Pre-processing Algorithm using Sub-band Decomposed Multiscale Retinex for Perceptual Video Coding

Kwang Yeon Choi, Ho Hyeong Ryu and Byung Cheol Song[1]
[1] Department of Electronic Engineering, Inha University
100 Inha-ro, Nam-gu, Incheon, 22212, Republic of Korea
E-mail: [1]bcsong@inha.ac.kr

Abstract: This paper presents a pre-processing algorithm based on sub-band decomposed multi-scale retinex (SD-MSR) for effective perceptual video coding. First, an input frame is decomposed into multiple sub-bands using SD-MSR. Then, a few high-frequency sub-bands are perceptually suppressed by controlling their gains based on a specific model. Finally, the pre-processed video sequence is encoded via a conventional video encoder. Experimental results show that the proposed algorithm can improve the coding efficiency of H.264 by about 12% on average in comparison with before-use.

Keywords-- **Perceptual, Sensitivity, Decomposition**

1. Introduction

With rapid development of digital display devices, the high-resolution video contents are becoming demanding more and more. Also, video coding schemes having high compression efficiency are required. On the other hand, video coding standards such as H.264/MPEG4 AVC and HEVC provide the best solution in terms of rate-distortion (RD) sense. Recently, so-called perceptual video coding (PVC) [1] taking into account human visual system (HVS) under real-life viewing environments is receiving attention. PVC is pursuing the best coding efficiency in terms of visual perception instead of RD cost. In general, PVC achieves coding efficiency by choosing and removing visually insensitive factors among various visual factors such as brightness, motion, edge, and texture.

Li et al. proposed a PVC mechanism using saliency map [2]. Kim et al. presented a JND model where several visual factors are integrated for PVC [3]. However, the above-mentioned algorithms are not compatible with conventional video coding standards because they should modify them.

For compatibility with existing video coding standards, this paper presents a pre-processing algorithm based on sub-band decomposed multi-scale retinex (SD-MSR) [4] for effective PVC. First, an input frame is decomposed into multiple sub-bands using SD-MSR. Then, a few high-frequency sub-bands are perceptually suppressed by controlling their gains based on a specific model. Finally, the pre-processed video sequence is encoded via a conventional video encoder. Experimental results show that the proposed algorithm can improve the coding efficiency of H.264 by about 12% on average in comparison with before-use.

2. Proposed Algorithm

Visually insenstive high-frequency components of an input video sequence is properly removed by the proposed pre-processing based on SD-MSR, and then the pre-processed video sequence is encoded as in Fig. 1.

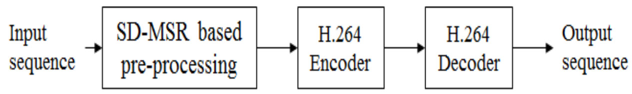

Figure 1. Block diagram of the proposed algorithm.

2. 1 Sub-band decomposition

The first step of the proposed algorithm is the decomposition by a conventional SR-MSR [1]. An input image is decomposed into multiple sub-bands by Eq. (1).

$$\begin{cases} R'_n(x,y) = I(x,y) - F_n(x,y) * I(x,y), & 1 \le n \le N-1, \\ R'_n(x,y) = I(x,y), & n = N, \end{cases} \quad (1)$$

where $R'_n(x,y)$ indicates the n-th scale retinex, and N stands for the number of scales, and $I(x,y)$ is an input image, and $F_n(x,y)$ means a Gaussian filter. * indicates convolution. In order to avoid frequency overlapping between adjacent sub-bands, Eq. (2) is applied.

$$\begin{cases} \overline{R}_n = R'_1, & n = 1, \\ \overline{R}_n = R'_n - R'_{n-1}, & 2 \le n \le N, \end{cases} \quad (2)$$

where $\overline{R}_n$ indicates the final n-th scale retinex. A space constant σ_n for $F_n(x,y)$ must meet the following condition.

$$\sigma_{n+1} > \sigma_n, \; n = 1,2,\ldots,N-1 \quad (3)$$

2. 2 Gain control

A proper gain for each scale retinex is derived, and then the pre-processed image is obtained by Eq. (4).

$$I' = \sum_{n=1}^{N} \overline{R}_n \cdot G_n \quad (4)$$

In Eq. (4), if G_n is less than 1, the frequency component for the n-th sub-band is suppresed. Note that as n becomes larger, $\overline{R}_n$ indicates lower frequency sub-band. We define the sub-bands beyond the *M*-th sub-band as the visually insensitive bands, and the possible minimum of G_1 is set to *t*. Like Fig. 2, we propose a smooth gain control for suppression of $\overline{R}_{M-1}$ to $\overline{R}_1$.

