Estimation of Structural Irregularity Based Temporal Masking for Color Video

Kuo-Cheng Liu
Information Educating Center, Taiwan Hospitality & Tourism University, Taiwan
Email: kcliu@mail.tht.edu.tw

***Abstract*—In this paper, we present a new temporal masking adjustment for designing DCT-based method to estimate color spatio-temporal Just Noticeable Distortion (JND) of color video sequences. The sensitivity of human perception to spatio-temporal structural irregularity in video sequences is investigated. The variation of structural properties and luminance change between the successive frames are taken into account simultaneously to deduce a new structural irregularity based temporal masking adjustment. Then, the spatio-temporal JND profiles are assessed by incorporating the temporal masking adjustment into the mathematical model of estimating the DCT-based spatial JND profiles for luminance and chrominance components of color images. The method of estimating spatio-temporal JND profiles is verified by a subjective test in which noise contamination is carried out by adding to or subtracting from the video sequence. The simulation results demonstrate that the PSNR of the JND-contaminated video sequence is low enough while maintaining nearly the same visual quality of its original video sequence.**

***Index Terms*—structural irregularity, just-noticeable distortion, temporal masking**

I. Introduction

By exploiting visibility thresholds that are experimentally measured for quantization errors of the DCT coefficients, the quantization matrices for the use in DCT-based compression were designed [1]. 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 [2]. In [3], a nonlinear additive model to estimate the spatial JND profiles for perceptual coding of color images was proposed. Liu [4] proposed a wavelet-based color visual model to increase the efficiency of image coders in compressing color images.

The JND model incorporated with temporal properties of the HVS is applied to the video codec and can be found in [5]-[9]. The JND models proposed in [6] introduce the overlapping effect between the luminance adaptation and the spatial contrast masking for better performance. By combining the luminance adaptation, contrast masking, spatio-temporal CSF, and eye movements, Jia *et al.* [7] estimate the subband just noticeable distortion for video. In [8], a new perceptual model by considering the gamma correction and temporal CSF was designed. In [9], an alternative perceptual video coding method was introduced to improve upon the current H.264/advanced video control framework.

In the latest research efforts [10]-[13], the visual models that analyze the structures of the image in the spatial domain to deduce the visual masking for accurately estimating the visibility thresholds of the image have been successfully proved. If the temporal masking effect takes the inter-frame variation of structural properties into account, the improvement and enhancement of the existing spatio-temporal JND model are reasonable to be expected. In this paper, we extend the structures of images in the spatial domain to that of videos in the temporal domain and analyze the relation between the inter-frame variation of structural properties and the temporal masking. By utilizing the relation, a scale factor is designed for temporal masking to extend the new masking function to visibility threshold estimation of video signals. To avoid underestimating or overestimating the masking effect in the region with alteration of the structural characteristics, the inter-frame variation of structural properties and luminance change between the successive frames are taken into account simultaneously to deduce the new masking function based on spatio-temporal structural irregularity of video sequences in this paper. By using this masking function, the model of estimating the visibility threshold of difference in video sequences is presented.

II. Structural Irregularity Based Temporal Masking

The temporal redundancy in video sequences is due to motion related blurring and resolution reduction, or so called the temporal masking effect. In [5], 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. Wang *et al.* [14] explored the spatio-temporal structural features that are implicitly existed in a local gradient, when the video signals are regarded as a pixel volume to show the spatio-temporal structural

Manuscript received April 8, 2015; revised August 31, 2015.

doi: 10.18178/joig.3.2.149-152

characteristics in the local time region. To approach the compound effect of multiple maskings and avoid underestimating or overestimating the temporal masking effect in the region with alteration of structural characteristics, two scale factors f_1 and f_2, contributed by luminance change and spatio-temporal structural irregularity, respectively, are simultaneously taken into account to deduce the new temporal masking function that takes the addition form of the two factors. 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 m-th block between the t-th and (t-1)-th frame in the O (O=Y, Cb, and Cr) color component is given by:

$$f_{O,t}(t,m) = \max\{f_{O,1}(t,m), f_{O,2}(t,m)\} \quad (1)$$

where $\max\{\cdot\}$ means that the temporal adjustment is mainly controlled by the stronger masking caused by luminance change and spatio-temporal structural irregularity, respectively. Since the visual tolerance is measured by using the visual masking effects, this indicates that more error tolerance is allowed while the stronger masking is adopted. The factor contributed by the luminance change is given by:

