VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

A NOVEL PSYCHOVISUAL MODEL ON A STANDARD RESOLUTION FOR VIDEO COMPRESSION

N. A. Abu, A. S. H. Basari and B. Hussin
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Durian Tunggal, Melaka, Malaysia
E-Mail: nura@utem.edu.my

ABSTRACT

A psychoacoustic model is well established in an audio coding. A psychoacoustic quite threshold is a basic tool on audio coding. This paper investigates on a quantitative experimental impact on almost lossless image compression based on the concept of psychovisual threshold. This paper proposes a practical image coding on a standard resolution two-dimensional discrete cosine transform (DCT). A psychovisual model is presented following the psychoacoustic quite threshold as a just noticeable difference instead of a quantization table. An experimental result on small and large rectangular images shall be presented. This framework can easily produce higher visual quality images at a competitive compression rate in comparison to an extended adaptive JPEG compression standard.

Keywords: psychovisual threshold, just noticeable difference, video coding.

INTRODUCTION

A video consists of a sequence of images. A small raw video of 320 pixels by 240 pixels in 30 frames per second has a data rate at more than 50 Mbit/s. At such high data rate, a video compression is certainly called for.

Current popular image and video coding schemes operate on a small square image block. These frameworks are basically due to a much lower processing power from the previous century. A large high quality image or video frame is typically rectangular. A visual perception on a display resolution can be affected by the display screen's rectangular shape, which is expressed in aspect ratio. The traditional popular ratio is 4:3.

An independent video frame is encoded using only information present in the frame itself. It typically uses only transform coding which provides moderate compression of two bits per pixel per colour channel.

Traditionally, each i-frame is a standalone frame. A video frame is divided into 8 by 8 or 16 by 16 image blocks and then transformed into a frequency domain on each small block. It is an efficient mode of coding. It is even faster to do it on 4 by 4 image block [1].

A typical encoded block consists of 16 by 16 luminance (Y) and two 8 by 8 chrominance (Cb and Cr) components. In MPEG-4, each macroblock's 16 by 16 luminance may be partitioned into 16 by 16, 16 by 8, 8 by 16, and 8 by 8 as discussed in [2].

Previous research mainly focuses on a small block by applying a discrete orthonormal transform on which an optimal quantization table is assigned to the frequency image signals. Recently, the employment of a psychovisual threshold has been investigated on an 8 by 8 JPEG baseline coding [3, 4].

Traditionally, 8 by 8 block transforms have been used for image and video coding. The 8 by 8 size has the advantages of being dyadic, large enough to capture trends and periodicities while being small enough to minimize spreading effects due to local transients over the transform area [5].

In this paper, a novel psychovisual model is proposed and practically tested on a small rectangular image block. One of the major drawbacks of the block-based DCT compression methods is that it may result in visible artifacts at block boundaries. An adaptive post-filtering algorithm has been proposed in [6] to remove coding artifacts in block-based DCT compressed images. A presence of blocking artifacts may be detected using block activity based on human visual system and block statistics.

A psychovisual model is presented following the psychoacoustic quite threshold as a quantization threshold. An experimental result on small 6 by 8 from a large 240 by 320 image frames shall be presented.

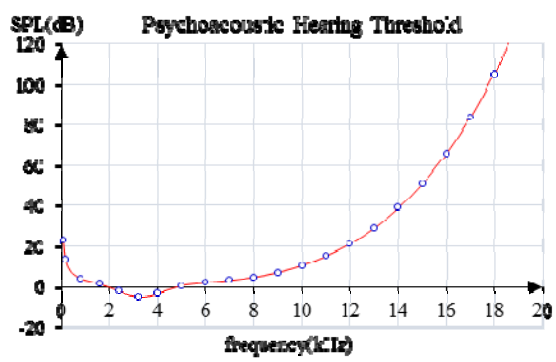

Figure-1. An absolute hearing threshold across audio spectrum.

A Psychoacoustics hearing threshold

The field of psychoacoustics has made significant progress toward characterizing human auditory hearing threshold and particularly the time-frequency analysis capabilities of the inner ear. Nevertheless, the same advancement has not been achieved in the psychovisual model. At the same time, an experiment on a psychovisual threshold is subjective to the frequency transform being applied on the image signals.

VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

This paper will extend on the concept of scalable psychovisual threshold from an adaptive JPEG image compression to intra and background frames on video compression. The concept of the psychovisual error threshold can be obtained from quantitative experiments by evaluating the just noticeable difference (JND) of the compressed image from the original image at various frequency orders [2].

