

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," in Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 3, pp. 1945-1948, 1989.
3. N. Damera-Venkata, T. D. Kite, W. Geisler, B. L. Evans and A. C. Bovik, "Image quality assessment based on a degradation model," IEEE Transaction 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, T. A. Michel and R. O. Hinds, "Supra-threshold perceptual image coding," in Proc. International Conf. on Image Processing (ICIP), Vol. I, pp. 237-240, 1996.
6. Z. Wang, A. C. Bovik and L. Lu, "Why is image quality assessment so difficult?," in Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol. 4, pp. 3313-3316, 2002.
7. Z. Wang and A. C. Bovik, "Universal image quality index," IEEE Signal Processing Letters, Vol. 9, No. 3, pp. 81-84, March 2002.
8. J. S. Lim: "Two-dimensional signal and image processing," Prentice-Hall, Engelwood Cliffs, New Jersey, 1990.
9. J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," IEEE Transactions on Information Theory, Vol. IT-20, No. 4, pp. 525-536, July 1974.
10. T. N. Pappas and D. L. Neuhoff, "Least-squares model based halftoning," IEEE Transactions on Image Processing, Vol. 8, No. 8, pp. 1102-1116, August 1999.