IEIE Transactions on Smart Processing and Computing, vol. 5, no. 5, October 2016
http://dx.doi.org/10.5573/IEIESPC.2016.5.5.366

Linear Sub-band Decomposition–based Pre-processing for Perceptual Video Coding

Kwang Yeon Choi and Byung Cheol Song

Department of Electronic Engineering, Inha University / Incheon, Korea bcsong@inha.ac.kr

* Corresponding Author: Byung Cheol Song

Received September 23, 2016; Accepted October 24, 2016; Published October 30, 2016

** Regular Paper*

** Extended from a Conference: Preliminary results of this paper were presented at the ITC-CSCC 2016. This paper has been accepted by the editorial board through the regular review process that confirms the original contribution.*

***Abstract*:** This paper proposes a pre-processing algorithm to improve the coding efficiency of perceptual video coding. First, an input image is decomposed into multiple sub-bands through linear sub-band decomposition. Then, the sub-bands that have low visual sensitivity are suppressed by assigning small gains to them. Experimental results show that if the proposed algorithm is adopted for pre-processing in a High Efficiency Video Coding (HEVC) encoder, it can provide significant bit-saving effects of approximately 12% in low delay mode and 9.4% in random access mode.

***Keywords*:** Perceptual video coding, Visual sensitivity, CODEC, Sub-band decomposition

1. Introduction

As digital media applications such as digital cameras, smartphones, tablet PCs, and ultra-high definition TVs proliferate rapidly, and the need for high-quality broadcasting and video on demand (VoD) service increases, high-performance video encoding standards such as H.264 and High Efficiency Video Coding (HEVC) are in greater demand [1-3]. Note that the above-mentioned video compression standards have been optimally designed in terms of rate-distortion (R-D) cost. Beyond the need for a video coder–decoder (CODEC) optimized in terms of R-D cost, video compression optimization in terms of subjective image quality has recently emerged. For example, the so-called perceptual video coding (PVC) technology was developed, which uses the characteristics of the human visual system (HVS) to remove perceptual redundancy [4].

PVC with an emphasis on visual sensitivity is usually categorized into saliency map–based methods [5, 6], and just noticeable difference (JND) model–based methods [7, 8]. Saliency map–based algorithms allocate a larger amount of the bit to the area having higher visual sensitivity. JND model–based methods usually remove the coding elements with low visual sensitivity by coding loop modeling. On the other hand, some methods utilize auxiliary information, such as information pertaining to the viewing environment [9]. Coded elements with low visual sensitivity are removed from an input video by pre-processing prior to video encoding, considering the viewing condition. However, the above-mentioned algorithms still cause visual-quality degradation for areas having low visual sensitivity, and the standardized video coding scheme has to be modified. Furthermore, the use of additional information, such as viewing environment, may be burdensome.

This paper presents a pre-processing technique to remove the psycho-visually redundant high-frequency components that require neither modification of the coding standard nor any prior information about viewing conditions. First, an input image is decomposed into multiple frequency sub-bands through linear sub-band decomposition (LSD). Next, by assigning low gains only for the visually redundant sub-bands, unnecessary high-frequency components are suppressed. After this gain control, the decomposed sub-bands are again synthesized to obtain the final output image. Experimental results show that if the proposed algorithm is employed for pre-processing in a conventional video CODEC, it results in a noticeable bit-saving effect without any perceived visual

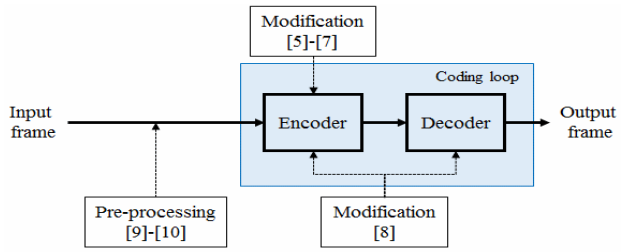

Fig. 1. Conventional PVC algorithms.

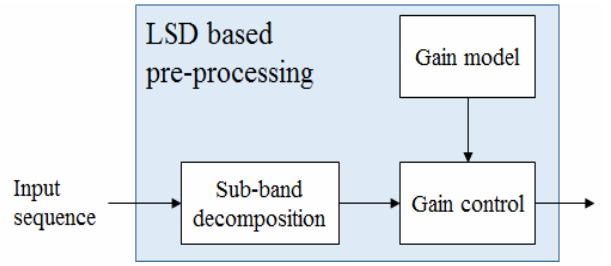

Fig. 2. Detailed processing of LSD-based pre-processing.

