SPATIAL SELECTIVITY MODULATED JUST-NOTICEABLE-DISTORTION PROFILE FOR VIDEO

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

Institute for Infocomm Research, Agency for Science, Technology and Research,
21 Heng Mui Keng Terrace, Singapore, 119613
Email: {zklu,wslin,xkyang,epong,ssyao}@i2r.a-star.edu.sg

ABSTRACT

Both visual sensitivity and spatial selectivity determine the overall visibility threshold at each pixel in an image, according to the physiological and psychological evidence towards the human vision system (HVS). Visual sensitivity can be decided by an existing estimator for Just-Noticeable-Distortion (JND). In this paper, a computational model is proposed for incorporating a selectivity measure into the JND profile so that more effective noise shaping is possible in various applications. Experimental results with noise-embedded video sequences confirm that introduction of spatial selectivity enhances the performance of JND profile used in noise shaping.

1. INTRODUCTION

Noise-shaping refers to re-allocating the inevitable noise or distortion into some domains or areas so that the resultant visual variation is the least noticeable to human eyes, in the applications like visual compression, communication, displaying, and data hiding. Therefore, visual sensitivity is an important issue in the relevant research [1, 2, 3]. Just-Noticeable-Distortion (JND) determines the visibility thresholds in pixels [4, 5] or subbands [6, 7]. Approximately, a JND threshold can be regarded as the inverse of visual sensitivity.

Another factor affecting our perception towards visual signal is the human vision system (HVS)'s spatial selectivity (or visual attention) [8, 9, 10] on contents in visual field. Spatial selectivity can enhance or reduce the actual visual sensitivity and consequently JND profiles need to be adjusted inside and outside of fovea area.

In this paper, we will exploit the methodology to combine visual sensitivity and spatial selectivity for better noise shaping related applications. In Section [**?**], an analysis of the relationship between spatial selectivity and visual sensitivity is given, based upon the HVS' mechanisms. A computational model is then discussed to modulate the JND profile with spatial selectivity measures, in Section 3. Experimental results are given in Section 4 for noise shaping, and the conclusions are given in Section 5.

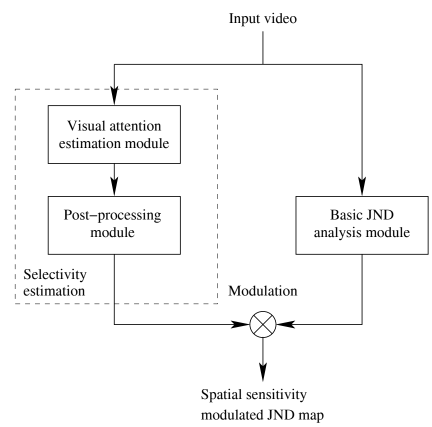

Fig. 1. Flowchart of the spatial selectivity modulated JND profile model.

2. VISUAL SENSITIVITY AND SPATIAL SELECTIVITY

Both visual sensitivity and spatial selectivity result from the biological mechanisms of the HVS. The affecting factors include: 1) distance between eyes and display, 2) optical property of eyes, 3) the structure of neurons lay behind photoreceptor cells (ganglion cells), 4) the noise introduced in vision path, 5) optic-electric properties of all kinds of photoreceptor cells on retina, and 6) fixation position and density distribution of photoreceptor cells on retina [11]. Visual sensitivity is affected mainly by 1) $\sim$ 5) [12, 13], while spatial selectivity is determined by 6).

The HVS perception has the highest spatial resolution and sensitivity at the point of fixation (fovea area) and the resolution/sensitivity decreases dramatically with increasing eccentricity. A large number of recent psychological researches reveal that visual attention is the fundamental reason of the enhancement of visual resolution and sensi-

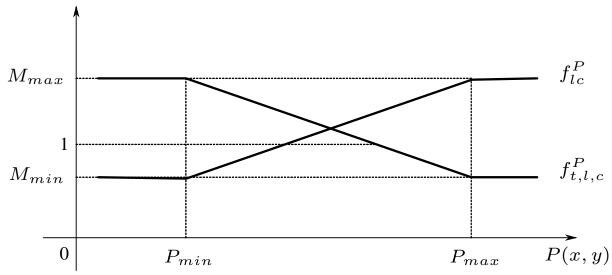

Fig. 2. Diagram of mapping function from PQSL value to modulation value.

