A Wavelet Approach to Compressed Image Quality Measurement*

Yung-Kai Lai, Jin Li, and C.-C. Jay Kuo
Integrated Media Systems Center
Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, CA 90089-2564

Abstract

The understanding and characterization of various forms of image degradation due to compression is important. The most commonly used image distortion measure is the mean squared error (MSE). However, the MSE value does not provide a good subjective performance measure, since it does not take human perception into account. In this research, we present a wavelet approach to measure the visual quality of compressed images based on a psycho-visual human vision model. Wavelet filter banks are shown to have a good space-frequency localization property and can be directly linked to the Michelson contrast and compactly supported features. Experiment results are provided to demonstrate the effectiveness of the Haar wavelet.

1. Introduction

Various compression schemes have been used in lossy image compression. Different schemes produce different compression artifacts such as the blocking artifact for block-based compression methods and the ringing artifact for frequency-based compression methods. Although they behave quite differently in visual appearances, there were very few subjective metrics proposed to measure the visual performances of compressed images. The most commonly used image distortion measure so far is still the mean squared error (MSE) or peak signal-to-noise ratio (PSNR), which computes the pixel-by-pixel difference between the original and the compressed images. However, these objective metrics do not take human perception into account, and do not provide a strong correlation of human visual experience.

To propose a subjective metric for image quality analysis, human perception models, which have been extensively studied in psychophysic science, should be incorporated. Some efforts have been made to incorporate Human Visual System (HVS) models into subjective metrics in 70's and 80's [5], [6], [9], [10]. However, the development of HVS models back then was still immature such that these models could not interpret human visual perception very well. Recently, the development of new image compression techniques became much more faster then before, and psychophysical researches also lead to more consistent HVS models. New distortion measures were therefore proposed with better prediction of image visual quality. Karunasekera [7] proposed a distortion measure for blocking artifacts, while van den Branden Lambrecht [1], [13] proposed a more complete model and extended its use to compressed video sequences.

The major problem of the basic HVS model, however, is that the bandpass filtering and localization property can not be satisfied simultaneously according to the Principle of Uncertainty. Although it was shown that Gabor filtering is the optimal solution, it remained unspecified how the Gaussian parameters of the Gabor envelope should be chosen. In addition, it is difficult to obtain the contrast from the frequency responses of individual channels. In this paper, we proposed a wavelet-based approach which solves both problems while still possesses good space-frequency localization property.

This paper is organized as follows. In Section 2, we introduce the use of Haar basis wavelets, our proposed HVS model, and image quality assessment system. Experimental results are shown in Section 3, where the performance of Haar wavelet filtering are shown in comparison with that of the commonly-used Gabor filtering.

2. The HVS model and wavelet approach

2.1. The HVS model and compression artifacts

The visual mechanism of human beings concentrates on light changes of the surroundings. Contrast, therefore, is used to discriminate between objects or/and the background.

*This work was supported by Intel and the Integrated Media Systems Center, a National Science Foundation Engineering Research Center.

Numerous experiments have been conducted in the psychophysics field based on human visual perception of contrasts. Experimental results showed that the human visual system is comprised by many units, each of which is sensitive to the contrast in certain frequency bands and functioning independently. The overall visual perception of object luminance or contrast is therefore the aggregate performance of each cell's frequency response [2]. Since human eyes cannot provide indefinite luminance resolution, a contrast threshold can be found as a function of spatial frequency via experiments, and the subthreshold (contrast which is lower than the threshold) stimuli cannot be sensed by human beings. (although they can be in certain circumstances, which will be described below as summation effect.) A typical contrast sensitivity (defined as the reciprocal of the contrast) curve is shown in Fig. 1 [2], which shows that the HVS has the highest luminance resolution around 3-10 cycles per degree.

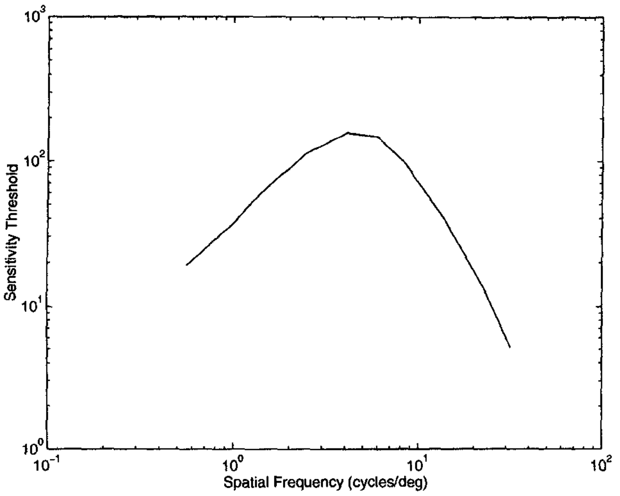

Figure 1. A typical contrast sensitivity curve of human beings.

