An Image Quality Assessment Scheme based on HVS using Gabor Function

Minyoung Eom and Yoonsik Choe
School of Electrical & Electronic Engineering, Yonsei University,
134 Shinchon-dong, Seodaemoon-ku, Seoul 120-749, Korea.
Tel : +82-2-2123-2774 Fax : +82-2-313-0960 E-meil : emy0606@yonsei.ac.kr, yschoe@yonsei.ac.kr

Abstract: In this paper, we propose a new image quality assessment scheme considering the human visual perception characteristics. A subjective quality assessment is obtained by the response of the receptive field in the primary visual cortex and a human's eye can't focus on all of the visual range in a moment. Take advantage of two facts above, we apply Gabor wavelet transform, which is well fit the receptive field in the cortex, to divided constant sized subblocks. Then a local distortion of the subblocks and a global distortion for the entire image are calculated in order. The proposed method has been evaluated using video test sequences provided by the Video Quality Experts Group (VQEG). The experimental results show that good correlation with human perception is obtained using the proposed metric, which is what we called GPSNR.

Keywords: Image Quality Assessment, PSNR, Gabor Wavelet, HVS

1. INTRODUCTION

Recent years have seen the introduction and rapid acceptance of digital image and video services. Most of the services rely heavily on a series of ITU standards, such as the JPEG and JPEG2000 for the still image and H.263, H.264 and MPEG-1/2/4 for the moving picture, which all have adopted the hybrid DPCM/DCT coding algorithm including quantization [1]. The objective of lossy image compression is to store image data difficultly by reducing the redundancy of image content and discarding unimportant information while keeping the quality of the image acceptable. Thus, the tradeoff in lossy image compression is between the number of bits required to represent an image and the quality of the compressed image. This is usually known as the rate-distortion tradeoff. The number of bits used to record the compressed image can be measured easily and objectively. However, the closeness between the compressed and the original images is not a purely objective measure, since human perception plays an important role in determining the fidelity of the compressed image.

The image quality degradation introduced by the digital coding algorithm differs fundamentally form the analog image quality distortions, thus calls for new measurement methods. Generally, two kinds of image quality assessment scheme are present. That is a subjective assessment and an objective assessment. The subjective assessment methods as defined in the ITU-R BT500 are still the most reliable instrument to evaluate the quality but the cost in terms of time, man power and required experience is very high. In the objective quality assessment cases, when an image is distorted from lossy image compression, the error between it and its original is typically measured by taking the average across the image of the squared pixel differences, known as the mean-squared error (MSE) or peak signal to noise ratio (PSNR). They can easily be computed to represent the deviation of the distorted image from the original image in the pixelwise sense. But it often correlates poorly with the subjectively judged distortion of the digital image because this measure has a physical and theoretical basis based on analog image. And human visual system does not process the image in a point by point fashion but extracts certain spatial, temporal, and chromatic features. Therefore, in order to obtain a good objective metric, models of the human visual system (HVS) should be taken into account.

Among the many models of the human visual system, our quality assessment scheme makes use of Gabor function which is mathematically described the spatial-response profiles of the receptive field in the cortical cells and has the property that it minimizes the joint uncertainty in a two-dimensional information space, whereupon the axes of the information space represent spatial-frequency components.

This paper organized as follows: the concepts of the HVS are discussed in section 2. Section 3 presents Gabor wavelets corresponding the receptive field. In section 4, our proposed metric for objective evalution is described and the remainders of this paper present experiments and conclusion.

2. Overview of Human Visual System (HVS)

2.1. The Visual Pathway

The computational role of the various anatomical structures along the visual pathway may be analyzed in terms of the neuronal interactions. Neurophysiologists have attempted to experimentally determine what patterns of visual stimulation (light patterns), when presented to the eye, will increase the firing of a particular groups of neurons within the neuronal layers of each anatomical structure. The visual pattern that best activates a given layer of nerve cells is called a trigger feature. The area of the photoreceptor surface

that elicits a strong response for this activated neuronal layer is called the receptive field of that neuron. Both the notions of trigger features and receptive fields are being used extensively to explain the computational behavior of the neurons in the various neuronal layers such as the retina, the LGN, the primary visual cortex, and the higher cortical regions of the visual cortex shown in Fig. 1.