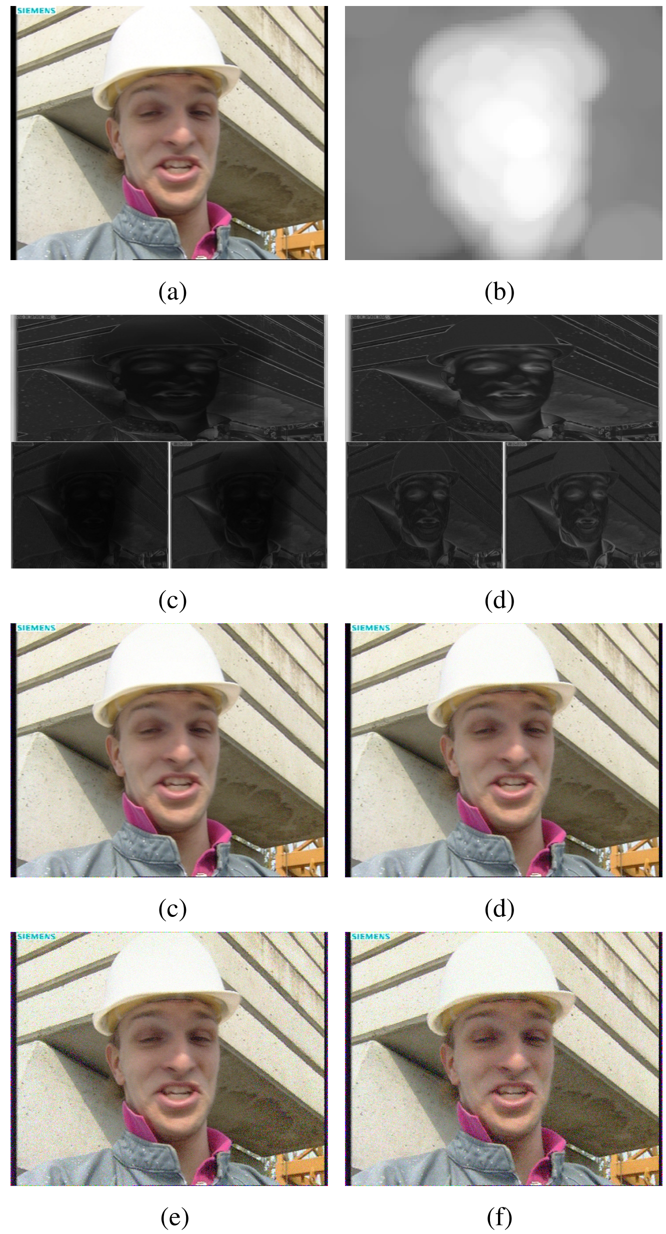

Fig. 3. Embedding noise into the 30th frame of video 'Foreman': (a) original image; (b) spatial selectivity measure P; (c) Θ^P generated by proposed model; (d) Θ generated by Yang's model; (e) noise-embedded image with $\Theta^P(x,y,t)$, $\alpha = 1$, $PSNR = 32.27dB$; (f) noise-embedded image with $\Theta(x,y,t)$, $\alpha = 1$, $PSNR = 32.84dB$; (g) noise-embedded image with $\Theta^P(x,y,t)$, $\alpha > 1$, $PSNR = 24.61dB$; and (h) noise-embedded image with $\Theta(x,y,t)$, $\alpha > 1$, $PSNR = 25.18dB$;

tivity in attentional area [8, 9, 10]. Visual attention and eye fixation position/movement in visual field is highly correlated; the eye fixation always follows the shift of visual attention in visual field to capture more useful information, although they don't have physical link [14]. Some psychological researches on motion perception further reveal that contrast sensitivity of surrounding areas is suppressed by motion of fixative objects [15]. Generally, fast motion has a stronger suppression on visual sensitivity, and the suppression by slow motion is quite weak. From another point of view, the perception of fast motion requires a large amount of brain's computational resources.

Visual sensitivity indicates the HVS' threshold with fixation, and spatial selectivity gives the extent of visual attention. For optimal noise shaping, these two have to be considered simultaneously. In our earlier work [16, 17], spatial selectivity has been estimated for perceptual quality evaluation purpose, with consideration of various salient visual features (e.g., motion, color, luminance, texture); the JND thresholds can be derived for each pixel as the compound effect of luminance adaption and texture masking [5] (as an improved model of [4]). In the next section, the methodology is to be investigated for the spatial selectivity modulated JND profile.

3. COMPUTATIONAL MODEL OF SPATIAL SELECTIVITY MODULATED JND PROFILE

Flowchart of the computational model is given in Figure 1. The spatial selectivity estimation part is adopted from [17], and the JND model is adopted from [5].

