MODELLING VISUAL ATTENTION AND MOTION EFFECT FOR VISUAL QUALITY EVALUATION

Zhongkang Lu, X. K. Yang, W. S. Lin, E. P. Ong and Susu Yao

Institute for Infocomm Research, Agency for Science, Technology and Research,
21 Heng Mui Keng Terrace, Singapore 119613

ABSTRACT

The perceptual visual quality evaluation of Human Visual System (HVS) is very complex. It concerns almost all aspects of visual processing in vision path, from low-level neuron activities to high-level visual perception. Existing perceptual Visual Quality Metrics (VQMs) only considered several of the mechanisms of HVS, and many others are ignored. In this paper, two global modulatory factors, visual attention and motion suppression, are modelled and combined to form a mathematic expression - Perceptual Quality Significant Level (PQSL). To a certain extent, it is believed that PQSL value reflect the processing ability of human brain on local visual contents. To evaluate their effects on visual quality evaluation, two VQMs are proposed. One is a MSE-like VQM based on PQSL-modulated JND profile, which was proposed in [1]; the other VQM is based on Wang's visual quality assessment [2], PQSL values are used to adjust the weights of his structural similarity index. Experimental results show that introducing of the global modulatory factors can improve the performance of current visual quality metrics.

1. INTRODUCTION

Various visual distortions are introduced in images/video by different visual processing techniques. Some distortions cannot be perceived by the HVS, and obviously do not have any contribution on visual quality. Visual Quality Metrics (VQMs) should be able to reflect this fact. Visible distortions can be separated into two categories: near-threshold distortions and suprathreshold distortions. Near-threshold distortions are the distortions around the threshold of visual detection, and suprathreshold distortions are much stronger to the threshold. Generally, suprathreshold distortions introduced by video/image compression techniques appear in certain forms, such as blockiness, blurrness and ringing. These distortions may destroy the structure of objects and scene in image/video, and evaluating the annoyance by suprathreshold distortions is difficult since the processing involves high level activities of human brains, such as pattern matching, object recognition and perception.

The thresholds that determines the visibility of distortions is also defined as Just-Noticeable-Distortion (JND), which are motivated from psychophysical and physiological vision research. Because the thresholds are different from person to person, JND is technically defined as the amplitude of noise that has a 50% possibility of detection by the HVS.

The evaluation of perceptual quality scores on near-threshold distortions are usually given by measuring the detectability of distortions, while on suprathreshold conditions, the score is given by measuring the annoyance of distortions. Correspondingly, current perceptual VQMs also can be separated into two categories. The first kind of VQMs share a error sensitivity-based philosophy [3, 4, 5]. This kind of techniques rely on the accurate estimation of detection thresholds, basically assuming that perceived quality depends significantly on the error threshold. By the benefits of exclusive psychophysical explorations on measuring the visibility thresholds of the HVS, this kinds of techniques work well on image/videos with near-threshold distortions. The second kind of VQMs only measure the structural distortions. In [6, 7, 2], prior knowledge of known coding artifacts is used in VQMs specific for decoded images/video. Blockiness is the most oft-used coding artifact (for JPEG, H.26x and MPEG-x applications), while others (e.g., ringing, blurring, damaged edge, correlated error) are additional candidates for the purpose.

All these two kinds of VQMs adapts several local characteristics of visual contents as the input, but there are some global modulatory factors are not considered. One of the factors that previous perceptual VQMs do not included is the HVS's visual attention [8, 9] on contents in visual field. Visual attention is a very important mechanism of HVS, and it is believed to be the result of several millions of years' evolution [10]. Visual attention can be defined as a set of strategies that attempts to reduce the computational cost of the search processes inherent in visual perception [11], and it also can be regarded as an active mechanism of human brain to re-allocate its computational resources on visual field. Comparatively, more computational resources of HVS are to be re-allocated to high attentional areas than low at-

0-7803-8554-3/04/$20.00

tentional areas. The formation of visual attention is very complex. It concerns all aspect of visual processing in human brain, and its aftereffects also influence the perception of visual contents in all levels. One of visual attention's aftereffect is that it can enhance or reduce the actual visual sensitivity inside and outside of fovea area [8, 9]. It is believed that the re-allocation of computational resources is the reason of the modulation of visual sensitivity in different areas.