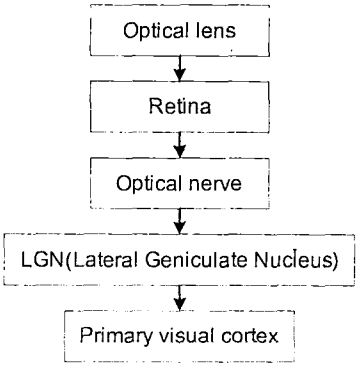


Fig. 1. A visual information transmission process of HVS

2.2. The Retina

The primary sensory system for biological vision is the retinal layer located at the back of each eye. The light stimulus from a visual object is initially focused on a layer of photoreceptor cells. The retinal receptors include approximately 120 million rods and 6million cones. The cones are sensitive to bright light, whereas the rods are sensitive to dim light. Electrophysiological studies suggest that a rod can respond to a single photon of light, whereas cones respond to a broader range of wavelengths in the electromagnetic spectrum. Thus, rods respond primarily to dim light and are good for night vision, whereas the cones respond to bright light and are good for color vision. These two types of photoreceptors have an uneven distribution in the retinal layer. The receptor density is a function of visual angle, or eccentricity. Cones are most dense in the central part of the retina, called the fovea, and the rods dominate the peripheral region of the retina.
In terms of information processing, the photoreceptors in the retina act as transducers that convert light energy to equivalent electrical signals. These signals is converted into a train of action potentials by the ganglion cells, and this train of action potentials is transmitted along the optic nerve to the visual cortex

2.3. The Primary Visual Cortex

The converted signals by the retina are transmitted to the primary visual cortex located at the rear of the brain. The primary visual cortex has three types of orientation-selective nerve cells. The first type of neuron, called simple cortical cells, responds best to bars of light projected into the retinal surface that have specific orientations. The second type, called complex cortical cells, responds strongest to correctly oriented bars of light that move over the entire receptive field. That is, most complex cells will respond to a specific direction of stimuli motion. The Third type of cell, called a hypercomplex cortical cell, responds only to visual stimuli that are moving lines with specific length of the stimuli has no effect on the response of a complex cell, it does have a direct effect on hypercomlex cells, such that if the stimulus is too long, the nerve cell will not fire.
These three types of cortical cells have broadly overlapping receptive fields respectively and visual perception is occurred by the response of the receptive fields [3][4].

3. The Mathematical Model of the Receptive Field

The localized spatial frequency decomposition due to the receptive fields may be modeled using elementary Gabor functions. The localized time-frequency analysis of arbitrary signals proposed by Gabor is based on the notion that the optimal set of basis functions for analyzing signals consists of the product of sinusoidal and Gaussian functions of time. The sinusoidal component of this function introduces a waviness, whereas the Gaussian component localizes the signal to a region in time surrounding the mean time value of the Gaussian. Gabor showed that the use of such a basis function for a signal of fixed duration will minimize the joint uncertainty associated with the product of the effective time duration of the signal and its effective bandwidth. These Gaussian weighted sinusoids are termed Gabor functions. The spatial-response profiles of the cortical cells may be mathematically described by a Gabor function given in the direction perpendicular to the optimal orientation of the neural cell. The Gabor function has the property that it minimizes the joint uncertainty in a two-dimensional information space, whereupon the axes of the information space represent spatial frequency components. There exist an infinite number of Gabor functions that can be used to model time signals. All possible types of Gabor function consist of sine and cosine waveforms with amplitudes modulated by Gaussian functions.
In 1987, Jones and Palmer demonstrated that the receptive fields in the primary visual cortex having the symmetry properties as an odd function and even function can be well fit by Gabor elementary functions[5]. According to Daugman, one suitable model of the two-dimensional (2D) receptive field profiles measured experimentally in mammalian cortical simple cells is the parameterized family of 2D Gabor filters [6]. These 2D Gabor function is a nonorthogonal wavelet, and can be specified by the frequency of the sinusoid W and the standard deviations of the Gaussian σ_x and σ_y as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}) + 2\pi jWx] \quad (1)$$

