Block Based Temporal Masking Estimation for Video Sequences

Kuo-Cheng Liu
Information Educating Center
Taiwan Hospitality & Tourism College
Taiwan

ABSTRACT: *A method which reveals the visual masking in the human visual system (HVS) is useful in perceptual data coding. In this paper, we propose a color DCT-based method to estimate color spatio-temporal just noticeable distortion (JND) for video coding. The spatio-temporal JND profiles are assessed by incorporating a new temporal masking adjustment into the mathematical model of estimating the DCT-based spatial JND profiles for luminance component and chrominance components of color images. In this paper, the new block-based temporal masking adjustment mainly considering the variation of local temporal statistics in luminance component between successive video frames is proposed. The spatio-temporal JND profiles are used to tune the H.264 video codec for higher performance. The simulation results demonstrate the performance of the perceptual video coding in terms of bit rates and visual quality. The bit rates required by the perceptually tuned video codec are lower than that required by the un-tuned video codec at nearly the same visual quality.*

Keywords: Just-Noticeable Distortion (JND), Luminance, Chrominance, Temporal Masking

Received: 12 November 2014, Revised 18 December 2014, Accepted 24 December 2014

1. Introduction

In order to achieve higher performance of compressing the image, the estimation of approaching actual *JND* of images is always an important topic and utilized into the perceptual coding schemes while considering that human eyes are the ultimate receiver of the visual data. Watson *et al*. [1], [2] presented that quantization matrices for the use in DCT-based compression were designed by exploiting visibility thresholds that are experimentally measured for quantization errors of the DCT coefficients. In [3], the *JND* threshold was determined by the dominant between the luminance masking and the texture masking to adapt the step size of a uniform quantizer in the proposed subband image coder. In [4], the masking thresholds derived in a locally adaptive fashion based on subband decomposition were applied to the design of a locally adaptive perceptual quantization scheme for achieving high performance in terms of quality and bit rate. Yang *et al*. [5] proposed a nonlinear additive model to estimate the spatial *JND* profiles for perceptual coding of color images. In [6], Liu proposed a wavelet-based color visual model to increase the efficiency of image coders in compressing color images.

For investigating the compact representation of video data in field of video coding, the *JND* model incorporated with temporal properties of the HVS are applied to the video codec with high coding efficiency to maintain acceptable visual quality at low bit rates. In [7], the spatial *JND* model in the pixel domain was extended by exploiting the temporal masking effect to obtain the spatio-temporal *JND* for video coding. The *JND* models proposed in [8] are the improvement of [7] for higher performance video coding by introducing the overlapping effect between the luminance adaptation and the spatial contrast masking. In [10], the estimation of the subband just noticeable distortion for video was developed by combining the luminance adaptation, contrast masking, spatio-temporal CSF, and eye movements. In order to improve the performance demonstrated in [10], Wei and Ngan [11] design a new perceptual model by considering the gamma correction and temporal CSF. In [12], the proposed *JND* model is constructed by taking not only the spatial and luminance properties of the HVS but also the temporal and chroma properties into account.

Accurate and effective estimation of the *JND* profiles of video signals is helpful to increase the efficiency of coding the video data. In this paper, a color DCT-based spatio-temporal *JND* model is proposed with the integrated formulations for video signals. By utilizing on the base detection threshold for each DCT coefficient in luminance and chro-minance components of color images, the proposed model uses the masking adjustment presented in our prior work [14] to compute the spatial *JND* profiles of three color components of the color image. Then, a new block-based temporal masking adjustment mainly considering the variation of local temporal statistics in luminance component between successive video frames is proposed to extend the above spatial *JND* profiles for assessing the color spatio-temporal *JND* profiles. The estimated spatio-temporal *JND* profiles are used to tune the H.264 video codec.