The JND for a pixel can be expressed as [5]:

$$\Theta_s = \Theta_l + \Theta_c - C_{lc} \cdot \min(\Theta_l, \Theta_c) \quad (1)$$

where Θ_s is the spatial threshold, Θ_l is the threshold caused by luminance adaptation, Θ_c is the threshold caused by texture masking, and C_{lc} is the coupling factor. With temporal masking, the final JND is obtained as :

$$\Theta = \Theta_s \cdot \Theta_t \quad (2)$$

where Θ_t is the re-correction function for temporal masking. The advantages of Yang's model are: 1) low complexity; 2) better accuracy; 3) pixel based JND profile; and 3) in $YCbCr$ domain.

The proposed spatial selectivity modulated JND profile can be expressed as:

$$\Theta_s^P = \Theta_l^P + \Theta_c^P - C_{lc}^P \cdot \min\left(\Theta_l^P, \Theta_c^P\right) \quad (3)$$
$$\Theta^P = \Theta_s^P \cdot \Theta_t^P \quad (4)$$

where Θ^P, Θ_s^P,Θ_t^P, Θ_l^P, Θ_c^P and C_{lc}^P denote the modulated versions of the variables defined earlier in this section.

The modulation functions can be expressed as:

$$\Theta_t^P(x,y) = \Theta_t(x,y) \cdot f_t^P(P(x,y)) \quad (5)$$
$$\Theta_l^P(x,y) = \Theta_l(x,y) \cdot f_l^P(P(x,y)) \quad (6)$$
$$\Theta_c^P(x,y) = \Theta_c(x,y) \cdot f_c^P(P(x,y)) \quad (7)$$
$$C_{lc}^P(x,y) = C_{lc}(x,y) \cdot f_{lc}^P(P(x,y)) \quad (8)$$

where $P(x,y)$ is the estimated spatial selectivity measure. $f_t^P()$, $f_l^P()$, $f_c^P()$ and $f_{lc}^P()$ are the corresponding modulation functions, as exemplified in Figure 2. In general, with higher spatial selectivity measure, $f_t^P()$, $f_l^P()$ and $f_c^P()$ take lower values, and $f_{lc}^P()$ takes higher value. The actual parameters of these modulation functions may be adjusted.

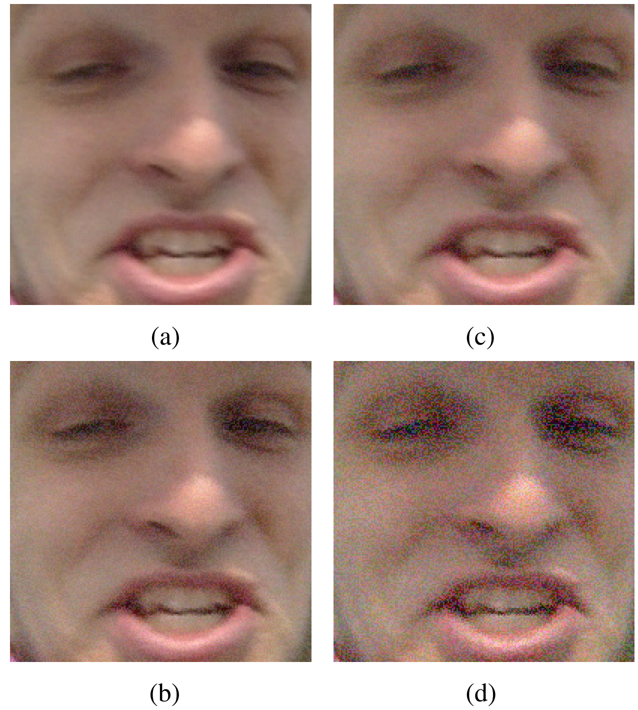

Fig. 4. Closer comparison for the region with the highest selectivity of Figure 3: (a) details of Figure 3(e); (b) details of Figure 3(f); (c) details of Figure 3(g); and (d) details of Figure 3(h);

4. EXPERIMENTAL RESULTS

To evaluate the proposed computational model for noise shaping, a noise embedding scheme is used. The embed noise for the images in our tests is obtained by $\Theta^P(x,y,t)$:

$$d(x,y,t) = \alpha \cdot \Theta_o(x,y,t) \cdot sgn(random()) \quad (9)$$

where Θ_o can be $\Theta^P(x,y,t)$ or $\Theta(x,y,t)$; $\alpha \le 1$ for perceptually lossless noise (if the visibility threshold is correctly determined), and $\alpha > 1$ for perceptually lossy noise; $random()$ is a random noise generator, and it is used here just to control the way to embed noise: addition or subtraction.

Such a noise embedding scheme can be used to examine the performance of $\Theta^P(x,y,t)$ against $\Theta(x,y,t)$ (or any other relevant models for visibility threshold determination). A more accurate JND model should derive a noise embedded image (or video) with better visual quality under a same level of noise (controlled by α), because it is capable of shaping more noise onto the less perceptually significant regions in the image.