And Eq. (1) can be written in the frequency domain as

$$G(u,v) = \frac{1}{2\pi\sigma_u\sigma_v}\exp\{-\frac{1}{2}[\frac{(u-W)^2}{\sigma_u^2}+\frac{(v-W)^2}{\sigma_v^2}]\} \quad (2)$$

and their examples are shown in Fig. 2. where $\sigma_x = \sigma_y = 5$ and $W = 0.2$

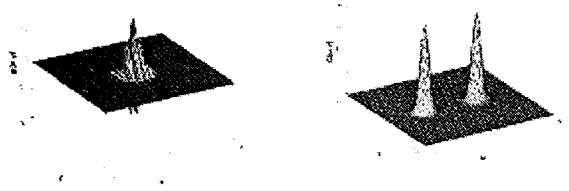
In this work, we use the version of the Gabor wavelets proposed by Manjunath and Ma [7]. The Gabor wavelets in this representation are obtained by dilation and rotation of $g(x,y)$ as above by using the generating function

$$g_{m,n}(x,y) = a^{-m}g(x',y') \quad (3)$$

where m, n is integers and

$$x' = a^{-m}(x\cos\theta + y\sin\theta)$$
$$y' = a^{-m}(-x\sin\theta + y\cos\theta)$$

and $\theta = n\pi / K$ with m and n indicating the scale and orientation, respectively. K is the total number of orientations desired and $a = (U_h / U_l)^{1/(S-1)}$ with U_h and U_l denoting the lower and upper center frequencies of interest. The scale factor a^{-m} in Eq. (3) is meant to ensure that the energy is independent of m. Examples of Gabor wavelets are shown in Fig. 3.



(a) Time domain (b) Frequency Domain

Fig. 2. 2D Gabor elementary function

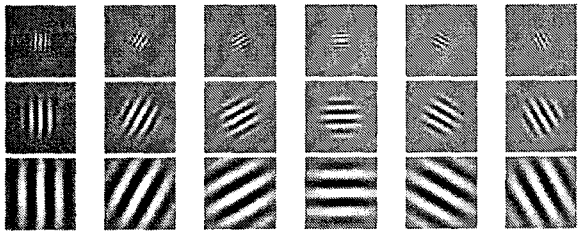


Fig. 3 Illustration of Gabor wavelets (S=3, K=6)

4. A New Image Quality Assessment Scheme

The procedure of our image quality assessment using 2D Gabor wavelet transform has the three main phases as shown in Fig.4. The First, Gabor wavelet transform applies to the divided constant sized subblocks of the original image and compressed image. The second, the sum of the squared difference of the coefficients in the subblocks calculated that is a local distortion. Lastly, we obtain the global distortion from the local distortion of the subblocks in terms of an average

In the present work, the input signal is the original image $(SI(x,y))$ and the compressed image $(PI(x,y))$ and they are divided into the subblocks of N whose size is $Fs \times Fs$. Then the $Fs \times Fs$ 2D Gabor wavelets are projected by using six scales $(S = 6)$ and four directions $(K = 4,\ 0^\circ, 45^\circ, 90^\circ, 135^\circ)$ with the lower and upper center frequencies specified as $U_l = 0.05$ and $U_h = 4.5$ cycles/pixel, respectively. The filtering process is performed in the time domain. Let sc_n^{sk} and pc_n^{sk} are the coefficients of Gabor wavelet transform of the n^{th} original and compressed subblock respectively when $S = s$ and $K = k$, then