Based upon this frequency model, the visual perception of compression artifacts is the direct result of band-limited cell perception. When an image is compressed and then decompressed, the decompressed image will have a difference frequency response than the original image as long as pixel differences exist. Visual artifacts will be perceived as frequency anomaly in certain regions of the decompressed image. In detail, They can be interpreted with this localized frequency model as follows:

- Overall smoothness
  The overall smoothness, which appears as edge smoothness or texture blur, is due to the loss of high frequency components. It will be perceived as the lack of high frequency components when compared with the original image.
- Blocking effect
  The blocking effect, which appears in all block-based compression techniques, is due to the coarse quantization of frequency components. It can be observed as surface discontinuity (edge) at block boundaries. These edges are perceived as abnormal high frequency components in the spectrum.
- Ringing effect
  The ringing artifact is observed as periodic pseudo-edges around original edges. It is due to improper truncation of high frequency components. It is natural to interpret this artifact from the frequency viewpoint since this artifact mainly happens in frequency-based compression techniques.
- Texture deviation
  This artifact is observed as some granular noise or "dirty window" in texture regions in model-based or segmentation-based compression schemes. It is due to the loss of fidelity of mid-frequency components, and is perceived as frequency inconsistency in textured regions. It can be expressed as frequency anomaly in the compressed image.

The traditional objective error measures, such as PSNR, mainly focused on the spatial domain pixel differences. The perceptual error measures, however, put emphasis on frequency domain mismatch between two images. Since the HVS channels are also localized in the space domain, we expect objective error measures to provide prediction of the visual quality to a certain degree, but a good perceptual quality assessment system should take frequency domain factors into account.

2.2. Wavelet approach

It is widely accepted that human spatial vision can be represented by a localized multiple spatial frequency channel model descrbed above. Since the whole visual system is synthesized by many spatial frequency selective channels each narrowly tuned to different locations of the vision field and different frequency bands, we can decompose the visual stimuli into their spatial frequency components and examine individual channel responses. Unfortunately, the commonly-used Fourier analysis is a global process, thus the channel localization property cannot be justified. Although it was shown that the Gabor function has the optimal combination of frequency and spatial domain localization, the Gaussian envelope parameters are left unspecified. Furthermore, since Gabor filters are essentially still IIR filters, proper filter length has to be carefully chosen to keep the localization property in both domains. In addition, the

HVS is characterized by the contrast curve as a function of spatial frequencies. The contrast C, defined by Michelson as

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}} \tag{1}$$

where L_{max} and L_{min} are the maximum and minimum luminances of the waveform around the point of interest, cannot be directly obtained by Gabor filtering process. Peli [11] tried to redefine the contrast in Fourier domain by dividing the lowpass band response by the highpass band response using cosine filters, but neither can this new definition be directed linked to (1).

On the other hand, wavelet approaches were also proved to have good space-frequency localization property. The passbands are well defined, and many compactly supported wavelets are available. Another advantage of wavelet transform is that the wavelet packets can successively decompose the signal to any band of required bandwidth. Furthermore, with Haar wavelet, whose base scale filter coefficients are given as

$$h_0[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{n=-1,0,} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$h_1[n] = \begin{cases} \frac{1}{\sqrt{2}} & \text{n=0,} \\ -\frac{1}{\sqrt{2}} & \text{n=-1,} \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

it is easily seen that the contrast C in (1) can be represented by the responses at the i^{th} scale or resolution:

$$C_i = \frac{H_i}{L_i}, \tag{4}$$

where H_i and L_i are lowpass and highpass responses of the i^{th} band, respectively. With (4), we can define the contrast of any pixel on the image at arbitrary resolution. Multiresolution analysis also facilitate us to deal with the retinal inhomogeneity phenomenon, which describes that the frequency resolution of foveal and peripheral vision decreases with increasing eccentricity from the fixation point [3], [8]. Dyadic wavelet transform provides a good tool for this analysis since it is shown that each visual frequency channel has the bandwidth of about 1 octave [11], and the equal-sensitivity neighborhood radius is inversely proportional to the spatial frequency [3].

2.3. Proposed image quality assessment system

With this wavelet approach, the proposed HVS model is described as follows:

1. Suprathreshold contrast perception model
   Visible artifact occurs when the contrast of the artifact exceeds the threshold. The suprathreshold perception model consists of three stages [4]:

   (a) A linear spatial filter with gain G:

   $$r = GC, \quad \text{where } C \text{ is the actual contrast.}$$

   (b) A static, nonlinear transducer:

   $$R' = (r - r_0)^m,$$

   where r_0 is the response of the contrast threshold.

   (c) A linear normalization stage:

   $$R = \frac{R'}{R_{norm}},$$

   where where R_{norm} is obtained by introducing a normalized contrast into the first two stages in order to consider the contrast constancy phenomenon.