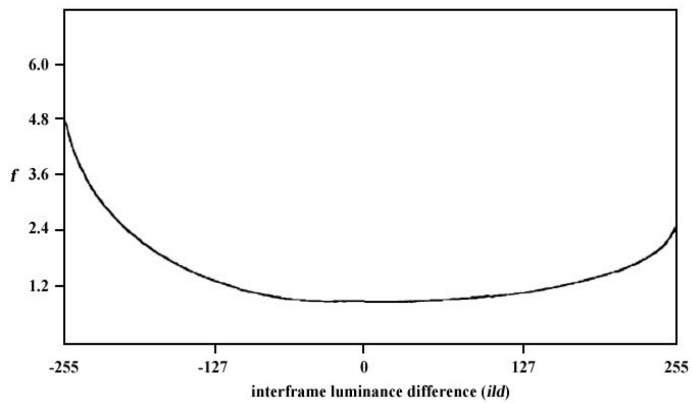

Figure 1. Relation between scale factor α and interframe luminance difference

2. Estimation of Color DCT-based Spatio-temporal JND

Based on the Ahumada's model proposed in [13], the proposed DCT-based spatio-temporal JND for color video is expressed as

$$\boxed{JND_{O,ST}(C^k_{O,b,u,v}) = JND_{O,S}(C^k_{O,b,u,v}).\, a_{Y,t}(k,b)} \quad (1)$$

where $JND_{O,ST}(C^k_{O,b,u,v})$ and $JND_{O,S}(C^k_{O,b,u,v})$ are the spatio-temporal JND value and the spatial JND value, respectively, of the DCT coefficient $C^k_{O,b,u,v}$ at location (u,v) in the b^{th} block of the k^{th} frame in the O ($O = Y$, Cb, and Cr) color component of the video sequences. The spatial JND value of a specified coefficient is evaluated by the corresponding base visibility threshold and the masking adjustment that are described in [14]. $a_{Y,t}(k,b)$ is the proposed temporal masking adjustment that is estimated by using the variation of local temporal statistics of the b^{th} block between the successive frames (the k^{th} frame and the $(k\text{-}1)^{\text{th}}$ frame) in luminance component (Y) of the video sequence.

Temporal masking effect refers to that the temporal redundancy in video sequences is due to motion related blurring and resolution reduction [15], [16]. In [7], the perceptual experiment was investigated to find that human eyes are not sensitive to the changes of luminance on the time axis while video signals are displayed. Based on the variation of local temporal statistics in luminance component between successive video frames, the corresponding temporal masking adjustment used with the spatial *JND* to compute the spatio-temporal *JND* is investigated in this paper. The experiments designed to obtain the adjustment is similar to the experiment presented in [7]. In this paper, the inter-frame variation of local temporal statistics in luminance component is based on the DCT 8×8 block. The relation between the temporal masking adjustment and the inter-frame variation at the b^{th} block between the k^{th} and $(k-1)^{\text{th}}$ frame in luminance component (Y) is given by

$$a_{Y,t}(k,b) = max\left(\alpha_Y(k,b), \beta_Y(k,b)\right) \tag{2}$$

$$\alpha_Y(k,b) = f\left(ild_Y(k,b)\right) \tag{3}$$

$$\beta_Y(k,b) = \begin{cases} 1, & bk(k,b) \textit{ and } bk(k-1,b) \textit{ are all Type } 1 \\ 2.3, & bk(k,b) \textit{ and } bk(k-1,b) \textit{ are Type } 1 \textit{ and } 2, \textit{ respectively} \\ 3.6, & bk(k,b) \textit{ and } bk(k-1,b) \textit{ are all Type } 2 \end{cases} \tag{4}$$

where $ild_Y(k,b) = c^{k}_{Y,b,0,0} - c^{k-1}_{Y,b,0,0}$ means the inter-frame average luminance difference and $bk(k,b)$ a is the b^{th} block in the k^{th} frame of luminance component. The scale factor α_Y is used to describe the human visual sensitivity to the variation of average background luminance between successive sequence frames. f is the function between the scale factor α_Y and the inter-frame average luminance difference. It is closed to the curve shown in Figure 1. The scale factor β_Y is used to describe the human visual sensitivity to the variation of block content between successive sequence frames. To simplify the change of block content, the block is categorized to two types in this paper. *Type* 1 represents plane blocks and edge blocks in which human eyes are usually very sensitive to the distortion in the smooth area or around the edge, while *Type* 2 represents texture blocks where human eyes are not sensitive to the distortion in the texture area.

Also, masking effects exist in chrominance components of the color image and affect the sensitivity to chrominance components of a target color pixel. It cannot be easily identified since masking effects in chrominance components involves complex human vision mechanisms. This makes the estimation of noise detection thresholds in chrominance become difficult. To reduce the complexity of measuring the temporal masking in chrominance components, the temporal masking adjustment used in luminance component is directly applied to chrominance components in this paper while considering that human visual perception is less sensitive to chrominance components than luminance component.