Motion suppression is another mechanism that reduce contrast sensitivity of surrounding areas by motion of fixative objects. Motion suppression is caused by the motion on retina image [12, 13], which is a combinatory effect of both object motion and eye movement. Computationally, it is believed that motion on retina image increases the processing cost of visual perception, thus suppresses the visual sensitivity. The motion suppression always happens on the low attentional areas when their motion are different to high attentional areas.

Visual attention and motion suppression are both global modulatory factors that modulate visual perception on visual field.

Because the eye movement always follow the shift of visual attention, and motion suppression is caused by relative motion to eye movement, motion suppression and visual attention are highly correlated. So, an computational expression, Perceptual Quality Significant Level (PQSL), which combines visual attention's aftereffect and motion suppression numerically, is first proposed in [14]. In the work, PQSL has been estimated by a nonlinear combination of both high level visual features (e.g., face, skin color), and low level stimuli (e.g., motion, luminance, color and texture contrast). To a certain extent, it is believed that PQSL value reflect the processing ability of human brain on local visual contents.

By experimental results on embedding near-threshold noise in video with JND profiles, it is proven that PQSL-modulated JND profile can improve the accuracy on JND prediction [1]. The PQSL-modulated JND profile is an improvement from Yang's JND profile [15]. An example of PQSL-modulated JND profile is shown in Figure 1. Figure 1(a) is the original frame of sequence 'Foreman', Figure 1(b) is the estimated PQSL map, Figure 1(c) is the JND profile by Yang's model [15], and Figure 1(d) is the estimated PQSL-modulated JND profile. Please note that the upper half of Figure 1(c) and (d) is JND profile in Y channel, the left lower part is in Cb channel and the left lower part is in Cr channel. Moreover, they are both normalized for better visualization. Generally, the modulation effect of PQSL is that the detectability threshold is pushed to a lower level when PQSL value is high, and the tolerance is increased when PQSL value is low.

However, can PQSL improve the visual quality evaluation with suprathreshold distortions? Or, in other words, if the same suprathreshold distortion on different position in visual field have different contribution to visual quality evaluation? Based on the assumption that PQSL value reflect the ability of visual perception on local contents, the suprathreshold distortions in high PQSL value areas should be more annoyance to human eyes than those in low PQSL value areas. To prove it, a modified VQM is proposed to introduce PQSL into Wang's quality assessment, which is based on a structural similarity index to measure suprathreshold distortion [2]. Moreover, to further evaluate the performance of the PQSL-modulated JND profile [1], a MSE-like VQM is also proposed.

In this paper, the details of the two VQMs are given in Section 2; the experimental results are presented and analyzed in Section 3; Section 4 is the conclusions.

2. PQSL-BASED VQMS

The PQSL-modulated JND profile, which is proposed in [1], is applied in both VQMs. Let I, $\hat{I}$ and Θ denotes original video sequences, degraded video sequence and the PQSL-modulated JND profile. The VQM A can be expressed as:

$$Q_A = \sum_{(x,y,t)} f_a(|I(x,y,t) - \hat{I}(x,y,t)|, \Theta(x,y,t)) \quad (1)$$

where

$$f_a(a,b) = \begin{cases} \frac{a}{b} - 1 & : \text{ if } a > b \\ 0 & : \text{ if } a \leq b \end{cases} \quad (2)$$

The meaning of equation 1 is that the distortion below detectability threshold (JND) can be ignored, and only the distortion that is stronger to JND contribute to visual quality score. The concept is similar to Chou's PPSNR [16], but the distortions are scaled by JND values. This is based the assumption that contribution of distortion to visual quality is affected by the sensitivity level. Theoretically, this VQM is more suitable to near-threshold distortion conditions.

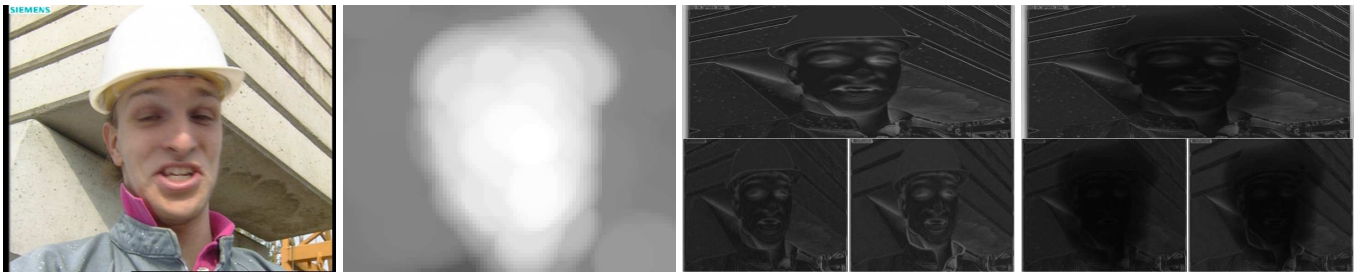
VQM B adapts Wang's structural similarity index (SSIM) [2], which includes luminance, contrast and structure comparison measures on $YCbCr$ color space. The VQM B can be expressed as:

