HUMAN PERCEPTION ORIENTED IMAGE ENHANCEMENT

Seongbae Bang, Wonha Kim

Kyung-Hee University
College of Information and Electronic Engineering
Yongin Korea

ABSTRACT

We developed a DCT domain image enhancement method that not only improves human perceived quality, but also overcomes artifacts that inherently occur from image signal processing. Exploiting an human visual perception model, we developed a perceived contrast measure in the DCT domain. Based on the measure, we designed a filter to make images seem to be vivid as well as finely resolved without boosting up noise. And after investigating how blocking and ringing artifacts occur during processing image signals, we devised a scaling filter that avoids such artifacts. The objective and subjective evaluations verify that the proposed method consistently produces competitive performances.

Index Terms— Image Enhancement, Perceived Contrast, DCT

1. INTRODUCTION

Many psychological and physiological studies have described human visual perception in the frequency domain. The works in [1] represented the perceived sensitivity of an image as perceived contrast and developed a contrast measure at each frequency component. Human visual preference is higher for the image quality with higher perceived sensitivity [2]. And the more perceptually contrasted image signals are more vivid as well as finely resolved so that they seem to be higher resolution [3]. Therefore, the image enhancement method should be developed as a way of increasing the perceived contrast.

Besides human perception, we should treat the issues inherently induced during signal processing. The typical issues include block and ringing artifacts [4]. The ringing artifact occurs when the direction of the image signal processing direction is not in parallel with the edge directions. The block artifact occurs in block processing, such as the Discrete Cosine Transform (DCT). To avoid the artifacts, the processing method is designed to adapt to local image signal properties.

The local contrast enhancement methods are roughly categorized by the spatial, frequency domain and learning-based methods. In the spatial domain, Kou *et al.* [5] and Deng [6] increased local contrast with achieving satisfactory enhancements for well-dedicated images. But, they produced unnatural or over-saturated results in non-dedicated images. As an frequency domian method, Celik [7] conducted the DCT over a whole and increases the sharpness of texture by emphasizing higher frequency components. The multi-band energy scaling method (MESM) in [8, 9] recursively scaled up local contrasts which are measured from frequency band energy ratios. These frequency methods produce robust texture sharpness. However, whole image DCT of Celik is next to impossible in real system, and the MESM causes blocking and ringing artifacts. Recently, the learning-based methods have been developed [10] with data base which is generated manually or by the existing enhancement methods. Therefore, in order to generate the ground-truth enhanced images, the most advanced enhancement method that does not use a learning algorithm must be developed.

In this study, we developed the DCT domain enhancement method considering both human perception and signal processing aspects. In regard to human perception, the proposed method measures the perceived contrast that is higher at higher sensitive frequency components and recursively updates the frequency components to increase the perceived contrast. With respect to image signal processing, we avoid the ringing artifact by tuning the enhancement directions and reduce the block artifact by not modifying low frequency components. Consequently, the proposed image enhancement consists of the filters that improves human perceived quality and prevents artifacts. Objective and subjective tests verify that the proposed method provides reliable and superior performances for various high-resolution images.

2. IMAGE ENHANCEMENT ORIENTED TO HUMAN PERCEPTION

2.1. Human Perceived Contrast in the DCT Domain

Previous studies have modeled human physiological sensitivity as the contrast sensitivity function (CSF) [11] in DCT domain such as

$$CSF(u,v) = \frac{1}{4} \cdot \left[\frac{1}{\phi_u \phi_v} \cdot \frac{\exp(0.18 \cdot \omega(u,v))}{1.33 + 0.11 \cdot \omega(u,v)} \right] \qquad (1)$$

This work was supported by the Korean Minister of Industry under Grant 10067205

where

$$\phi_p = \begin{cases} \sqrt{N} & \text{for p=0,} \\ \sqrt{N/2} & otherwise. \end{cases}$$