Figure 3 shows the experimental result on video clip 'Foreman'. Figure 3(a) is the original 30$^{\text{th}}$ frame, Figure 3(b) is the estimated spatial selectivity measure with the scheme devised in [16, 17]. As can be seen, the human face and the helmet have been correctly identified as the areas with high selectivity (visual attention). Skin color and relative motion are the main reasons for the face area to be identified as the region with the highest selectivity by the said detection scheme. As for the helmet area, relative motion and the gradient of spatial selectivity makes it the region with second highest selectivity in the image. Such salient regions have been confirmed by human observation by different subjects. Figure 3(c) is the JND profile Θ^P generated by proposed spatial selectivity modulated JND profile model. The upper half is the JND profile on Y channel, the lower left part is on Cb channel, and the lower left part is on Cr channel. They are scaled to fit in this paper. Figure 3(d) is the JND profile Θ generated by Yang's model.

Figure 3(e)(f) and (g)(h) are two pairs of comparison between $\Theta^P(x,y,t)$ and $\Theta(x,y,t)$, with very similar resultant PSNR. Detailed comparison of them on the highest-selectivity region are shown in Figure 4 for better appreciation. The pair as shown in Figure 3(e)(f) is with similar PSNR values around 32 dB, and the subjective distortion in Figure 3(e) is almost not noticeable in comparison with original frame; For the pair in Figure 3(g)(h), the overall subjective quality of the video associated with the proposed model ($\Theta^P(x,y,t)$) is better than that associated with $\Theta(x,y,t)$. The reason is, as aforementioned, that more noise is allocated to the areas with low spatial selectivity values under the guidance of $\Theta^P(x, y, t)$. Experimental results on other video clips, which are provided for download in [18], also confirms that spatial selectivity-modulated JND model improves overall perceptual quality of the noise-embedded video.

5. CONCLUSIONS

In this paper, the physiological and psychological evidence is firstly introduced towards the visual sensitivity and spatial selectivity in the human vision system (HVS). The overall visibility threshold at each pixel in an image is affected by both the visual sensitivity and spatial selectivity. In the current estimators for Just-Noticeable-Distortion (JND), spatial selectivity (mainly luminance adaptation and texture masking with the pixel-based approaches) is usually considered. We devise a computational model for incorporating a spatial selectivity measure into the JND profile so that more effective noise shaping is possible in various applications. Experimental results with noise-embedded video sequences confirm that introduction of visual sensitivity enhances the performance of JND profile used in noise shaping.

The benefits of a more accurate visibility threshold determination process can be translated into resource (computation, bitrate, etc.) saving in image/video compression, and performance (e.g., resultant visual quality) enhancement in both image/video compression and visual data hiding (such as watermarking).

6. REFERENCES

[1] Hector Yee, Sumanta Pattanaik, and Donald P. Greenberg, "Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments," in *ACM Transactions on Graphics*, pp. 39–65. ACM Press, 2001.

[2] Shyh shiaw Kuo and J. D. Johnston, "Spatial noise shaping based on human visual sensitivity and its application to image coding," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 509 –517, May 2002.

[3] 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.

[4] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 467–476, Dec. 1995.

[5] 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.

[6] A. J. Ahumada and Heidi A. Peterson, "Luminance-model-based dct quantization for color image compression," in *SPIE proceedings of Human Vision, Visual Processing, and Digital Display III*, 1992, vol. 1666, pp. 365–374.

[7] Andrew B. Watson, "Dctune: A technique for visual optimization of dct quantization matrices for individual images," in *Society for Information Display Digest of Technical Papers XXIV*, 1993, pp. 946–949.

[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] M. Carrasco, C. Penpeci-Talgar, and M. Eckstein, "Spatial covert attention increases contrast sensitivity across the csf: support for signal enhancement," *Vision Research*, vol. 40, no. 10-12, pp. 1203 – 1215, 2000.

[10] Harriet A. Allen and Tim Ledgeway, "Attentional modulation of threshold sensitivity to first-order motion and second-order motion patterns," *Vision Research*, vol. 43, no. 27, pp. 2927 – 2936, December 2003.

[11] James W. Kalat, *Biological psychology (7th ed.)*, Brooks/Cole Publishing, New York, 2001.

[12] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–535, July 1974.

[13] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*, SPIE Press, Bellingham, WA, 1999.

[14] 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.

[15] T. Takeuchi and K. K. De Valois, "Modulation of perceived contrast by a moving surround," *Vision Research*, vol. 40, no. 20, pp. 2697 – 2709, Sept. 2000.

[16] 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.

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