$$Q_B = \sum_{(x,y,t)} f_b(SSIM, \max_{(x,y,t)}(|I - \hat{I}|), \Theta, PQSL) \quad (3)$$

where

$$f_b(a,b,c,d) = \begin{cases} a \cdot (d-\beta)^\alpha & : \text{ if } b > c \\ 0 & : \text{ if } b \leq c \end{cases} \quad (4)$$

where α and β are constant parameters, which is tuned to map PQSL values into a suitable scope. The values of α and β are dependent on the PQSL estimation algorithm [17], and



(a) (b) (c) (d)

Fig. 1. An example of PQSL and PQSL-modulated JND profile estimation: (a) the 30^{th} frame of video 'Foreman'; (b) estimated map of PQSL from (a); (c) JND profile by Yang's model (d) PQSL-modulated JND profile.

reflect the weighting of the PQSL on suprathreshold distortion evaluation. Because SSIM values are block-based, the position (x, y, t), and correspondent Θ ,$PQSL$ are all block-based. Moreover, the operation $\max_{(x,y,t)}(|I - \hat{I}|)$ is to find the maximum difference between I and $\hat{I}$ within the block (x, y, t).

Same as Equation 1, the distortion below JND is ignored in equation 3.

3. EXPERIMENTAL RESULTS

Experiments are based on VQEG [18] Phase-I test dataset, which includes 20 SDTV test sequences, subjective rating results and evaluation methods to evaluate the performance of proposed video quality metrics. Three methods, which are adopted by VQEG [18], are used to evaluate the prediction accuracy of the VQMs. They are:

- Method 1 (M_1): The Pearson linear correlation coefficient between $DMOS_p$ and $DMOS$;
- Method 2 (M_2): Spearman rank order correlation coefficient between $DMOS_p$ and $DMOS$;
- Method 3 (M_3): Outlier ratio of "outlier-points" to total points N.

Where N is the total number of testing sequences, $DMOS$ are Difference Mean Opinion Scores (DMOS), the output of subjective testings. $DMOS_p$ are predicted DMOS. A three-parameter logistic function is used to estimate $DMOS_p$ from output of VQMs (Video Quality Rating, VQR):

$$DMOS_p = \frac{b1}{1 + \exp(-b2 * (VQR - b3))} \qquad (5)$$

where $b1$, $b2$ and $b3$ are fitted parameters from $[DMOS, VQR]$.

The experimental results are list in Table 1. 20 VQEG Phase-I testing sequences are re-categorized into three sets for VQMs' performance evaluation: 1), 50HZ set with 180 testing sequences in PAL SDTV format; 2), 60HZ set with 180 testing sequences in NSTC SDTV format; and 3), both of them. We don't have the experimental results of Wang visual quality assessment on 50Hz and 60Hz data sets, but we still can see from table 1 that the accuracy of VQM B is improved compare with Wang's visual quality assessment. As for VQM A, its performance is also improved compare to PSNR, which is equivalent to MSE.

In the experiment, the accuracy of VQM B is better than VQM A. The reason is that the testing sequences used in VQEG Phase-I work are degraded by DCT-based video compression softwares, and blockiness is the major distortion. This is a suprathreshold situation.

4. CONCLUSION

Two global modulatory factors, visual attention and motion suppression that may influence visual quality evaluation, are explored in this paper. To combine the two factors, a numerical expression, Perceptual Quality Significant Level (PQSL), is also introduced. To a certain extent, it is believed that PQSL value reflect the processing ability of human brain on local visual contents. Based on the analysis, two visual quality metrics are proposed. One is a MSE-like visual quality metric based on PQSL-modulated JND profile; and the other adapts Wang's structural similarity index, with weighting adjustment based on PQSL. Experimental results based on VQEG Phase-I test dataset show that introducing of the two factors into visual quality metrics can improve their performance.