Denoting the viewing distance and pixel pitch as v_d and P, the spatial frequency $\omega(u,v)$ in actual viewing conditions is convertible from the DCT frequency as follows [11]:

$$\omega(u,v) = \frac{1}{2N\theta}\sqrt{u^2+v^2} \ \text{ and } \ \theta = 2\cdot\arctan(\frac{\mathrm{P}}{2\mathrm{v}_d}). \quad (2)$$

The background energy against the DCT-frequency component $F(u,v)$ is the accumulation of energy whose frequencies are lower than (u,v) [1], [8]. That is, the background energy $B(u,v)$ at (u,v) is calculated as

$$B(u,v) = \sum_p \sum_q |F(p,q)|^2, \ \sqrt{p^2+q^2} < \sqrt{u^2+v^2}. \quad (3)$$

Then, since humans perceive better the frequency components having higher CSF values and higher energy compared to the background energy, we devised the DCT-domain PC measure in following way:

$$\mathrm{PC}(u,v) = \frac{|F(u,v)|^{2.4}}{CSF(u,v)^{-2} + B(u,v) + |F(u,v)|^2}. \quad (4)$$

2.2. Enhancement of Human Perceived Contrast (PC)

In order for an enhanced image to have better visual quality than the original image, the perceived contrasts of the enhanced image are higher than those of the original image. That is, the enhanced perceived contrast, $\overline{\mathrm{PC}}(u,v)$ must be

$$\overline{\mathrm{PC}}(u,v) > \mathrm{PC}(u,v). \quad (5)$$

By substituting $\overline{\mathrm{PC}}(u,v)$ and $\mathrm{PC}(u,v)$ with (4), the enhanced DCT coefficient, $\bar{F}(u,v)$, is related to $F(u,v)$ as follows:

$$\bar{F}(u,v) > R(u,v)\cdot F(u,v) \quad (6)$$

where

$$R(u,v) = \left\{\frac{CSF(u,v)^{-2} + \bar{B}(u,v) + |\bar{F}(u,v)|^2}{CSF(u,v)^{-2} + B(u,v) + |F(u,v)|^2}\right\}^{1/2.4}.$$

Since $B(u,v)$ and $\bar{B}(u,v)$ are dominant over $|F(u,v)|^2$ and $|\bar{F}(u,v)|^2$, respectively, $R(u,v)$ is well approximated as

$$R(u,v) \approx \left\{\frac{CSF(u,v)^{-2} + \bar{B}(u,v)}{CSF(u,v)^{-2} + B(u,v)}\right\}^{1/2.4}. \quad (7)$$

Introducing the coefficient scaler, $\lambda(u,v)(>1)$, $\bar{F}(u,v)$ would be obtained as follows:

$$\bar{F}(u,v) = \lambda(u,v)\cdot R(u,v)\cdot F(u,v). \quad (8)$$

In texture images, the energy at middle and high frequency components is large, leading to the background energy $\bar{B}(u,v)$ dominating $CSF(u,v)^{-2}$. Then, $R(u,v)$ becomes large at the middle and high frequency components. Conversely, in plain images having low energy at the middle and high frequency components, $CSF(u,v)^{-2}$ dominates $\bar{B}(u,v)$, leading to $R(u,v)$ being almost 1 over the entire frequency range. So, $R(u,v)$ enhances texture images and does not evoke noise in plain images.

3. SIGNAL ADAPTIVE COEFFICIENT SCALING

3.1. Signal Energy Distribution Adaptive Scaling

Occurrences of block artifact depend on local image signal. The edge images having high low frequency energy require wide low frequency ranges to avoid the blocking artifacts, whereas the texture images having small low frequency energy allow narrow low frequency ranges. Thus, the frequency components in the low frequency range should not be modified according to energy distribution in frequency domain.

