Foveated Frequency Sensitivity의 구현

*Tran, Nhat Huy, **Bui, Minh Trung, ***김원하
경희대학교
*huytn@khu.ac.kr, **minhtrung@khu.ac.kr ***wonha@khu.ac.kr

Desgin of Foveated Frequency Sensitivity

*Tran, Nhat Huy, **Bui, Minh Trung, ***Kim, Wonha
Kyung Hee University
*huytn@khu.ac.kr, **minhtrung@khu.ac.kr ***wonha@khu.ac.kr

요약

We develop the signal processing method for implementing the human perceptual variant on frequency and space. The human visual perceptual sensitivity varies as frequency components and the human perceivable resolution diminishes as the distances further from the eye-focused point. For realizing the frequency sensitivity, we developed the signal direction adaptive multiband energy scaling method to weight the frequency components. The low-pass filtering is designed on the developed energy scaling method for diminishing perceivable resolutions as the deviated distance from the eye-focused point. The developed method not only enhances the frequency components of image signals at the eye-focused region but also smoothes non-perceivable detailed image signals at non-focused regions. The proposed method is verified by the subjective and objective evaluations that it can improve human perceptual visual quality.

1. Introduction

The photoreceptors, in the human eyes, sensing the color and lightness non-uniformly distributes over the retina. The highest density of their distribution is on the fovea aligned with the straight line from the center of eye, that is, visual axis. The photoreceptors density exponentially decreases with respect to the distance away from the fovea [1]. When human view a still image, the attention seems to spread over the whole image causes human eyes have enough time to scan over the whole image. Whereas, motion becomes the most stimuli for human attention in movies, thus human tends pay attention to small regions on the screen during watching movies. In other words, human visual system localizes and tracks some narrow regions on the screen due to the non-uniform distribution of photoreceptors in retina and limited spatial-temporal resolution capacity. Based on this characteristic of human eyes, the conventional foveation filters try to drop out the detail in the non-focused region without enhancing the focused region [2], [3] to archive the compression ration in video coding. The solution for video delivery in case of insufficient network bandwidth is priority video coding. The principle is that the important information gain higher priorities order encoding. One of trend to accomplish this mission is to exploit HVS in coding framework. Wang et al. proposed a foveated visual sensitivity weighting model in wavelet domain to classify the importance of the wavelet coefficients [2]. In [3], Itti uses low level image information such as color, intensity, motion, orientation to automatically generate saliency map. Fixation point in image is corresponding to the high intensity region in saliency map, therefore, the most attractive region gain higher order coding (MPEG-1 and MPEG-4) than the less attractive regions. Another effort for simple generating saliency map in natural images using phase spectrum of Fourier transform was introduced in [4]. By using that saliency map, Guo and Zhang propose a multiresolution foveation model in Wavelet domain to perform priority coding in image and video compression. A foveation-based DCT domain inverse motion compensation video transcoder framework was proposed in [5]. Foveation helps to reduce the required bit rate of video transcoding to the receiver. However, this method depends on the accuracy of local bandwidth estimation, which can cause heavy distortion in their result.

From our viewpoint, a foveated image does not only contain low local contrast in the non-focused regions but also need high local contrast in the focused regions. This desire can be archived through the energy multiband scaling methods, which can control the local contrast efficiently. The distinction of the developed foveation filter is that it simultaneously performs smoothing and enhancing in different regions, whereas the existing foveation filters perform only smoothing. The proposed filter transfers the bandwidth saved in the smoothed regions to the enhanced regions more efficiently. The experiments confirm that the proposed method provides reliable perceptual quality improvements for various sequences and at various bandwidths.

The outline of the paper is organized as follows. In section 2, we show the typical method to control the local contrast. Section 3 elaborates our adaptive multiband energy scaling proposal. The performances of the developed method are discussed in section 4. Finally, the conclusion is given in section 5.

2. Revisit to multiband energy scaling

The goal of foveation filter is to maintain the high visual quality of the locally designated regions and suppress details of the surrounding regions. Thus, it is essential for a foveation filter to perform local processing. Due to ability of local processing and general usages for image processing, we are motivated to use DCT for realizing the foveation filter. The section addresses how the signal energy distribution in the DCT domain is related to the human visual system.