5. REFERENCES

[1] Z. K. Lu, W. Lin, X. K. Yang, E. Ong, and S. Yao, "Spatial selectivity modulated just-noticeable-distortion profile for video," in *The 2004 IEEE International conference on Acoustics, Speech, and Signal Processing (Accepted)*, 2004.

[2] Zhou Wang, Ligang Lu, and Alan C. Bovik, "Video

Table 1. Experimental results of VQM A and B on VQEG Phase-I data set. Here M_1 is the output of Metric 1, M_2 is the output of Metric 2, and M_3 is the output of Metric 3.

	PSNR			VQM A			Wang's VQA			VQM B		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
50Hz	0.786	0.810	0.728	0.820	0.812	0.567	-	-	-	0.889	0.859	0.539
60Hz	0.760	0.711	0.583	0.795	0.762	0.628	-	-	-	0.914	0.897	0.556
All	0.779	0.786	0.678	0.812	0.805	0.603	0.849	0.812	0.578	0.895	0.871	0.541

quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, February 2004.

[3] Scott Daly, "The visible differences predictor: An algorithm for the assessment of image fidelity," in *Digital Images and Human Vision*, Andrew B. Watson, Ed., pp. 179–206. MIT Press, Cambridge, Massachusetts, 1993.

[4] Stephan Winkler, *Vision models and quality metrics for image processing applications*, Ph.D. thesis, Ecole Polytecnique Federale De Lausanne (EPFL), Swiss Federal Institute of Technology, Thesis No. 2313, Lausanne, Switzerland, December 2000.

[5] A. B. Watson, J. Hu, and J. F. McGowan III, "Dvq: A digital video quality metric based on human vision," *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[6] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in image based on human visual sensitivity," *IEEE Transaction Image Processing*, vol. 4, no. 6, pp. 713–724, 1995.

[7] H. R. Wu, "A new distortion measure for video coding blocking artifacts," in *Proceedings of 1996 International Conference on Communication Technology*, May 1996, vol. 2, pp. 658 – 661.

[8] Harold L. Hawkins, Steven A. Hillyard, Steven J. Luckand Mustapha Mouloua, Cathryn J. Downing, and Donald P. Woodward, "Visual attention modulates signal detectability," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 16, no. 4, pp. 802–811, November 1990.

[9] L. Itti, J. Braun, and C. Koch, "Modeling the modulatory effect of attention on human spatial vision," in *Advances in Neural Information Processing Systems, Vol. 14*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds. MIT Press, Cambridge, MA, 2002.

[10] M. M. Chun and J. M. Wolfe, "Visual attention," in *Blackwell Handbook of Perception*, B. Goldstein, Ed., pp. 272–310. Blackwell Publishers Ltd., Oxford, UK, 2001.

[11] J. K. Tsotsos, "Motion understanding: Task-directed attention and representations that like perception with action," *International Journal of Computer Vision*, vol. 45, no. 3, pp. 265–280, December 2001.

[12] Brian J. Murphy, "Pattern thresholds for moving and stationary gratings during smooth eye movement," *Vision Research*, vol. 18, no. 5, pp. 521–530, 1978.

[13] D. H. Kelly, "Visual processing of moving stimuli," *Journal of the Optical Society of America A - Optics Image Science and Vision*, vol. 2, no. 2, pp. 216–225, February 1985.

[14] Z. K. Lu, W. Lin, E. Ong, S. Yao, and X. K. Yang, "Perceptual-quality significance map (pqsm) and its application on video quality distortion metrics," in *Proceedings of ICASSP'2003*, Hong Kong, April 2003, vol. 3, pp. 617–620.

[15] X. K. Yang, W. S. Lin, Z. K. Lu, E. P. Ong, , and S. S. Yao, "Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images," in *Proceedings of ICASSP'2003*, Hong kong, April 2003, vol. 3, pp. 609 – 612.

[16] Chun-Hsien Chou and Yun-Chin Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transaction on Circuits System on Video Technologies*, vol. 5, no. 6, pp. 467–476, May 1995.

[17] Z. K. Lu, W. Lin, X. K. Yang, E. Ong, and S. Yao, "Pqsm based rf and nr video quality metrics," in *SPIE Proceedings of VCIP'2003*, Lugano, Switzerland, July 2003, vol. 5150, pp. 633–640.

[18] VQEG (Video Quality Expert Group), "Final report from the video quality expert group on the validation of objective models of video quality assessment," March 2000, http://www.vqeg.org.