In [12], the energy distribution is defined as energy ratio(ER) of low, middle and high frequency as follows:

$$\mathrm{ER} = \frac{E_L + E_M}{E_H}. \quad (9)$$

where E_L, E_M and E_H are low, middle and high frequency energy, respectively, which are defined in [12].

Unlike the existing methods fixing low frequency range, we adjusted the low frequency range from the ER. We empirically chose the LR to achieve the maximum enhancement effect while avoiding the block artifact, in the following way:

$$\mathrm{LR} = \begin{cases} \lfloor N/7 \rfloor, & \text{for } 0 \le \mathrm{ER} < 15 \\ \lfloor N/4 \rfloor, & \text{for } 15 \le \mathrm{ER} < 30 \\ \lfloor N/3 \rfloor, & \text{for } \mathrm{ER} \ge 30 \end{cases}. \quad (10)$$

In order to suppress block artifacts, we fixed the coefficient scaler to be 1 for the frequency components in LR, that is, $\lambda(u,v) = 1$ for $\sqrt{u^2+v^2} < \mathrm{LR}$.

The DCT coefficients should be scaled appropriately to features of the frequency components. The coefficient scaler should increase up to middle frequency range to emphasize sharpness, and decrease at the high frequency range so as not to boost noise. It is also necessary to control the enhancement strength in accordance with local signals. When the ER is low, images are likely to be plain images and should be rarely enhanced. Conversely, since images with higher ER are closer to texture, the images are to be more enhanced to achieve better sharpness. Satisfying these demands, we modified the Butterworth function so that the coefficient scaler, $\lambda(u,v)$, and enhancement strength, α, are

$$\lambda(u,v) = 1 + \alpha\cdot\frac{\sqrt{u^2+v^2}/(2N)}{1+\{\sqrt{u^2+v^2}/(2DN)\}^8} \quad (11)$$

and

$$\alpha = \frac{\alpha_{max}}{1+\exp\{-0.1\cdot(\mathrm{ER}-10)\}} + 1$$

where D adjusts the noisy high frequency range. empirically, we set $D = 0.65$.

3.2. Signal Direction Adaptive Scaling

Real image signals are often mixed with non-directional and directional image signals. The plain or texture are usually the non-directional signals and enable full enhancement effect by scaling up frequency components in all directions. The edge is a typical directional signal and needs the directional scaler in parallel to their directions to avoid ringing artifacts. Therefore, we develop the coefficient scaler to be composed of the non-directional and directional parts.

For non-directional image, the non-direction scaler λ^{ND} avoiding block artifacts is

$$\lambda^{\mathrm{ND}}(u,v) = \begin{cases} 1, & \sqrt{u^2+v^2} < \mathrm{LR} \\ 1+\alpha \cdot \dfrac{\sqrt{u^2+v^2}/(2N)}{1+\{\sqrt{u^2+v^2}/(2DN)\}^8}, & \sqrt{u^2+v^2} \geq \mathrm{LR} \end{cases} \quad (12)$$

When an image signal directs vertically, the DCT coefficients at the first row dominate. In horizontal signals, the DCT coefficients at the first column dominate. Therefore, the magnitudes of the DCT coefficients at the first column and row are equivalent to the gradients at each direction [13]. Let ∇_{ver}, ∇_{hor} be the vertical and horizontal gradients, respectively. Then, the gradients of each direction are

$$\nabla_{ver} = \frac{1}{\Gamma}\sum_{u=0}^{N-1}|F(u,0)|, \;\; \nabla_{hor} = \frac{1}{\Gamma}\sum_{v=0}^{N-1}|F(0,v)| \quad (13)$$

where

$$\Gamma = \sum_{u=0}^{N-1}\sum_{v=1}^{N-1}|F(u,v)| - |F(0,0)|.$$

To steer the coefficient scaling direction to be parallel with the edge direction, we decomposed the scaler to horizontal and vertical directions and then weighted each direction gradient to each direction scaler. So, we invented the directional scaler $\lambda^{\mathrm{D}}(u,v)$ in following way:

$$\lambda^{\mathrm{D}}(u,v) = \frac{\nabla_{ver}}{\nabla}\cdot\lambda(u) + \frac{\nabla_{hor}}{\nabla}\cdot\lambda(v) \quad (14)$$

where $\lambda(\ell)$ is the one-dimensional scaler. From (11),

$$\lambda(\ell) = \begin{cases} 1, & \ell < \mathrm{LR}/\sqrt{2} \\ 1+\alpha\cdot\dfrac{\sqrt{2}\ell/N}{1+\cdot\{\sqrt{2}\ell/(DN)\}^8}, & \ell < \mathrm{LR}/\sqrt{2}. \end{cases} \quad (15)$$

To maximize enhancement performance while reducing artifacts, we regulated the ratio between the non-directional and directional coefficient scaler. Since the images having larger ∇ contain more directional signals, ∇ could be a proper barometer estimating the ratio. Borrowing the optimal framework [14], we designed the signal adaptive coefficient scaler, $\lambda(u,v)$, in terms of ∇ in following way:

$$\lambda(u,v) = \frac{\nabla}{\nabla+\epsilon}\cdot\lambda^{\mathrm{D}}(u.v) + \frac{\epsilon}{\nabla+\epsilon}\cdot\lambda^{\mathrm{ND}}(u,v) \quad (16)$$

where ϵ is the fuzzy threshold. We empirically set $\epsilon = 0.25$.

Therefore, the proposed method consists of signal adaptive scaling and perceived contrast elevating parts as follows:

$$\begin{aligned} \bar{F}(u,v) &= \lambda(u,v)\cdot R(u,v)\cdot F(u,v) \\ &= \text{Signal Adaptive Coefficient Scaling} \\ &\quad \cdot\ \text{Human Perception Elevation}\cdot F(u,v) \end{aligned} \quad (17)$$

Table 1. Image quality scores using CPBD

		Spatial Domain		Frequency Domain		
	Original	Unsharp	CAIDE	MESM	CWM	Proposed
Aerial	0.437	0.541	0.538	0.568	0.588	0.686
Bar	0.682	0.780	0.760	0.811	0.797	0.812
Boat	0.223	0.358	0.454	0.413	0.486	0.613
Cross Walk	0.510	0.724	0.778	0.741	0.800	0.850
Market	0.457	0.654	0.702	0.678	0.772	0.811
Square	0.419	0.552	0.598	0.616	0.657	0.689
Tango	0.669	0.795	0.783	0.806	0.843	0.862
Fountain	0.385	0.500	0.530	0.557	0.593	0.680

4. EXPERIMENTS AND DISCUSSIONS

We compared the proposed method and recently developed local contrast enhancement methods including the unsharp masking [6], content adaptive image detail enhancement (CAIDE) [5], multiband energy scaling method (MESM) [9] and coefficient weight method (CWM) [7]. Test images are ultra HD images in [15]. The block size of DCT is 16×16 that is usually used for HD and UHD.

For objective quality evaluation, we employed the cumulative probability of blur detection (CPBD) metric. As the metric score approaches to 1, humans perceive the image quality to be better. Table 1 compares the CPBD scores of the enhanced images. the proposed method consistently produces competitive performances for most of the test images. In particular, the proposed method produces higher values in images, such as Aerial and Market that contain the more detail signals.