$$f_{O,1}(t,m) = L(c^{t}_{O,m,0,0} - c^{t-1}_{O,m,0,0}) \quad (2)$$

where $c^{t}_{O,m,i,j}$ is the DCT coefficient at location (*i*, *j*) in the *m*-th block of the *t*-th frame in the O (O=Y, Cb, and Cr) color component of the video sequences. This means that the factor is varied with the average difference of each block between successive frames. The factor contributed by the spatio-temporal structural irregularity is given by:

$$f_{O,2}(t,m) = \Psi(si_O(x,y,t))$$

$$si_O(x,y,t) = \sum_{O,0<i\le m,i\in\Re} (LBP(x,y,t), I_i, g_x, g_y, g_t) \quad (3)$$

where $si_O(x, y, t)$ is the spatio-temporal structural irregularity of the pixel (*x*, *y*) in the *t*-th frame, *LBP*(*x*, *y*, *t*) is the local binary pattern (LBP) that characterizes the spatio-temporal structure for the pixel and is calculated by the local neighborhood of the pixel (*x*, *y*, *t*), $\Re$ is the local neighborhood of the pixel (*x*, *y*, *t*), *m* is its size, and g_x, g_y and g_t are the local gradients of the pixel (*x*, *y*, *t*) in current frame along *x*, *y*, and *t* direction of two successive frames by applying the 3D Sobel kernels (Fig. 1) to calculate for describing the spatio-temporal structural characteristics.

Also, masking effects exist in chrominance components of the color image and affect the sensitivity to chrominance components of a target color pixel in the visual information. 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.

Herein, the proposed DCT-based spatio-temporal JND for color video is simply computed as the product of the spatial JND value and the corresponding proposed temporal masking adjustment based on the Ahumada's model proposed in [15]. In this paper, we use the model proposed in [16] to calculate the spatial JND value of a specified coefficient, where the corresponding base visibility threshold and the masking adjustment are combined for measurement.

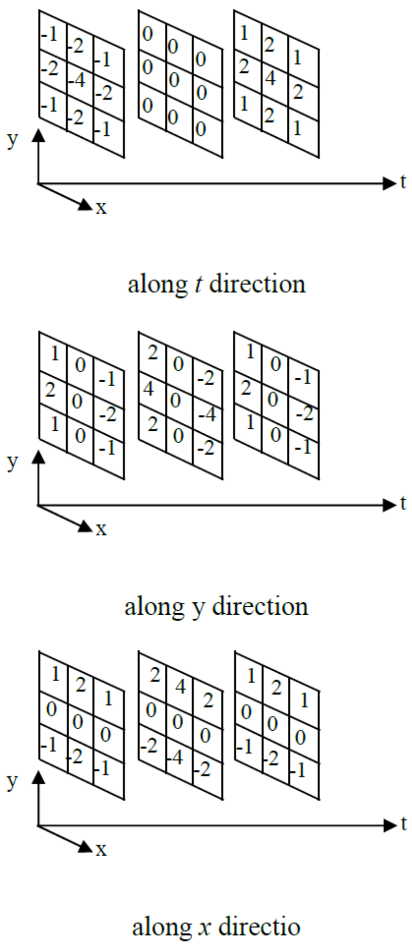

Figure 1. 3D Sobel kernels along x, y, and t direction.

III. Justification of 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:

$$\tilde{c}^{t}_{O,m,i,j} = c^{t}_{O,m,i,j} + \delta^{t}_{O,m,i,j} \cdot \varphi \cdot D_{O,ST}(c^{t}_{O,m,i,j}) \quad (4)$$

where $c^t_{\text{O},m,i,j}$ and $\tilde{c}^t_{\text{O},m,i,j}$ are the DCT coefficient and its JND-contaminated DCT coefficient at location (i, j) in the m-th block of the t-th frame in the O (O=Y, Cb, and Cr) color component of the video sequences, $D_{\text{O},ST}(c^t_{\text{O},m,i,j})$ is the spatio-temporal JND value, $\delta^t_{\text{O},m,i,j}$ 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.