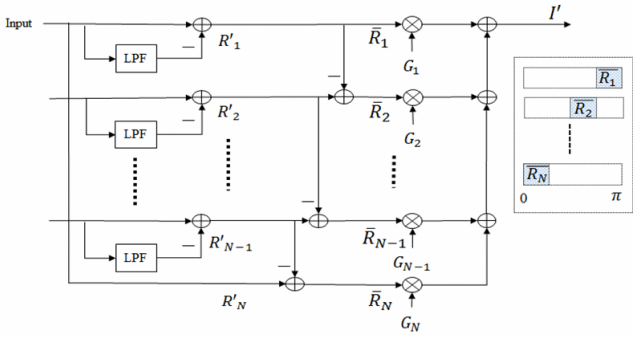

Fig. 3. Schematic of LSD and synthesis.

quality degradation.

This paper is organized as follows. Section 2 reviews the previous work regarding PVC. Section 3 describes the proposed algorithm and Section 4 evaluates it. Finally, Section 5 gives the concluding remarks.

2. Previous Work

Several PVC algorithms considering the HVS [5-10] were developed to improve coding efficiency from the visual perception perspective. As shown in Fig. 1, PVC algorithms can be categorized into in-loop processing and pre-processing approaches. In-loop processing approaches may modify only the encoder [5-7], or both encoder and decoder [8]. For example, Li et al. proposed a visual attention-based bit allocation strategy for video compression [6]. They used saliency-based attention prediction to detect regions of interest (ROIs) in the input video, and generated a guidance map to guide the bit allocation strategy. Finally, perceptual bit allocation is performed based on the guided image. Naccari and Pereira proposed PVC tools, notably decoder-side JND model estimation to allocate the available rate perceptually, with the finest level of granularity, while avoiding the extra rate associated with coding the varying quantization steps [8].

The pre-processing approach employs an adaptive low pass filter (LPF) [9] or foveation model–based filtering [10]. For instance, Vanam et al. proposed a perceptual pre-filter for adaptive VoD content delivery [9]. The filter is based on some parameters of the reproduction setup, such as the viewing distance, pixel density, and display contrast ratio, to remove the spatial oscillations that are invisible under such viewing conditions.

In sum, the use of the above-mentioned approaches results in severe degradation of the perceptual visual quality, due to the unbalanced bit allocation, and they may be incompatible with coding standards. Furthermore, acquisition of prior information about the viewing condition is a very unusual practice.

3. The Proposed Algorithm

In order to overcome the drawbacks of the previous works, this paper presents a novel pre-processing algorithm that decomposes the input image into a few sub-bands using the LSD technique, and suppresses the high-frequency sub-bands having low visual sensitivity.

As shown in Fig. 2, the proposed algorithm consists of LSD, gain control, and synthesis. Gain control is based on the proposed gain model. Since the proposed algorithm is a genuine pre-processing technique to be used prior to video encoders, it is fully compatible with any video coding standards. Furthermore, because the proposed algorithm only handles high-frequency components with low visual sensitivity, it can achieve perceptual bit-saving.

3.1 Sub-band Decomposition

For sub-band decomposition, we employed and modified one of the multi-scale retinex algorithms, i.e., the sub-band decomposed multiscale retinex (SD-MSR) algorithm [12], as shown in Fig. 3. We did not adopt the modified log (mlog) function. Since mlog may change the average luminance level of the input image, it is unfit for the purpose of the proposed algorithm. This is a different thing from the conventional SD-MSR. Fig. 3 describes the LSD technique. In Fig. 3, the n-th scale sub-band for input image $I(x,y)$ is defined as follows:

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

where $R'_n(x,y)$ denotes the n-th scale sub-band, N is the number of scales (i.e., sub-bands), and * indicates convolution. Surround function $F_n(x,y)$ is defined as follows:

$$F_n(x,y) = K_n \cdot e^{-(x^2+y^2)/\sigma_n^2} \quad (2)$$

where σ_n is the Gaussian surround space constant, and K_n is determined such that $F_n(x,y)$ satisfies Eq. (3).

$$\iint F_n(x,y)dxdy = 1 \tag{3}$$

σ_n satisfies the following condition:

$$\sigma_{n+1} > \sigma_n,\ n = 1,2,...,N-1 \tag{4}$$

However, the LSD output of Eq. (1) has overlapping spectral ranges. Hence, it is not efficient to apply gain to LSD output according to spectral characteristics. To alleviate this problem, we decompose the output into nearly non-overlapping spectral bands. This sub-band decomposition can be achieved by using the following equation:

$$\begin{cases} \overline{R}_n = R'_1, & n = 1, \\ \overline{R}_n = R'_n - R'_{n-1}, & 2 \le n \le N, \end{cases} \tag{5}$$

where $\overline{R}_n$ is the result of the n-th scale sub-band. As a result, we can effectively apply gain to each LSD output.

3.2 Gain Control

If the input image is decomposed into several sub-bands, each of them is multiplied with proper gain, and the final image is synthesized as seen in Eq. (6):

$$I' = \sum_{n=1}^{N} \overline{R}_n \cdot G_n \tag{6}$$

where I' denotes the synthesized image, and G_n indicates the gain for the n-th sub-band. According to Eq. (6), a specific sub-band may be emphasized or suppressed by controlling its G_n. Without loss of generality, the HVS tends to be insensitive to high-frequency sub-bands. Therefore, by assigning gain of less than 1 only for the high-frequency sub-bands, we can suppress unrecognizable high-frequency components within the input image. Therefore, this pre-processing technique can save the coding bit-rate without degrading perceptual visual quality.