The absolute threshold of hearing maps the minimum amount of energy needed in a pure tone such that it can be detected by a young sensitive listener at various audible frequencies. The quiet (absolute) threshold [7] is well approximated by the nonlinear function,

$$T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4 \text{dB} \quad (1)$$

As depicted in Figure-1. In a high fidelity audio compression, the inaudible frequency signal below this energy threshold will not be encoded at all. While the threshold curve captured in Figure-1 is associated with pure tone sound, the same concept is not well established in a digital image processing. The curve is often referred to in an audio coding system by equating the lowest point (i.e., near 4 kHz) to the energy in 1 bit of signal amplitude. The same concept has been applied here on a psychovisual threshold.

First, a psychovisual curve as a function of frequency order is proposed. Second, a psychovisual experiment will be conducted to generate the contour map from DCT to establish the level of 1-bit reconstruction error. Third, a set of parameters is chosen on the luminance and chrominance psychovisual curves as JND curves. Lastly, the performance of these noticeable visual thresholds is tested on real and graphical images.

A Psychovisual threshold on discrete cosine transform

Traditionally, the DCT provides a good support on a small image block [8, 9]. Thus, the DCT also has energy compactness properties on a large image. For a given set of input value $f(x, y)$ known as image intensity values of sizes M by N, the forward DCT of order $p + q$ is given as follows:

$$T_{pq} = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} \alpha_p \cos\frac{\pi(2x+1)p}{2M} \cdot f(x,y) \cdot \beta_p \cos\frac{\pi(2y+1)q}{2N}, \quad (2)$$

for $p = 0, 1, 2, ..., M-1$ and for $q = 0, 1, 2, ..., N-1$ where $f(x, y)$ denotes the intensity value of an image at the pixel position x and y.

The frequency coefficients are

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \frac{\sqrt{2}}{\sqrt{N}}, & p > 0 \end{cases} \quad \text{and} \quad \beta_q = \begin{cases} \frac{1}{\sqrt{M}}, & q = 0 \\ \frac{\sqrt{2}}{\sqrt{N}}, & q > 0 \end{cases} \quad (3)$$

The inverse of DCT on a perfect image reconstruction from its frequency coefficients is given as follows:

$$\tilde{f}(x,y) = \sum_{p=0}^{M-1}\sum_{q=0}^{N-1} \alpha_p \cos\frac{\pi(2x+1)p}{2M} \cdot T_{pq} \cdot \beta_p \cos\frac{\pi(2x+1)q}{2N}, \quad (4)$$

for $x = 0, 1, 2, ..., M-1$, and $y = 0, 1, 2, ..., N-1$, where $\tilde{f}(x, y)$ denotes the reconstructed intensity value and $M+N$ denotes the maximum frequency order being used.

Initially, an input RGB image is converted into YUV colour space. An input image is a raw M by N block of image pixels. In the previous work, the 8 by 8 DCT psychovisual threshold has been generated for image compression[3]. The same technique shall be applied here to generate a new 6 by 8 DCT psychovisual threshold.

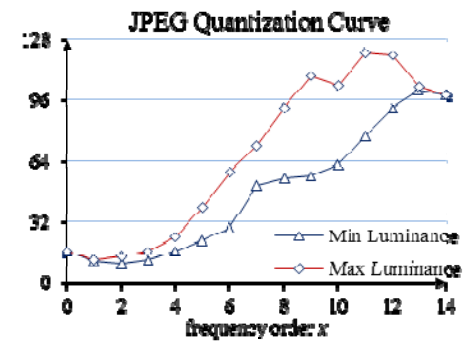

Figure-2. The luminance quantization curves from classic JPEG compression give a good impression on a psychovisual threshold.

A noticeable visual threshold

A quantization table approximates the sensitivity and critical capabilities of the human visual system. The irrelevant image information is removed by the quantization table without any discretion. An adaptive JPEG quantization step does not take the rate of distortion into consideration. Most of the time, the quantization process will limit the scope of the high frequency signals and discard information which is not visually significant for better compression.

Table-1. Standard JPEG quantization values for luminance channel.

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-2. Standard JPEG quantization values for chrominance channel.

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

In general, a human visual system is more sensitive to the low frequency signals than to the high frequency signals. Nevertheless, it is discovered in this research, a human visual system is very sensitive to the far end high-frequency signals. Traditionally, it is also well known that the young ears are more sensitive to the high frequency sound. However, they do not have the buying power in the music industry. As shown in Figure-1, the audio hearing threshold is not going down on the far end of high frequency audio signals in favor of gaining higher compression rate.