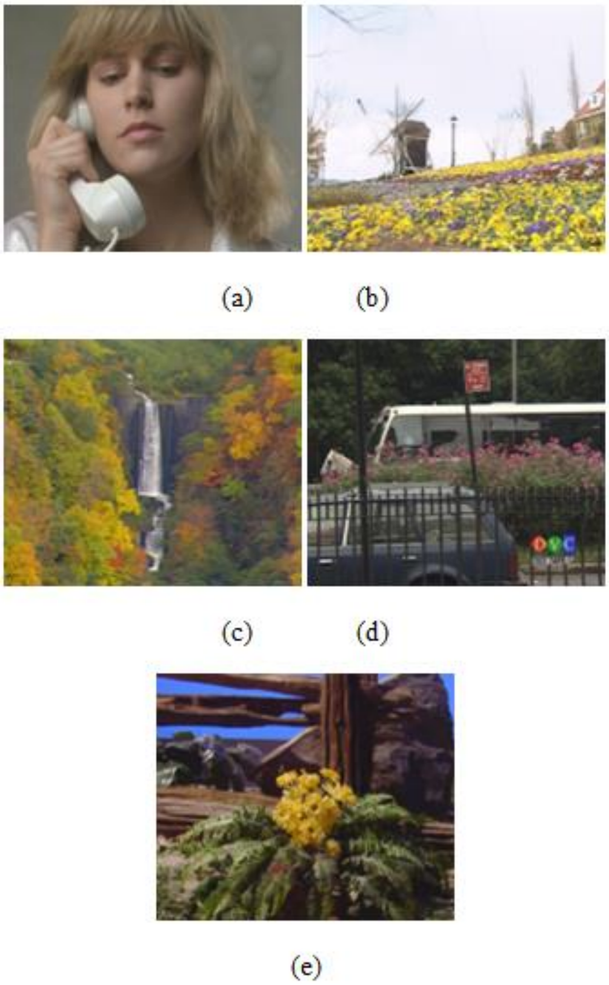

Figure 2. Thumbnail of the test sequences: (a) Suzie, (b) flower, (c) waterfall, (d) bus, and (e) tempete.

IV. Simulation Results

In the simulation, the similar subjective viewing test and viewing condition based on the method presented in [17], [18] is used in order to evaluate the visual quality of the JND-contaminated color video sequences. Five color video sequences (Fig. 2) are used in the test for verifying the proposed method. In the viewing test, the pair color videos of composing of the original color video and its JND-contaminated video are displayed side by side on the screen of the monitor for comparison. The subjects are asked to evaluate the perceptual difference between the pair videos. 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 order to achieve a fair comparison, we chose the representative model proposed in [19] for color images and videos to implement in the simulation results. Watson designed perceptual color quantization matrices for DCT-based image compression. By using the Watson's model and the proposed method, the comparison of average PSNR values (dB) for JND-contaminated sequences is shown in Table I. The average PSNR values given by the proposed method are lower than that given by the Watson's model while maintaining nearly the same visual quality of the original video sequence. It is obviously shown that the proposed method incorporated with the structural irregularity based temporal masking adjustment is able to measure the more accurate visual error tolerance and outperform the existing model.

TABLE I. Comparison of Average PSNR Values for JND-Contaminated Sequences

	PSNR (dB) values	
Sequences	Watson's model	Proposed method
Suzie	32.56 dB	31.63 dB
Flower	32.18 dB	31.41 dB
Waterfall	31.43 dB	30.73 dB
Bus	30.74 dB	30.05 dB
Tempete	31.77 dB	30.89 dB

V. Conclusions

In this paper, the sensitivity of human perception to spatio-temporal structural irregularity in video sequences is investigated for temporal masking. A new structural irregularity based temporal masking adjustment is successfully designed for the better estimation of color spatio-temporal JND of color video sequences.

Acknowledgement

The work was supported by the Ministry of Science and Technology, Taiwan (R.O.C.), under contract MOST 103-2221-E-278-001, and Image & Video Processing Laboratory of Department of Computer Science and Information Engineering, National Dong-Hwa University, Taiwan.

References

[1] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. SPIE Int. Conf. Human Version, Visual Processing, and Digital Display-IV*, Feb. 1993, vol. 1913, pp. 202-216.

[2] I. Höntsch and L. J. Karam, "Locally adaptive perceptual image coding," *IEEE Trans. Image Processing*, vol. 9, no. 9, pp. 1472-1483, Sep. 2000.