2. Subthreshold summation effect
   Spatial summation theory showed that the neighboring frequency channels contribute to the total contrast threshold. Therefore, a subthreshold contrast may still produce a small response if there exist other excitory stimuli in nearby frequency channels. This effect can be modeled as a contrast threshold C_T decrease [8]:

   $$C_T = \left(\sum_{i=0}^{N} |C_{Ti}|^{s_i} \right)^{\frac{1}{s_i}}, \tag{5}$$

   where C_{Ti} is the contrast sensitivity threshold of the i^{th} closest channel, N is the total neighboring channels affecting the perception of the target channel, C_{T0} represents the original contrast without summation, and s_i's are the exponential component to be determined.

3. Suprathreshold masking effect
   Another inter-channel interaction is the masking effect. The visibility of a signal at some frequency could be impaired by the presence of other stimuli in nearby frequency channels. One well-known example in compressed image artifact analysis is that the blocking artifact of block-based compression schemes, which consists high-frequency edge components, is less visible in the mid-to-high-frequency-based texture regions. This effect could be modeled as a exponential function of both their frequency and contrast differences [1], [12]:

$$C_T = C_{T0}\left(\sum_{i=0}^{N}\left(\frac{C_i}{C_{T0}}\right)^{m_{1,i}}\left(\frac{f_i}{f_0}\right)^{m_{2,i}}\right) \quad (6)$$

where C_{Ti}, C_{T0}, N have the same definition as in the subthreshold summation case, and C_i and f_i are the actual contrast and spatial frequency of the i^{th} closest channel, respectively. $m_{1,i}$'s and $m_{2,i}$'s are the coefficients to be determined.

The system block diagram is shown in Fig. 2. First the contrast sensitivity threshold C_{T0} in each band is adjusted by the summation and masking effects. The C_{T0}'s will be then fed into the suprathreshold perception model to compute the suprathreshold (i.e. perceived) contrast response. The perceptual distortion $D(i,j)$ at pixel (i,j) of the image is then the weighted sum of frequency mismatch in each band k:

$$D(i,j) = \sum_k a_k |R_o(i,j) - R_c(i,j)|^t \quad (7)$$

where R_o and R_c are the perceived contrast responses of the original and compressed images, respectively. The total distortion will be the sum of distortion of each pixel in the image.

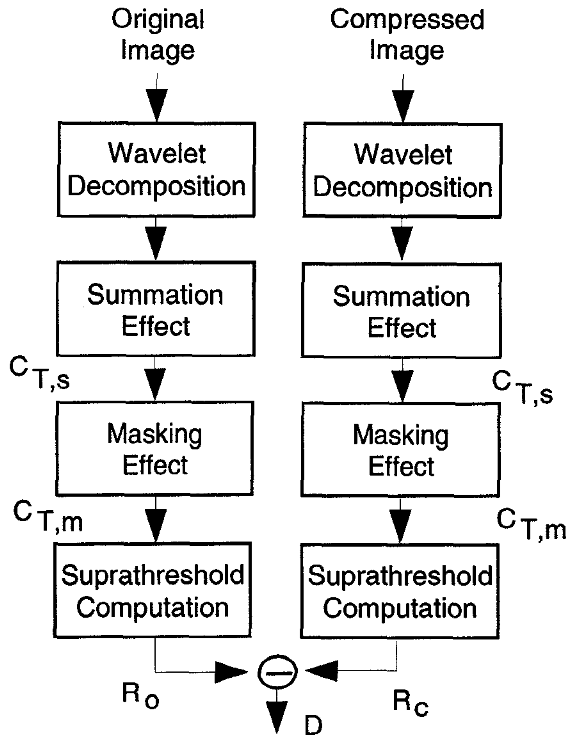

Figure 2. Block Diagram of the proposed quality measurement system.

3. Experimental results

We applied the Haar wavelet family (Eqns. (2) (3)) to contrast sensitivity experiments. The experiments were conducted on a 17" Silicon Graphics GDM-17E11 color graphic display. The luminance range of the display was adjusted from 0 to 80 cd/m^2 (candela/squared meter) using a Photoresearch spectroradiometer. 256 discrete gray scales were presented during the experiments. The curve of luminance versus gray scale is shown in Fig. 3. This curve was used to compute the actual contrast in the experiments.

Both Gabor and Haar filtered patches were used in the experiments. The spatial frequency range of the test patches is from 0.069 to 19.2 cycles per degree, which covers virtually the whole frequency band we would sense from digital images. The result is shown in Fig. 4, where the sensitivity threshold, which is defined as the reciprocal of the contrast threshold, was plotted as a function of spatial frequency. The closeness of these two curves showed that Haar filtering has comparable performance to that of Gabor filtering, and demonstrated that the Haar filters can be used in the quality assessment system.