The sensitivity threshold of the human visual system at the very low order signals has hardly been taken into consideration. It is also noted that the standard luminance curve goes down slightly on low frequency signals as the counterpart in psychoacoustic threshold as depicted in Figure-1 between 2 kHz and 5 kHz frequency signals. This region is the most sensitive frequency signals to a human hearing system. A person can hear a whisper at a very low energy level. The classic JPEG luminance quantization curve follows the same concept as shown in Figure-2, the curve goes down below 16 within frequency order 1 and 3.

Similarly, human visual system is perceptually sensitive to certain high frequency image signals. For instance, human visual system is naturally highly sensitive to the smoothness of human skin. This effect has been modestly taken care by luminance JPEG quantization table only as the quantization values go down slightly at frequency order 13 and 14 in Figure-2.

The JPEG quantization tables have been known to offer satisfactory performance, on the average, over a wide variety of applications and viewing conditions. The traditional JPEG quantization tables for luminance and chrominance are shown in Table-1 and Table-2. These are the most commonly used quantization tables as they are published by the Independent JPEG Group in 1998. Since then there has been considerable progress in determining an optimal quantization value at various frequency orders.

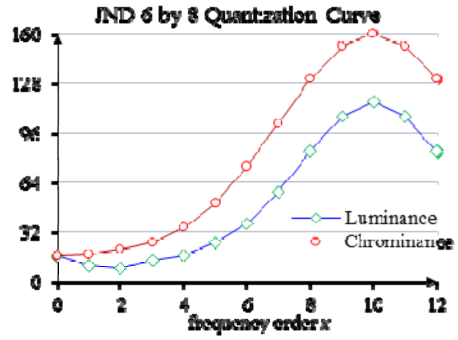

Figure-3. The quantization curves follow the concept of a just noticeable difference on a 6 by 8 image block.

A scalable JND curve on 6 by 8 DCT

This section will show how the JND quantization curves can be applied to a small rectangular DCT block. For a comparison purposes, a 6 by 8 block is chosen. These JND luminance and chrominance curves which represent the visual sensitivity of DCT basis function are given below respectively.

$$Q_L(x)=\begin{cases} 2^g - A_{L_0}\dfrac{x^2}{s_{L_0}^2}e^{-\frac{x^2}{s_{L_0}^2}}, & 0\le x\le \dfrac{M}{2} \\ 2^g + \dfrac{A_{L_1}}{s_{L_1}}e^{-(\frac{x-\mu}{s_{L_1}})^2}, & \dfrac{M}{2}<x<M+N-1 \end{cases}$$

$$Q_C(x)=2^g+\frac{A_{C_1}}{s_{C_1}}e^{-\frac{(x-\mu)^2}{s_{C_1}^2}},\quad 0\le x<M+N-1 \tag{5}$$

A sample set of parameters of the JND quantization curves on a 6 by 8 DCT is presented in Table-3. These new 6 by 8 JND psychovisual curves shall be used as the quantization values. The quantization tables are presented in Table-4 and Table-5 on luminance and chrominance respectively. They can be compared to traditional JPEG quantization tables for luminance and chrominance are shown in Table-1 and Table-2.

Table-3. A sample set of parameters of the JND on a 6 by 8 block quantization tables.

g	A_{L_0}	A_{L_1}	A_{C_1}
4	24	360	604
μ	s_{L_0}	s_{L_1}	s_{C_1}
10	1.6	3.6	4.2

The selection of parameters especially on A_{L_1} and A_{C_1} has been done so that the quality measure is equivalent to the standard JPEG image quality [10]

VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

according to the Absolute Reconstruction Error(ARE) near 3 and Self-Similarity Index Measure (SSIM) near 0.986 as highlighted in Table-6 and Table-7. Higher parameters will result in better compression at the expense of a lower quality.

The JND luminance curve is plotted in Figure-3 together with the JND chrominance curve for every frequency order. The red chrominance curve has been set consistently higher than the blue luminance curve. This is due to the fact that human eye is more sensitive to the change in luminance channel than the chrominance channel along the frequency order. Human contrast sensitivity is in general skewed to-the-left. Thus, JND is suitable to set higher quantization values at higher frequency signals.

Table-4. JND quantization values for the luminance channel.

16	10	8	13	16	25	37	58
10	8	13	16	25	37	58	84
8	13	16	25	37	58	84	107
13	16	25	37	58	84	107	116
16	25	37	58	84	107	116	107
25	37	58	84	107	116	107	84

Table-5. JND quantization values for the chrominance channel.