[3] X. K. Yang, W. S. Lin, Z. Lu, E. P. Ong, and S. Yao, "Just-Noticeable-Distortion profile with nonlinear additivity model for perceptual masking in color images," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Apr. 2003, vol. 3, pp. 609-612.

[4] K. C. Liu, "Prediction error preprocessing for perceptual color image compression," *EURASIP Journal on Image and Video Processing*, vol. 1, no. 3, pp. 1-14, Mar. 2012.

[5] C. H. Chou and C. W. Chen, "A perceptually optimized 3-D subband codec for video communication over wireless channels," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 143-156, Apr. 1996.

[6] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-Compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742-752, Jun. 2005.

[7] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just noticeable distortion for video," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 7, pp. 820-829, Jul. 2006.

[8] Z. Wei and K. N. Ngan, "Spatio-Temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 3, pp. 337-346, 2009.

[9] Z. Luo, L. Song, S. Zheng, and N. Ling, "H.264/Advanced video control perceptual optimization coding based on JND-directed coefficient suppression," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 23, no. 6, Jun. 2013.

[10] A. Liu, W. Lin, M. Paul, C. Deng, and F. Zhang, "Just noticeable difference for images with decomposition model for separating edge and textured regions," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, no. 11, pp. 1648-1652, Nov. 2010.

[11] J. Wu, F. Qi, and G. Shi, "Self-Similarity based structural regularity for just noticeable difference estimation," *Journal of Visual Communication and Image Representation*, vol. 23, no. 6, pp. 845-852, 2012.

[12] J. Wu, W. Lin, and G. Shi, "Visual masking estimation based on structural uncertainty," in *Proc. IEEE Symp. Circuits and Systems (ISCAS)*, May 2013, pp. 933-936.

[13] J. Wu, W. Lin, G. Shi, X. Wang, and F. Li, "Pattern masking estimation in image with structural uncertainty," *IEEE Trans. Image Processing*, vol. 22, no. 12, pp. 4892-4904, Dec. 2013.

[14] Y. Wang, T. Jiang, S. Ma, and W. Gao, "Novel spatio-temporal structural information based video quality metric," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 7, pp. 989-998, Feb. 2012.

[15] A. Ahumada and H. Peterson, "Luminance-Model-Based DCT quantization for color image compression," in *Proc. SPIE Human Vision, Visual Process., Digit. Display III*, 1992, vol. 1666, pp. 365-374.

[16] K. C. Liu, "Perceptual JPEG compliant coding by using DCT-based visibility thresholds of color images," in *Proc. Int. Conf. on Digital Image Processing*, 2013, pp. 119-122.

[17] P. Longere, X. Zhang, P. B. Delahunt, and D. H. Brainaro, "Perceptual assessment of demosaicing algorithm performance," *Proc. IEEE*, vol. 90, no. 1, pp. 123-132, Jan. 2002.

[18] Y. T. Jia, W. S. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 820-829, Jul. 2006.

[19] A. B. Watson, "Perceptual optimization of DCT color quantization matrices," in *Proc. IEEE Int. Conf. Image Processing*, Singapore, 1994, pp. 100-104.

Kuo-Cheng Liu received the Ph.D. degrees from the Department of Electrical Engineering, Tatung University, Taiwan. He is President of Taiwan Hospitality and Tourism University, Taiwan and Professor with the Information Educating Center and Graduate Institute of Tourism & Hospitality Management. In 1997, he joined the Department of Information Management, Taiwan Hospitality and Tourism University, Taiwan, as a Lecturer. He became an Assistant Professor in the Department of Information Management in 2005, an Associate Professor in 2009, and is currently a Professor with the Information Educating Center. Professor Liu has participated into several Research Projects funded by the Ministry of Science and Technology (MOST) and Ministry of Education (MOE), Taiwan. Since 2009, he was also involved in the research activities and projects of Image & Video Processing Laboratory (IVP Lab.) guided by Professor Shinfeng D. Lin, Dean of College of Science and Engineering, in National Dong Hwa University, Taiwan. His research activities and interests are in the areas of perceptual coding of color images, multimedia signal processing, information application to tourism education and training, and information management. He was invited as referee for highly ranked scientific journals including IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Information Forensics & Security, IET Image Processing, Neurocomputing, Applied Computing and Informatics, and Computer Journal *et al*. and the committee member of various international conferences.