Let f(i;j) be the pixel value at position (i;j) of an N×N image block. Then, the DCT coefficient F(u;v) at position (u;v) of the corresponding image block is obtained such as

$$F(u,v) = C_u C_v \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)\cos\frac{(2i+1)u\pi}{2N} \cos\frac{(2j+1)v\pi}{2N},\quad 0 \le u,v \le N-1 \qquad (1)$$

where

$$C_p = \begin{cases} \sqrt{1/N} & \text{for}\, p=0 \\ \sqrt{2/N} & o.w \end{cases}$$

The collection of DCT coefficients at the same frequency band construct the diamond shape. Denote $\Omega_n\ (0 \le n \le 2N+1)$ as the vector of the DCT coefficients at the n^{th} frequency band. Then

$$\Omega_n = [F(0,n),\ldots F(u,n-u),\ldots F(n,0)] \quad (2)$$

$L(\Omega_n)$ denoting the number denoting the number of the DCT coefficients at Ω_n

$$L(\Omega_n) = \begin{cases} n+1 & \text{for}\, 0 \le n \le (N-1) \\ 2N-n-1 & \text{for}\, N \le n \le (2N-1) \end{cases} \quad (3)$$

A typical method for enhancing the local contrast is the multiband energy scaling [6]. The method exploits the fact that HVS recognizes the local contrast by the energy accumulation ratio. In the method, the local contrast at the n^{th} frequency band is defined as the ratio of the nth band energy versus the average energy of lower frequency bands. The local contrast c_n, at the n^{th} frequency band, is measured as

$$c_n = \frac{\|\Omega_n\|/L(\Omega_n)}{\sum_{k=0}^{n-1}\|\Omega_k\|/L(\Omega_k)} \qquad (4)$$

where ||.|| stands for L1 or L2 norm. The numerator is the energy of the n^{th} frequency band and the denominator is the average of energies at whose frequency bands are lower than the n^{th} band. For enhancing the local contrast, the existing multiband energy scaling method increases the energy at the middle frequency bands through weighted factor greater than 1. Let $\overline{F}(u,v)$ and $\overline{\Omega_n}$ be the DCT coefficient of the modified local contrast and the set of those DCT coefficients at the nth frequency band, respectively. From (4), the modified and original local contrasts are related such as

$$\frac{\|\overline{\Omega_n}\|/L(\Omega_n)}{\sum_{k=0}^{n-1}\|\overline{\Omega_k}\|/(n+1)} = \lambda\frac{\|\Omega_n\|/L(\Omega_n)}{\sum_{k=0}^{n-1}\|\Omega_k\|/(n+1)} \qquad (5)$$

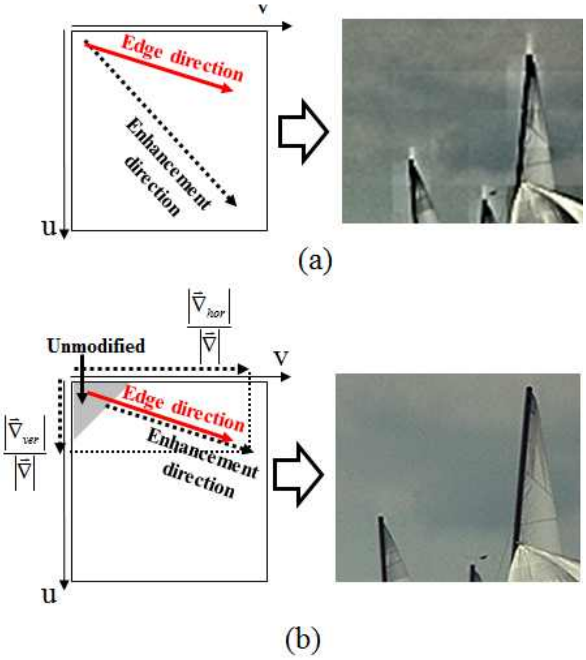

Fig 1. Ringing and blocking artifacts caused by existing method enhancing local contrast. (a) The exsiting method and (b) Proposed method.

where λ is the is the enery weight factor. As becomes larger than 1, the local contrast is more enhanced. More explicitly,

$$\|\overline{\Omega_n}\| = \lambda \circ R_n \circ \|\Omega_n\| \qquad (6)$$

where the energy variation ratio R_n is

$$R_n = \sum_{k=0}^{n-1}\|\overline{\Omega_k}\| / \sum_{k=0}^{n-1}\|\Omega_k\|$$

As seen in Eq.(6), the local contrast modified frequency bands are recursively calculated from the lower frequency bands with updates using the local energy variation ratio. Using L2-norm, the DCT coefficients of the modified contrast are driven from the original ones in following way:

$$\overline{F}(u,v) = \sqrt{\lambda} \bullet \sqrt{R_n} \bullet F(u,v) \text{ for } n = u + v \quad (7)$$

With enhancing the local contrast, the local image signals become more apparent against the overall luminance or the background. However, the existing method increases the local contrasts in all directions without considering he directions of the edge images. As a result, it may bring out overshooting in perpendicular direction to the edges and cause the ringing artifact around the edges. Also, the adjustment of the DCT coefficients at the frequency bands may cause the sudden variations in low frequency energies, which occurs the blocking artifacts making unnatural boundary lines between blocks. Fig. 1 shows the blocking and ringing artifacts caused by the existing multiband energy scaling method and the result of the proposed method.