16	17	20	25	35	51	74	102
17	20	25	35	51	74	102	131
20	25	35	51	74	102	131	152
25	35	51	74	102	131	152	160
35	51	74	102	131	152	160	152
51	74	102	131	152	160	152	131

The new smooth psychovisual thresholds for JND luminance and chrominance are mathematically defined by Eqn. (5). The variable g refers to the number of bit gain by the transform being used. Similar to the case of JPEG compression, the DCT on an 8 by 8 image block has gained g=4 bits. The luminance quantization Table-1 starts from 2g=24=16 at frequency order 0+0=0. The JND quantization curves can also be applied to a 6 by 8 rectangular block. The respective value of the bit gain g, however, may need further accurate investigation.

Table-6. The image reconstruction quality coming out of JND curves standard JPEG quantization process at high medium and low quality factors from 48 real images.

Quality Factor	JND	Q_3=75	Q_2=50	Q_1=25
ARE	2.9670	2.3356	2.9132	3.6293
MSE	25.8670	19.1974	25.3448	32.6212
PSNR	34.6265	36.4906	34.8771	33.4095
SSIM	0.9859	0.9912	0.9864	0.9793

Table-7. The image reconstruction quality coming out of JND curves standard JPEG quantization process at high medium and low quality factors from 48 graphical images.

Quality Factor	JND	Q_3=75	Q_2=50	Q_1=25
ARE	3.3034	2.0637	2.9077	3.7619
MSE	28.4942	16.9419	23.8278	30.3702
PSNR	34.6041	37.4733	35.4657	34.0879
SSIM	0.9859	0.9934	0.9863	0.9773

The design process on the JND quantization curves in this paper is slightly biased toward compression rate. A software developer may easily set the parameters higher to gain higher compression rate. As highlighted in the left-most column of Table-6 and Table-7, these JND quantization curves have been set slightly lower quality than standard JPEG quantization tables. The results on the average bitrate of Huffman codes in comparison to standard JPEG at Q2=50, these JND quantization curves has on average produced slightly lower bitrate, in both real images in Table-8 and graphical image in Table-9, on the chrominance channels ACU and ACV.

Table-8. The average bit length of Huffman codes coming out of JND curves and standard JPEG quantization process at high, medium and low quality factors from 48 real images.

Quality Factor	JND	Q_3=75	Q_2=50	Q_1=25
DCY	4.6196	5.3166	4.4192	3.5458
DCU	2.8776	3.4715	2.7099	2.0185
DCV	2.5545	3.1597	2.4427	1.8319
ACY	1.5687	1.5353	1.5593	1.5623
ACU	1.5433	1.6263	1.5994	1.5460
ACV	1.5338	1.6048	1.5781	1.5131

VOL. 11, NO. 5, MARCH 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Table-9. The average bit length of Huffman codes coming out of JND curve and standard JPEG quantization process at high, medium and low quality factors from 48 graphical images.

Quality Factor	JND	Q_3=75	Q_2=50	Q_1=25
DCY	5.0140	5.3402	4.5152	3.6922
DCU	3.6593	3.9278	3.2195	2.5200
DCV	3.5582	3.8609	3.1625	2.4474
ACY	1.6085	1.6027	1.5752	1.6008
ACU	1.5830	1.6578	1.6578	1.5729
ACV	1.5744	1.6606	1.6568	1.6178

A sample visual output image coming out of JPEG compression on the typical 8 by 8 image blocks and default quantization tables is shown in Figure-4. The output image coming out of the psychovisual threshold on rectangular 6 by 8 image blocks using JND curves is shown in Figure-5. The image in Figure-4 carries 8 by 8 blocking artifacts surrounding Lena's eyelid along the lower eyelashes. Whereas the image in Figure-5 carries a more similar but natural 6 by 8 blocking artifacts. Even though it carries slightly higher quantitative errors as depicted in Table-7, the output image is visually much better. A specific attention can be highlighted on the lower eyelash below the eye.

Figure-4. A Lena right eye coming out of 8 by 8 JPEG quantization tables is zoomed in 400%.

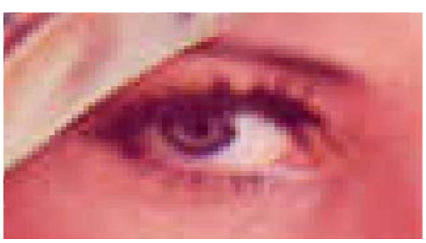

Figure-5. A Lena right eye coming out of 6 by 8 JND curves is zoomed in 400%.