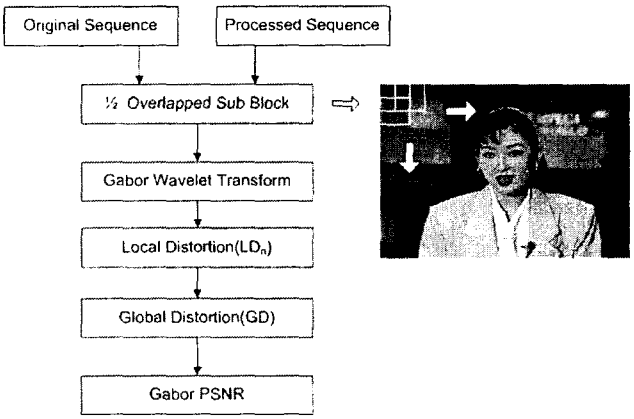


Fig. 4 Block diagram of proposed quality assessment scheme

$$sc_n^{sk} = \int_{-\frac{Fs}{2}}^{\frac{Fs}{2}}\int_{-\frac{Fs}{2}}^{\frac{Fs}{2}} SI_n(x,y)g^{sk}(x,y)dxdy \quad (4)$$

and

$$pc_n^{sk} = \int_{-\frac{Fs}{2}}^{\frac{Fs}{2}}\int_{-\frac{Fs}{2}}^{\frac{Fs}{2}} PI_n(x,y)g^{sk}(x,y)dxdy \quad (5)$$

where $1 \le s \le S$, $1 \le k \le K$, $1 \le n \le N$ and $g^{sk}(x,y)$ is Gabor filter with $S = s$ and $K = k$.

For all s and k, we can represent with the matrix form for the Gabor wavelet coefficients of the n^{th} subblock as below

$$SC_n = \begin{pmatrix} sc_n^{11} & sc_n^{12} & \cdots & sc_n^{1K} \\ sc_n^{21} & & & \\ \vdots & & \ddots & \\ sc_n^{S1} & & & sc_n^{SK} \end{pmatrix} \quad (6)$$

and

$$PC_n = \begin{pmatrix} pc_n^{11} & pc_n^{12} & \cdots & pc_n^{1K} \\ pc_n^{21} & & & \\ \vdots & & \ddots & \\ pc_n^{S1} & & & pc_n^{SK} \end{pmatrix} \quad (7)$$

Using Eq. (6)and (7), we can define a distortion matrix as below.

$$LCD_n = (DC_n)(DC_n)^T \quad (8)$$

where DC_n is

$$SC_n - PC_n = \begin{pmatrix} sc_n^{11} - pc_n^{11} & sc_n^{12} - pc_n^{12} & \cdots & sc_n^{1K} - pc_n^{1K} \\ sc_n^{21} - pc_n^{21} & & & \\ \vdots & & \ddots & \vdots \\ sc_n^{S1} - pc_n^{S1} & & \cdots & sc_n^{SK} - pc_n^{SK} \end{pmatrix} \quad (9)$$

From the distortion matrix above, a Local Distortion (LD) for the n^{th} subblock is computed as average squared difference error

$$LD_n = \frac{1}{S^2}\sum_{i=1}^{S}\sum_{j=1}^{S} LDC_n^{ij} \quad (10)$$

and a Global Distortion (GD) can be computed on the entire image as below

$$GD = \frac{1}{N}\sum_{n=1}^{N} LD_n \quad (11)$$

Finally, we can obtain the objective quality score, which is what we called GPSNR, using the global distortion above.

$$GPSNR = 10\log_{10}\frac{[\max(|sc_n^{sk} - pc_n^{sk}|, \forall(s,k,n))]^2}{GD} \quad (12)$$

5. Experiments and Results

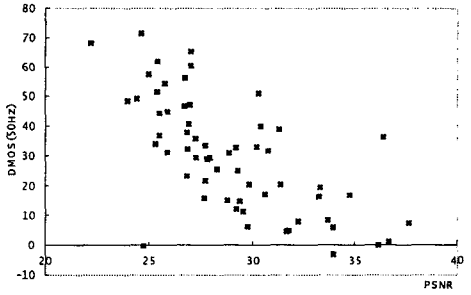
In order to evaluation the performance of the proposed objective image quality assessment scheme, experiments are conducted using the video sequences that are distributed and approved for the evaluation on performance of objective assessment models by VQEG. Table 1 and 2 show the list of the sequences [8]. And for comparison between the proposed scheme and other methods that is PSNR and EPSNR using difference image on the edge region [8], we use Pearson correlation as below.