Subsequent experiments to determine the system parameters are currently under way.

4. Conclusion

In this work, we proposed a wavelet approach to model human visual system. This approach has the advantage of space-frequency localization, ability to compute the contrast at arbitrary resolution, and compactly supported filtering. It was demonstrated in experiments that the performance of Haar filtering is comparable to that of Gabor filtering. A integrated quality evaluation system was outlined and further experiments will be conducted to show its quality evaluation ability.

References

[1] D. Costantini, C. J. van den Branden Lambrecht, G. L. Sicuranza, and M. Kunt. Motion rendition quality metric for MPEG coded video. In *Proceedings 1996 IEEE International Conference on Image Processing*, pages 889–892, Sept. 1996.

[2] R. L. De Valois and K. K. De Valois. *Spatial Vision*. Oxford University Press, 1988.

[3] M. A. García-Pérez. The perceived image: Efficient modelling of visual inhomogeneity. *Spatial Vision*, 6(2):89–99, 1992.

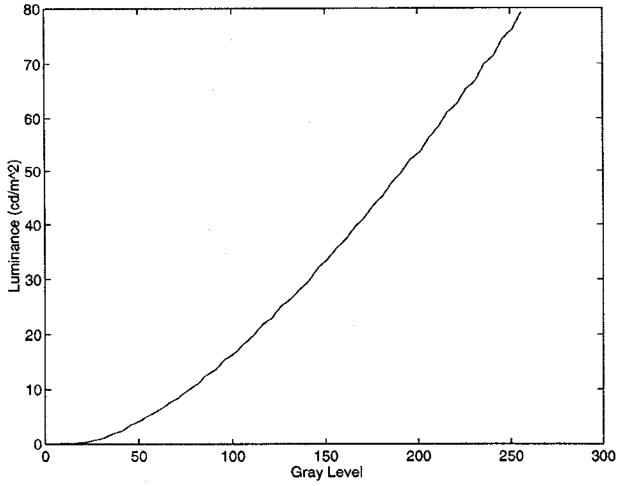

Figure 3. Luminance versus gray level for the display used in the experiments.

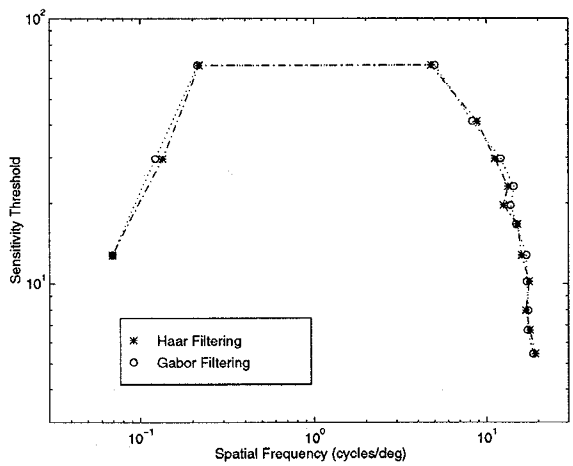

Figure 4. Contrast sensitivity threshold of Gabor and Haar filtering.

[4] M. A. Georgeson. Contrast overconstancy. *J. Opt. Soc. Am. A*, 8(3):579–586, Mar. 1991.

[5] D. J. Granrath. The role of human visual models in image processing. *Proceedings of the IEEE*, 69(5):552–561, May 1981.

[6] C. F. Hall and E. L. Hall. A nonlinear model for the spatial characteristics of the human visual system. *IEEE Trans. on SMC*, 7(3):161–170, Mar. 1977.

[7] S. A. Karunasekera and N. G. Kingsbury. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE Trans. on Image Processing*, 4(6):713–724, June 1995.

[8] D. H. Kelly. Retinal inhomogeneity: I. spatiotemporal contrast sensitivity. *J. Opt. Soc. Am. A*, 1(1):107–113, Jan. 1984.

[9] F. X. J. Lukas and Z. L. Budrikis. Picture quality prediction based on a visual model. *IEEE Trans. on Communications*, 30(7):1679–1692, July 1982.

[10] H. Marmolin. Subjective MSE measures. *IEEE Trans. on SMC*, 16(3):486–489, June 1986.

[11] E. Peli. Contrast in complex images. *J. Opt. Soc. Am. A*, 7(10):2032–2040, Oct. 1990.

[12] J. P. Thomas. Indepedent processing of suprathreshold spatial gratings as a function of their separation in spatial frequency. *J. Opt. Soc. Am. A*, 6(7):1102–1111, July 1989.

[13] C. J. van den Branden Lambrecht. A working spatio-temporal model of the human visual system for image restoration and quality assessment applications. In *Proceedings 1996 International Conference on Acoustics, Speech, and Signal Processing*, pages 2293–2296, May 1996.