3. Signal direction adaptive multiband energy scaling

In this section, we propose the new multiband energy scaling method that avoids both blocking and ringing artifacts and is able to enhance local contrast in the high visually stimulated regions and smooth the contrast in the regions with less visual stimuli, simultaneously. For reducing the ringing artifact, we devise a method estimating the gradients of underlying image signals and then enhance or smooth the local contrast in parallel with image signal directions. For avoiding the blocking artifact, we prevent the modification of signal energies at low frequency band. Finally, we control the energy weight factor for controlling the degree of enhancement and smoothness.

3.1 Block gradient

For estimating the gradient of the image signals directly from the DCT coefficients, we adopt the block gradient. The block gradient estimates the overall direction within a block. It is obtained from the variations of sums of half blocks. Previous works had reported the satisfactory accuracies of the blockgradient [7]. The block gradient of a N × N block is calculated as

$$\overrightarrow{\nabla} = \nabla_{ver} \bullet \hat{i} + \nabla_{hor} \bullet \hat{j} \quad (8)$$

where the horizontal block gradient is

$$\nabla_{hor} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N/2-1} [f(i,j) - f(i,N-j-1)]$$

and the vertical block gradient is

$$\nabla_{ver} = \frac{1}{N^2} \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} [f(i,j) - f(N-i-1,j)]$$

The block gradient can be directly derived from DCT coefficients. The DCT coefficients at odd rows of the first column can be derived from Eq. (1):

$$\begin{aligned} &F(2l+1,0) \\ &= \frac{\sqrt{2}}{N} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j) \cos\frac{(2i+1)(2j+1)\pi}{2N} \\ &= \frac{\sqrt{2}}{N} \sum_{i=0}^{N/2-1} \sum_{j=0}^{N-1} [f(i,j) - f(N-i-1,j)] \quad (9) \\ &\qquad \cos\frac{(2i+1)(2l+1)\pi}{2N} \\ &= \frac{\sqrt{2}}{N} \sum_{i=0}^{N/2-1} X_i \cos\frac{(2i+1)(2l+1)\pi}{2N} \end{aligned}$$

where

$$X_i = \sum_{j=0}^{N-1} [f(i,j) - f(N-i-1,j)]$$

3.2 Signal gradient adaptive multiband energy scaling

Based on the block gradient, we optimally adjust the direction of the local contrast modification. We decompose the DCT frequency bands in the horizontal and vertical directions. Let Ω_u^{ver} and Ω_v^{hor} be the u^{th} vertical and v^{th} horizontal frequency bands, respectively. Then,

$$\Omega_u^{ver} = [F(u,0), F(u,1), \ldots\ldots F(u,N-1)] \quad (10)$$
$$\Omega_v^{hor} = [F(0,v), F(1,v), \ldots\ldots F(N-1,v)]$$

Also, denote the local contrast modified uth and vth frequency bands to each direction as $\overline{\Omega_u^{ver}}$, $\overline{\Omega_v^{hor}}$, respectively. In similar way to Eq.(6),

$$\overline{\Omega_u^{ver}} = \lambda \bullet R_u^{ver} \bullet \| \Omega_u^{ver} \| \quad (11)$$
$$\overline{\Omega_v^{hor}} = \lambda \bullet R_v^{hor} \bullet \| \Omega_v^{hor} \|$$

where the energy ratio are

$$R_u^{ver} = \frac{\sum_{l=0}^{u-1} \| \overline{\Omega_l^{ver}} \|}{\sum_{l=0}^{u-1} \| \Omega_l^{ver} \|}, R_v^{hor} = \frac{\sum_{l=0}^{v-1} \| \overline{\Omega_l^{hor}} \|}{\sum_{l=0}^{v-1} \| \Omega_l^{hor} \|}$$

From Eq.(7), the local contrast modified DCT coefficients for each direction becomes

$$\overline{F^{ver}}(u,v) = \sqrt{\lambda} \bullet \sqrt{R_u^{ver}} \bullet F(u,v) \quad (13)$$
$$\overline{F^{hor}}(u,v) = \sqrt{\lambda} \bullet \sqrt{R_v^{hor}} \bullet F(u,v)$$

Then, we can set the directional local controlled DCT coefficient in such way as:

$$\overline{F}(u,v) = \gamma \bullet \overline{F^{ver}}(u,v) + (1-\gamma) \bullet \overline{F^{hor}}(u,v) \quad (14)$$

where γ is the direction weight.