For the subjective evaluation, we followed the comparison judgment method of ITU-R BT.500-11 [16]. Images enhanced by each method were displayed at two identical ultra HD(3840×2160) monitors in about 1.5 times monitor height. 20 subjects are invited to assign a score in range [-3, 3] to image pairs. A score of -3 indicates that the left sequence has significantly better visual quality than the sequence on

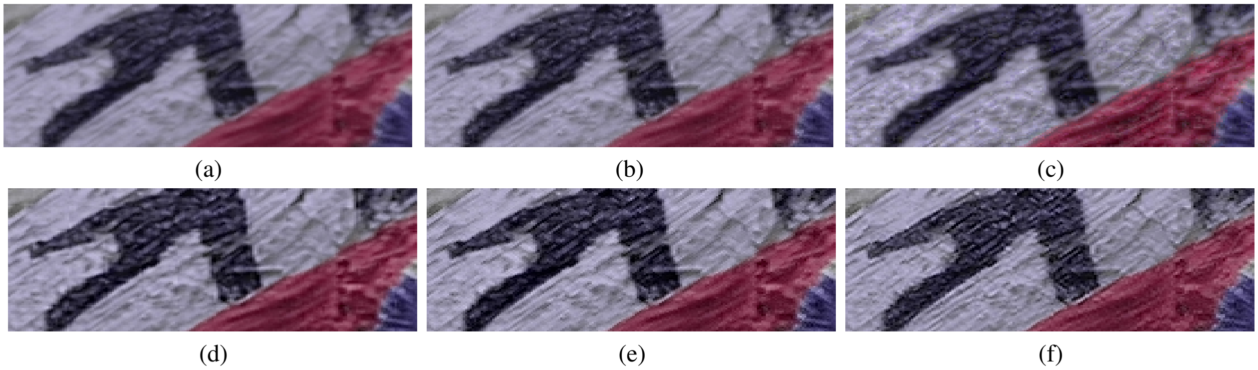

Fig. 1. Enhanced images. (a) Original image, (b) by Unsharp, (c) by CAIDE, (d) by MESM, (e) by CWM, (f) by the proposed method

Table 2. The MOS improvements of enhanced images

	Spatial Domain		Frequency Domain		
	Unsharp	CAIDE	MESM	CWM	Proposed
Aerial	0.150	-0.848	0.730	0.356	0.911
Bar	0.060	-0.208	0.080	0.063	0.080
Boat	0.254	-0.932	0.591	0.666	0.867
Cross Walk	0.140	-0.600	0.231	0.341	0.455
Market	0.132	-0.218	0.060	0.228	0.575
Square	0.041	-0.807	0.116	0.228	0.218
Tango	0.060	-1.182	0.063	0.183	0.254
Fountain	0.231	-0.667	-0.185	0.367	0.463

the right. A score of 3 signifies that the sequence on the right has significantly better quality than the left.

Table 2 shows the MOS improvements of the enhanced images. The proposed method outperformed other methods, especially in Arial and Boat, which contain lots of textures. Fig. 1 compares the images enhanced by each method. As shown in Fig. 1.(f), the proposed method provides the most natural and vivid pattern at wood board.

Fig. 2 analyzes how the proposed method enhances the texture signals. For a better understanding, horizontal signals of enhanced images were plotted. As shown in the signals within the rectangles, the proposed method well intensifies local variations, demonstrating how the proposed method produces better sharpness. The signals in circles show generation of the tiny signals embedded at locally variant signals. Such tiny signals make textures more finely resolved and so produce images that humans perceive to be higher resolution.

Fig. 3 shows occurrence of the blocking, ringing artifact and noise boost-up in images enhanced by MESM, CWM and the proposed method. The MESM causes all artifacts because it does not consider any properties of local signals. By taking DCT over the whole image, the CWM does not evoke block and ringing artifact. But, due to the global processing, it boosts noises and not quite realizable in real systems. On the contrary, in accordance with the local signals, the proposed method adjusts the low frequency range, scaling direction and enhancement strength, and so rarely evokes artifacts as shown in Fig. 3.(d).

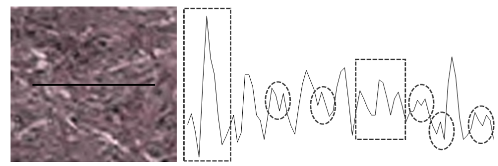

Fig. 2. Enhanced detail signals by the proposed method.