3. Justification of Estimating Color DCT -Based Spatio-temporal JND

To justify the color spatio-temporal *JND* estimation, a subjective test is conducted to inspect if the estimated JND is perceptually redundant to the human visual perception.

Suppose a test image represented in the YCbCr color space is contaminated by the associated JND profiles in the DCT domain. That is

$$\boxed{\widetilde{C}^{k}_{O,b,u,v} = C^{k}_{O,b,u,v} + \delta^{k}_{O,b,u,v} \cdot \varphi \cdot JND_{O,ST}(C^{k}_{O,b,u,v})} \tag{5}$$

where $\widetilde{C}^{k}_{O,b,u,v}$ is the *JND*-contaminated DCT coefficient at location (u,v) in the b^{th} block of the k^{th} frame in the O ($O = Y$, Cb, and Cr) color component of the video sequences, $\delta^{k}_{O,b,u,v}$ is a uniformly distributed random variable taking value of either 1 or –1, and φ is a scale factor whose value can be chosen such that the distortion is uniformly distributed over the contaminated image while the contaminating strength is slightly larger than *JND*. Herein, we use the scale factor φ to inspect how the estimated *JND* approaches the actual *JND* of the video sequences in the following experiment. If the proposed model can accurately estimate the perceptual redundancy, the PSNR of the video contaminated by the associated *JND* profiles in the DCT domain should be as low as possible while maintaining the visual quality.

4. Simulation Results

The proposed method is implemented with six CIF (352×288) color video sequences. In the experiment, the simulation is carried out to justify the proposed *JND* model with $\varphi = 1.0$. In order to evaluate the visual quality of the *JND*-contaminated color video sequences, we adopt the similar subjective viewing test and viewing condition based on the method presented in [12], [17], [18] in the simulation.

In each subjective viewing test, each subject is asked to observe the pair color videos that are displayed side by side on the screen of the monitor. The pair videos are composed of the original color video and its *JND*-contaminated video for evaluating