To avoid the blocking effect, we should prevent the modifications of the illuminance that is the overall brightness of a block. Since the energy at the low frequency bands controls the illuminance, we control the local contrast without modifying the components of which frequencies are lower than one third of the highest frequency. Finally, the signal direction adaptive local contrast controlled DCT coefficient is set as

$$\overline{F^{HVS}}(u,v) = \begin{cases} F(u,v)\ \mathrm{for}\, u+v \le \left\lfloor \dfrac{2N-1}{3} \right\rfloor \\ \overline{F}(u,v)\ otherwise. \end{cases}$$


(a) (b)

Sequences	Method	Bit rates (bps)		
		0.3M	0.5M	1M
Paris	Unprocessed	0.68	0.69	0.70
	Conv.	0.70	0.73	0.74
	Prop.	0.74	0.76	0.78
Silent	Unprocessed	0.48	0.52	0.57
	Conv.	0.52	0.57	0.63
	Prop.	0.58	0.63	0.69
Night(HD)	Unprocessed	0.43	0.44	0.47
	Conv.	0.46	0.47	0.50
	Prop.	0.48	0.49	0.53

Fig. 2. The "Paris" sequences (a) Original image, (b) Proposed method given the fixation point at the human faces..

Table I. The CPBD scores that presented scores of unprocessed, conventional and proposed method.

4. Experiments

In order to analyze how the proposed method affects the coding performance, we evaluate their decoded sequences by adopting H.264/AVC JM17.2 for the codec. Fig. 2 demonstrates the result of our proposal on the "Paris" sequences.

We compare three categories of decoded sequences: the sequences processed with the conventional foveation filter, the sequences processed with the proposed foveation filter, and the unprocessed sequences. We measure the CPBD [8] scores of each frame and take average of the scores. Table I shows the average CPBD scores. As seen in the table, the sequences processed by the proposed foveation filter produce the best scores. At the same bit rate, the CPBD scores of the proposed method are better by a minimum of 0.02 and a maximum of 0.07 than the CPBD scores of the conventional method.

5. Conclusion

The motivation of this study is based on the observation that the human visual sensitivity spectrum is realized as foveated frequency sensitivity. For implementing the foveated frequency sensitivity, we developed the DCT based signal direction adaptive multiband energy scaling method for realizing the CSF-oriented perceptual spectrum and embedded the low-pass filter onto the developed multiband energy scaling method for diminishing perceivable resolutions as the distance from the eye-focused point. Essence of the developed method is that the developed method not only enhances the detailed texture signals at the eye-focused region but also smoothes signals at non-focused regions so that it well processes image signals to be more suitable to human visual perception.

감사의 글

이 논문은 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임(No. 2010-0012331)

References

[1] B. Wandell, "Foundations of Vision". Sinauer Associates Inc, May 1995.

[2] Z. Wang, L. Lu, and A. Bovik, "Foveation scalable video coding with automatic fixation selection," IEEE Trans. Image Processing, vol. 12, no. 2, pp. 243 – 254, 2003.

[3] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," IEEE Trans. Image Processing, vol. 13, no. 10, pp. 1304 –1318, 2004.

[4] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," IEEE Trans. Image Processing, vol. 19, no. 1, pp. 185 –198, 2010.

[5] S. Liu and A. C. Bovik, "Foveation embedded dct domain video transcoding," J. Vis. Comun. Image Represent., vol. 16, no. 6, pp. 643 – 667, 2005.

[6] J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," IEEE Signal Process. Lett, vol. 10, pp. 289 – 292, 2003.

[7] C.-B. Wu, C.-Y. Yao, B.-D. Liu, and J.-F. Yang, "Dct-based adaptive thresholding algorithm for binary motion estimation," IEEE Trans. Circuits and Systems for Video Technology, vol. 15, no. 5, pp. 694 – 703, 2005.

[8] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)", IEEE Trans. Image Processing, vol. 20, no. 9, pp. 2678 – 2683, 2011