DISCUSSIONS

The principle of psychoacoustic model on audibility threshold is adopted here to measure the psychovisual threshold of the output images. Psychophysics is an approach to measure limited sensitivity and sensation to the physical characteristics of a stimulus. The psychovisual threshold curves in this research have been designed to replace the role of quantization tables.

The design process on the psychovisual threshold curves in this paper is slightly biased toward image quality. A software developer may easily set the parameters A_{L_1} and A_{C_1} higher to gain higher compression rate. At the same time, the psychovisual threshold provides a significant impact on the visual quality of the compressed image.

This research has embarked on the psychovisual threshold of DCT on small and large rectangular image blocks. The psychovisual threshold is practically the best measure on an optimal amount of frequency image signals to be encoded. This quantitative experiment shows the critical role of psychovisual threshold as a basic primitive to generate quantization table in an image lossy compression.

CONCLUSIONS

There is a need for an image and video compression to move up on a higher performance along the advancement of computing powers and multimedia networking technologies. A small block transform is a popular due to of its compact property and ease of implementation. This paper proposes a psychovisual design on quantization process based on the human visual system model. The smoother rectangular quantization tables from the psychovisual threshold produce a better visual quality on the standard resolution output image.

Utilizing a popular encoding technique from audio processing, the output images from the proposed psychovisual error threshold consume a relatively low average bit length of Huffman code. At the same time, they produce consistently better visual quality images than the standard JPEG quantization tables using uniform scaling quality factor. This psychovisual threshold has been designed to support an industrial video compression in the near future. A set of psychovisual curves has been proposed on a small rectangular block for an independent video frame. A further investigation should be made on the front object in video frames.

ACKNOWLEDGEMENTS

The authors would like to thanks the Ministry of Education, Malaysia, for providing financial support under the Fundamental Research Grant Scheme (FRGS/2013/FTMK/ICT07/02/2/F00162).

REFERENCES

[1] N.A. Abu, S.L. Wong, H. Rahmalan, and S. Sahib. 2010. Fast and Efficient 4x4 Tchebichef Moment Image Compression, Majlesi Journal of Electrical Engineering, 4(3): 037-045.

[2] A. Puri, X. Chen and A. Luthra. 2004. Video coding using the H.264/MPEG-4 AVC Compression

Standard, Signal Processing: Image Communication, 19(9): 793–849.

[3] N.A. Abu, F. Ernawan, and S. Sahib. 2013. A Psychovisual Model based on Discrete Orthonormal Transform, International Conference on Mathematical Sciences and Statistics, Kuala Lumpur, Malaysia, AIP Conference Proceedings: 309-314.

[4] N.A. Abu, F. Ernawan, and N. Suryana. 2013. A Generic Psychovisual Error Threshold for the Quantization Table Generation on JPEG Image Compression, 9th IEEE International Colloquium on Signal Processing and its Applications, Kuala Lumpur, Malaysia: 039-043.

[5] S. Srinivasan, P.J. Hsu, T. Holcomb, K. Mukerjee, S.L. Regunathan, B. Lin, J. Liang, M.C. Lee and J. Ribas-Corbera. 2004. Windows Media Video 9: Overview and Applications, Signal Processing: Image Communication, 19(9): 851–875.

[6] J. Singh, S. Singh, D. Singh and M. Uddin. 2011. A Signal Adaptive Filter for Blocking Effect Reduction of JPEG Compressed Images, International Journal of Electronics and Communications, 65(10): 827–839.

[7] T. Painter and A. Spanias. 2000. Perceptual Coding of Digital Audio, Proceedings of the IEEE, 88(4): 451–515.

[8] N.A. Abu, S. L. Wong, N. Suryana, and R. Mukundan. 2010. An Efficient Compact Tchebichef Moment for Image Compression, 10th International Conference on Information Science, Signal Processing and their Applications, Kuala Lumpur: 448-451.

[9] L.E. Tresa and M. Sundararajan. 2014. Video Compression Using Hybrid DWT-DCT Algorithm, International Journal of Emerging Technology and Advanced Engineering, 4(7): 918-922.

[10] N.A. Abu, and M.R.K. Ariffin. 2014. A Novel Psychovisual Model on An Independent Video Frame for Almost Lossless Compression, 10th IEEE International Conference on Information Assurance and Security, Okinawa, Japan: 060-065.