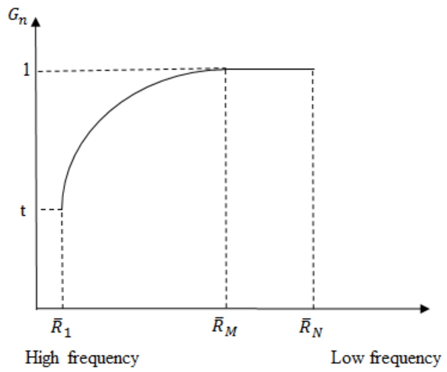

Figure 2. Gain control according to frequecny sub-band.

3. Experimental Results

For performance evaluation, four well-known full HD (1920×1080) video sequences, i.e., Cactus, Toys and calendar, Traffic, and Flag were employed. The first 30 frames of each sequence were employed for this experiment. Also, we adopted additional four full HD sequences which were acquired by the authors, which are defined as test1, test2, test3, and test4. We used Samsung NX1 to obtain these test videos. The first 15 frames of each test sequence were used. We used H.264 JM_VC9 as a video coding platform. Table 1 shows the encoder setting for the following experiments.

Table 1. Encoder setting

Frame rate	30
Quantization parameter	24, 28, 32, 36
Sequence type	IPPP
Motion estimation scheme	EPZS
Number of reference frames	2
RD optimization	Off
Rate control	Off

First, we evaluate the proposed algorithm in terms of subjective visual quality. For the same quantization parameter (QP), the reconstructed frames with pre-processing and without pre-processing are compared. Viewing environment is described in Table 2. Fig. 3 shows some samples in this experiment. Subjects seldom explore any difference between two results.

Table 2. Viewing enviroment

Display	Samsung UN46F8000AF
Size	46”
Type	LED
Resolution	1920×1080
Number of subjects	10
Viewing distance	2H

Next, for quantitative evaluation, we employed multiple scale-structural similarity (MS-SSIM) index [5]. Table 3 compares the result of a conventional H.264 without the pre-processing (JM) and that with the proposed pre-processing (PVC_{Pro}). The numerical values are the averages for all QP’s and all the test sequences. With the same MM-SSIM values, PVC_{pro} achieves bit-saving of about 12%.

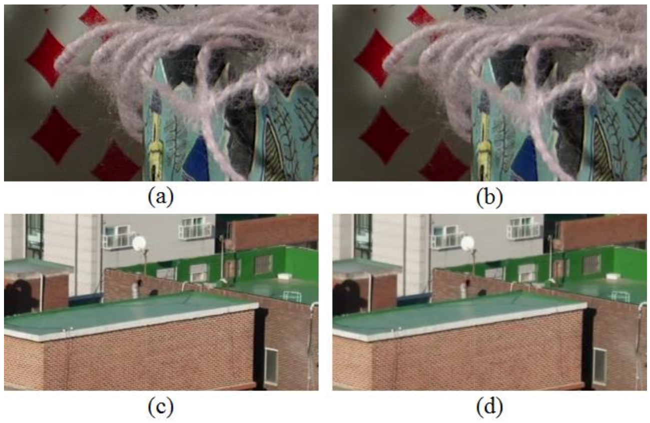

Figure 3. Comparison in terms of subjective viqual quality at QP of 24. (a) the 10-th reconstructed frame of Cactus sequence without pre-processing, (b) the 10-th reconstructed frame with pre-processing, (c) the 10-th reconstructed frame of test1 sequence without pre-processing, (d) the 10-th reconstructed frame with pre-processing.

Table 3. Comparison in terms of ojective visual quality

	JM	PVC_{Pro}
MM-SSIM	0.9795	0.9776
Bitrate(kbps)	6,990	5,918
$\Delta Bitrate$ (%)	-	-11.66

4. Conclusion

For compatibility with existing video coding standards, this paper presents a pre-processing algorithm based on SD-MSR for effective PVC. Experimental results show that the proposed algorithm can improve the coding efficiency of H.264 by about 12% on average in comparison with before-use.

Acknowledgement

This work was supported by the Inha ICT Corporate-connected R&D Personnel Training Project funded by the Korea Institute for Advancement of Technology (KIAT) [N026700017], and was supported by the Industrial Technology Innovation Program funded by the Ministry of Trade, industry & Energy (MI, Korea) [10052982, Development of multi-angle front camera system for intersection AEB].

References

[1] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding*. Boca Raton, FL: CRC Press, Nov. 2005.

[2] Z. Li, S. Qin, and L. Itti, “Visual attention guided bit allocation in video compression,” *Image Vis. Comput.*, vol. 29, no. 1, pp. 1–14, Jan. 2011.
[3] J. Kim, S. H. Bae, and M. C. Kim, “An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 11, pp. 1786-1800, Nov. 2015.
[4] J. H. Jang, B. Choi, S. D. Kim, and J. B. Ra, “Sub-band decomposed multiscale retinex with space varying gain,” *Proc. IEEE Int. Conf. on Image Processing*, pp. 3168–3171, 2008.
[5] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *Proc. IEEE Asilomar Conf. Signals, Systems, Comput.*, vol. 2, pp1398-1402, Nov. 2003.