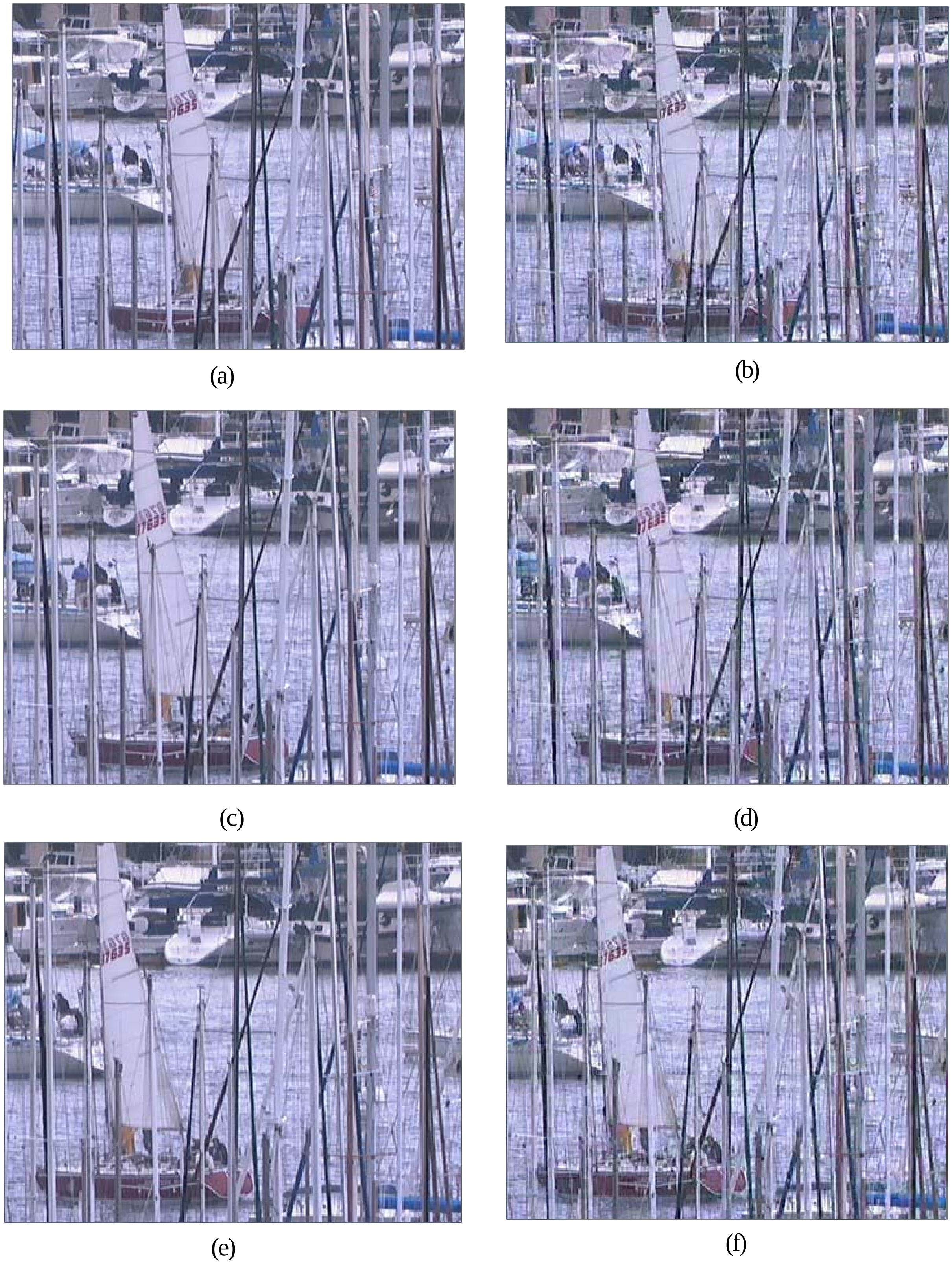

Figure 2. The experimental results of "*Harbour*" color video. (a), (c), (e) the 2^{th}, 12^{th}, and 28^{th} frames in the video sequences and (b), (d), (f) the *JND*-contaminated version of (a), (c), (e), respectively

the perceptual difference between the pair videos. For a fair comparison, the presentation of the video pairs is in randomized order. The viewing condition of observing the pair video sequences is in a dark room at a viewing distance of 6 times the video height.

In Figure 2 shows the experimental results of testing the "*Harbour*" color video at visually lossless quality. The 2^{th} 12^{th}, and 28^{th} frames in the video sequences are depicted for comparison. The original 2^{th} frame of the "*Harbour*" color video is shown in Figure 2a while its *JND*-contaminated version with PSNR of 28.87dB is shown in Figure 2b. The two frames are perceptually indistinguishable for human eyes as observing at the viewing condition described above. The same observation can also be found in other frames. Figure 2d demonstrates that the perceptually lossless visual quality of the *JND*-contaminated 12^{th} frame of the "*Harbour*" color video (28.92dB) is shown while comparing to the original 12^{th} frame (Figure. 2c). These experimental results indicate that the proposed model is justified to be able to estimate the JND profiles for color video sequences. The simulation results of the 28^{th} frame are shown in Figures 2e and f.

In this paper, the estimated color DCT-based spatio-temporal *JND* profiles are used to tune the prediction error signal of the H.264 codec for improving the coding performance. The dynamic range of the prediction error signals we can reduce, the less objective distortion of the reconstructed color image for a given bit rate we can achieve [19]. That is, we utilizes the *JND* profiles to process the prediction error signals such that the dynamic range of perceptually tuned signals can be reduced to achieve lower bit rate or better reconstructed image quality.

$$\hat{e}^{k}_{O,b,u,v} = \begin{cases} e^{k}_{O,b,u,v} + \varepsilon \,.\, JND_{O,ST}(C^{k}_{O,b,u,v}), \; if\; e^{k}_{O,b,u,v} < -\varepsilon \,.\, JND_{O,ST}(C^{k}_{O,b,u,v}) \\ e^{k}_{O,b,u,v} - \varepsilon \,.\, JND_{O,ST}(C^{k}_{O,b,u,v}), \; else\; if\; e^{k}_{O,b,u,v} > \varepsilon \,.\, JND_{O,ST}(C^{k}_{O,b,u,v}) \\ C^{k}_{O,b,u,v}, \; otherwise \end{cases} \tag{6}$$

where *e* is the prediction error signals. The experiments of coding the color video sequences by H.264 codec are carried out with and without tuning the prediction error signals, while all frames are intra coded. The experimental results demonstrate that the bit rates required by the H.264 codec with the perceptually tuned prediction error signals are low-er than that without the perceptually tuned prediction error signals, while both the coding video sequences are at nearly lossless visual quality.