$$\rho = \frac{Cov(OS, DMOS)}{\sqrt{Var(OS)Var(DMOS)}} \quad (12)$$

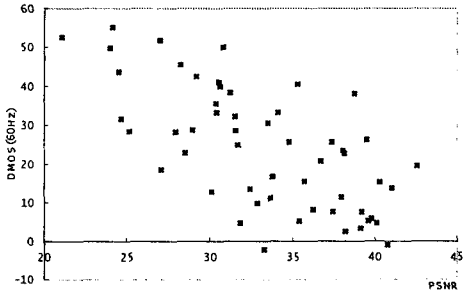
Where OS and $DMOS$ are objective score and subjective score respectively and $Cov(a,b)$ is a covariance between a and b.
And we obtain following several parameters used for Gabor wavelet transform empirically.

$$Fs = 24,\ S = 6,\ K = 4,\ U_l = 0.05,\ U_h = 4.5$$

Fig. 5,6 and 7 illustrate the scatter plots of DMOS versus the objective score respectively. Fig. 5(a) and (b) are the case of PSNR for 625/50Hz and 525/60Hz format respectively, Fig. 6 is the case of EPSNR and Fig. 7 shows the result of proposed scheme. As can be seen in the figures, the proposed objective quality assessment scheme gives much better correlation between objective scores and DMOS in terms of the correlation coefficient than PSNR and EPSNR. Especially, for 625/50Hz format, our method has the better performance than other one.

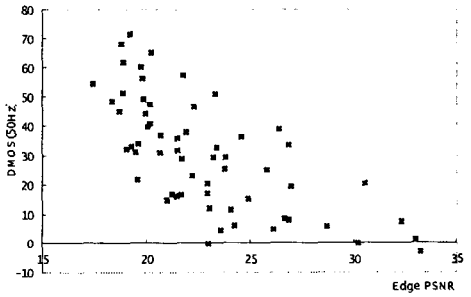


(a) ρ = -0.675107

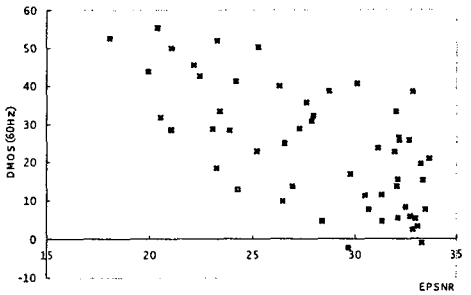


(b) ρ = -0.66278

Fig. 5. The scatter plots of DMOS vs PSNR. (a) is 625/50Hz format and (b) is 525/60Hz format

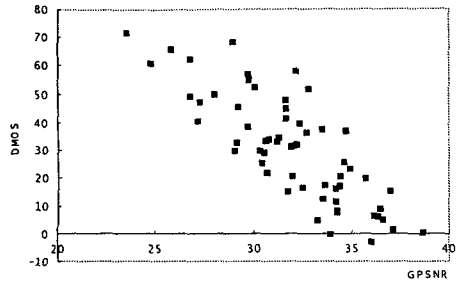


(a) ρ = -0.683964

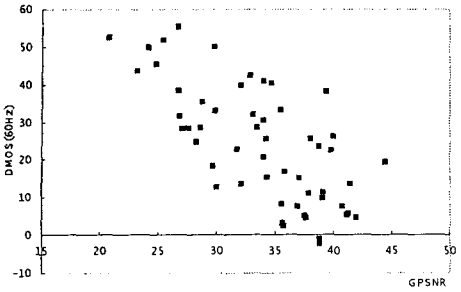


(b) ρ = -0.662757

Fig. 6. The scatter plots of DMOS vs EPSNR. (a) is 625/50Hz format and (b) is 525/60Hz format