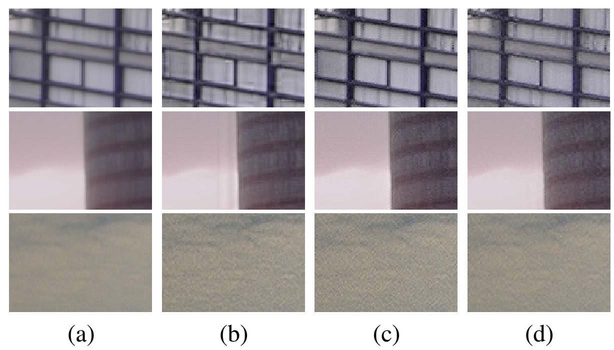

Fig. 3. Artifact occurrences in enhanced images. (a) Original images, (b) by the MESM, (c) by the CWM, (d) by the proposed method

5. CONCLUSION

We devised a filter that increases perceived contrasts. The filter makes textures be vivid as well as finely resolved and does not boost noise in plain images. We also designed a filter that scales frequency energy adaptively to the characteristics of image signals. In adaptation to energy distribution, the filter does not scale the low-frequency components to prevent block artifacts and controls the enhancement strength not to boost up noise. The filter also scales up the DCT coefficients in parallel with signal direction to avoid ringing artifacts. The cooperative integration of two filters enables the proposed enhancement method to not only improve image quality perceived by human but also to overcome artifacts.

6. REFERENCES

[1] A. M. Haun and E. Peli, "Perceived contrast in complex images," *Journal of Vision.*, vol. 13, no. 3, pp. 1–21, Nov. 2013.

[2] B. Spehar, S. Wong, S. van de Klundert, J. Lui, C. W. G. Clifford, and R. P. Taylor, "Beauty and the beholder: the role of visual sensitivity in visual preference," *Frontiers in Human Neuroscience.*, vol. 9, no. 514, Sept. 2015.

[3] O. K. Ersoy S. Aghagolzadeh, "Transform image enhancement," *Optical Engineering*, vol. 31, pp. 31 – 31 – 13, 1992.

[4] Zhen Li and E. J. Delp, "Block artifact reduction using a transform-domain markov random field model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1583–1593, Dec. 2005.

[5] F. Kou, W. Chen, Z. Li, and C. Wen, "Content adaptive image detail enhancement," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 211–215, Feb. 2015.

[6] G. Deng, "A generalized unsharp masking algorithm," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1249–1261, May. 2011.

[7] T. Celik, "Spatial entropy-based global and local image contrast enhancement," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5298–5308, Dec. 2014.

[8] Jinshan Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Processing Letters*, vol. 10, no. 10, pp. 289–292, Oct. 2003.

[9] S. Lee, "An efficient content-based image enhancement in the compressed domain using retinex theory," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 2, pp. 199–213, Feb. 2007.

[10] J. T. Barron S. W. Hasinoff M. Gharbi, J. Chen and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 118, 2017.

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

[12] H. H. Y. Tong and A. N. Venetsanopoulos, "A perceptual model for jpeg applications based on block classification, texture masking, and luminance masking," *Proceedings 1998 International Conference on Image Processing. ICIP98 (Cat. No.98CB36269).*, vol. 3, pp. 428–432, Oct. 1998.

[13] W. Wan, J. Wu, X. Xie, and G. Shi, "A novel just noticeable difference model via orientation regularity in dct domain," *IEEE Access.*, vol. 5, pp. 22953–22964, 2017.

[14] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transactions on Image Processing.*, vol. 18, no. 9, pp. 1921–1935, Sept. 2009.

[15] "Xiph.org foundation, video test media[derf's collection]," [Online]. Available: https://media.xiph.org/video/derf/.

[16] I. Recommendation, *500-11, methodology for the subjective assessment of the quality of television piceture recommendation itu-r bt. 500-11*, 2002.