5. Conclusions

In this paper, a color DCT-based spatio-temporal *JND* profile is proposed. It is based on the combination of a new temporal masking adjustment and the mathematical model of estimating the DCT-based spatial *JND* profiles for luminance component and chrominance components of color images. The model is inspected to compare the visual distortion between the video and its *JND*-contaminated version by using a subjective viewing test. By utilizing the spatio-temporal *JND* profile, the coding performance of the perceptually tuned H.264 codec is improved.

Acknowledgement

The work was supported by the National Science Council, Taiwan (R.O.C.), under contract NSC101-2221-E-278-001, and Image & Video Processing Laboratory of Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan.

References

[1] Peterson, H. A., Ahumada, A. J., Jr., A. B. Watson. (1993) Improved detection model for DCT coefficient quantization, in Proc. SPIE Int. Conf. Human Version, Visual Processing, and Digital Display-IV, 1913, Feb, 191-201.

[2] Watson, A. B. (1993). DCT quantization matrices visually optimized for individual images, *In:* Proc. SPIE Int. Conf. Human Version, Visual Processing, and Digital Display-IV, 1913, Feb, 202-216.

[3] Chou, C. H., Li, Y. C. A perceptually tuned sub band image coder based on the measure of just-noticeable-distortion profile, *IEEE Trans. Circuits and Syst. Video Technol*, 5 (6) 467-476, Apr. 1995.

[4] Höntsch, I., Karam, L. J. (2000). Locally adaptive perceptual image coding, *IEEE Trans. Image Processing*, 9 (9) 1472-1483, September.

[5] Yang, X. K., Lin, W. S., Lu, Z., Ong, E. P., Yao, S. (2003). Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images, In: Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, 3, Apr. 2003, 609-612.

[6] Liu, K. C. (2012). Prediction error preprocessing for perceptual color image compression, *EURASIP Journal on Image and Video Processing*, 1 (3) 1-14, Mar.

[7] Chou, C. -H., Chen, C.-W. (1996). A perceptually optimized 3-D sub band codec for video communication over wireless channels, *IEEE Trans. Circuits Syst. Video Technol*., 6 (2) 143–156, Apr.

[8] Yang, X. K., Lin, W. S., Lu, Z. K., Ong, E. P., Yao, S. S. (2005). Just noticeable distortion model and its applications in video coding, Signal Processing: Image Communication, 20 (7) 662-680.

[9] Yang, X., Lin, W., Lu, Z. Ong, E., Yao, S. (2005). Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile, IEEE Trans. *Circuits Syst. Video Technol*., 15 (6) 742–752, Jun.

[10] Jia, Y., Lin, W., Kassim, A. A. (2006). Estimating just noticeable distortion for video, *IEEE Transactions on Circuits and Systems for Video Technology*, 6 (7) 820–829, July.

[11] Wei, Z., Ngan, K. N. (2009). Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain, IEEE Transactions on *Circuits and Systems for Video Technology*, 19 (3) 337-346, 2009.

[12] Chen, H., Hu, R., Hu, J., Wang, Z. (2010). Temporal color just noticeable distortion model and its application for video coding, *In*: Proc. IEEE Int. Conf. *Multimedia and Expo*, July, 713-718.

[13] Ahumada, A., Peterson, H. (1992). Luminance-model-based DCT quantization for color image compression, *In*: Proc. SPIE Human Vision, Visual Process., Digit. Display III, 1666, 365–374.

[14] Liu, K.-C. (2013). Perceptual JPEG compliant coding by using DCT-based visibility thresholds of color images, *In*: Proc. Int. Conf. on Digital Image Processing, 119-122.

[15] Kelly, D. H. (1961). Visual responses to time-dependent stimuli, *J. Opt. Soc. Amer*., 51, 422-429.

[16] Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system, *J. Opt. Soc. Amer*., 56, 1141-1142, 1966.

[17] Longere, P., Zhang, X., Delahunt, P. B., Brainaro, D. H. (2002). Perceptual assessment of demosaicing algorithm performance, *In*: Proc. *IEEE*, 90 (1) 123–132, Jan.

[18] Jia, Y. T., Lin, W. S., Kassim, A. A. (2006). Estimating just-noticeable distortion for video, *IEEE Trans. Circuits Syst. Video Technol*., 16 (7) 820-829, July.

[19] Berger, T. (1971). Rate Distortion Theory. Englewood Cliffs, NJ: Prentice-Hall.