(a) ρ = -0.81254



(b) ρ = -0.702852

Fig. 7. The scatter plots of DMOS vs GPSNR. (a) is 625/50Hz format and (b) is 525/60Hz format

Table 1. 625/50Hz format test sequence

Assigned number	Sequence	Characteristics	Source
1	Tree	Still, different direction	EBU
2	Barcelona	Saturated color + masking effect	RAI/ Retevision
3	Harp	Saturated color, zooming, highlight, thin details	CCETT
4	Moving graphic	Critical for Betacam, color, moving text, thin characters, synthetic	RAI
5	Canoa Valsesia	water movement, movement in different direction, high details	RAI
6	F1 Car	Fast movement, saturated colors	RAI
7	Fries	Film, skin colors, fast panning	RAI
8	Horizontal scrolling 2	text scrolling	RAI
9	Rugby	movement and colors	RAI
10	Mobile & calendar	available in both formats, color, movement	CCETT

Table 2. 525/60Hz format test sequence

Assigned number	Sequence	Characteristics	Source
13	Baloon-pops	film, saturated color, movement	CCETT
14	NewYork 2	masking effect, movement	AT&T/ CSELT
15	Mobile& Calendar	available in both formats, color, movement	CCETT
16	Betes_pas_ betes	color, synthetic, movement, scene cut	CRC/CBC
17	Le_point	color, transparency, movement in all the directions	CRC/CBC
18	Autumn_ leaves	color, landscape, zooming, water fall movement	CRC/CBC
19	Football	color, movement	CRC/CBC
20	Sailboat	almost still	EBU
21	Susie	skin color	EBU

Table 3. Test condition (HRCs)

ASSIGNED NUMBER	A	B	BIT RATE A B	RES	METHOD	COMMENTS
16	X		1.5 Mb/s	CIF	H.263	Full Screen
15	X		768 kb/s	CIF	H.263	Full Screen
14	X		2 Mb/s	3/4	mp@ml	This is horizontal resolution reduction only
13	X		2 Mb/s	3/4	sp@ml	
12	X		4.5 Mb/s		mp@ml	With errors TBD
11	X		3 Mb/s		mp@ml	With errors TBD
10	X		4.5 Mb/s		mp@ml	
9	X	X	3 Mb		mp@ml	

6. Conclusion

In this paper, we proposed a new image quality assessment scheme based on HVS using Gabor function which is well fit the receptive field in the primary visual cortex. Also we used the fact that human's eye can't focus on all of the visual range in a moment. The performance of the proposed model has been evaluated with the sequences provided by VQEG. It is demonstrated that the new method achieves high correlation with human perception compared with PSNR and EPSNR.

Reference

[1] K. R. Rao, J. J. Hwang, " Techniques and Standards for image, Video, and Audio Coding, " Prentice Hall PTR, 1996.

[2] S. Comes and B. Macq, "Human Vision Quality Criterion, " in SPIE Visual Communications and image Processing, Vol. 1360, pp. 2-7, 1990.

[3] Madan M. Gupta and George K. Knopf, "Neuro-Vision Systems : A Tutorial, " IEEE Press, 1993.

[4] Yao Wang, Jorn Ostermann and Ya-Qin Zhang, "Video Processing and Communications," Prentice Hall, pp.1-12, 2002.

[5] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter medel of simple receptive fields in cat striate cortex," J. Neurophysiol., vol. 58, pp. 1233-1258, 1987.

[6] John G. Daugman, "Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression," IEEE Trans. Acoust., Speech, Sinal Processing, vol. 36, no. 7, pp. 1169-1179, July 1988.

[7] Manjunath BS and Ma WY, "Texture features for browsing and retrieval of image data," IEEE transactions on Pattern Analysis and Machine Intelligence, 18(8):837-842, 1996.

[8] Video Quality Experts Group, "Final Report from the Video Quality Experts Group on the Validation of Objective models of Video Quality Assessment, Phase II(FR-TV2)", Aug. 2003.