Let us define the M-th band as the boundary between the unnecessary high- and necessary low-frequency sub-bands. Then, Eq. (6) is modified to Eq. (7):

$$I' = \sum_{n=1}^{M} \overline{R}_n \cdot G_n + \sum_{n=M+1}^{N} \overline{R}_n \tag{7}$$

Note that the gain of the low-frequency sub-bands above the M-th sub-band is set to 1 in Eq. (7). Therefore, the low-frequency sub-bands above the M-th sub-band can be agglomerated as a single sub-band without any decomposition. Finally, the synthesis can be represented by

$$I' = \sum_{n=1}^{M} \overline{R}_n \cdot G_n + \overline{R}_{M+1} \tag{8}$$

where $\overline{R}_{M+1}$ stands for the single low-frequency sub-band above the M-th sub-band base layer.

Finally, appropriate values of σ_n for effective sub-band decomposition, M, (i.e., the boundary between the high-frequency and low-frequency sub-bands) and the separate band-specific gain need to be determined. The values of σ_n and M are empirically determined to avoid any artifacts around the edges, and the related experimental result is depicted in Section 4. Moreover, the model for band-specific gain is described in detail in the following subsection.

3.3 Gain Model

In Eq. (8), the gain for the high-frequency sub-bands below the M-th sub-band should be set to values less than 1, such that bit-saving can be achieved while preserving the same perceptual visual quality. Since an abrupt gain change between neighboring sub-bands may deteriorate visual quality, we need to design a smoothly decreasing function. Therefore, we evaluated several well-known non-linear functions (see Section 4), and finally chose the Gaussian function. Based on a conventional Gaussian function, we modeled the band-specific gain as follows:

$$G_n(x,y) = \exp(-1 \times (\frac{NR_n(x,y)}{\lambda})^{\overline{\sigma_n}}) \tag{9}$$

where λ is a constant used to determine the minimum gain value. Here NR_n and $\overline{\sigma}_n$ denote the normalized $\overline{R}_n$ and σ_n, respectively, and are defined as follows:

$$NR_n(x,y) = \frac{\left|\overline{R}_n\right|}{\max(\left|\overline{R}_n\right|)} \tag{10}$$

$$\overline{\sigma_n} = \frac{\sigma_n}{\max(\sigma_n)} \tag{11}$$

The gain model of Eq. (9) has the following features. First, if NR_n of a certain pixel increases, its gain decreases, because the pixel is regarded as a high-frequency pixel.

Second, cut-off frequency λ decreases gain slowly while maintaining it above a certain value.

Finally, $\overline{\sigma}_n$ accelerates the suppression of high-frequency bands by further decreasing the gain of high-frequency pixels. Fig. 4(a) shows that the gain given by Eq. (9) gradually decreases according to NR_n. It also shows that as n increases, gain decreases. Fig. 4(b) describes the gain model's dependence on λ. In Section 4, proper σ_n values and λ are empirically determined. As a result, the NR_n-adaptive gain given by Eq. (9) effectively suppresses the high-frequency components without any perceptual degradation.

Table 1. Some important parameters for video encoding.

	H.264	HEVC (LD)	HEVC (RA)
Coding structure	IPPP	GPB structure	Clean random access (CRA)
I-frame period	Only first	Only first	8
Quantization parameter	22, 27, 32, 37	22, 27, 32, 37	22, 27, 32, 37
Motion estimation scheme	EPZS	TZ search, Hadamard measure	TZ search, Hadamard measure
Search range	32	64	64
Number of reference frames	2	4	3, 4
Rate control	Off	Off	Off
Entropy coding	CABAC	CABAC	CABAC
RD optimization	Off	Default (QP factor)	Default (QP factor)

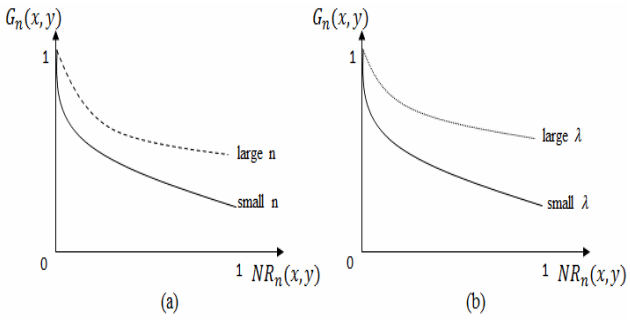

Fig. 4. (a) Gain model according to NR_n and (b) gain model according to λ.

4. Experimental Result

We adopted 10 MPEG test video sequences, which included six 1920×1080 videos (*Cactus, Toys and Calendar, Traffic, Flag Shoot, Park Scene, BQ Terrace*) and four 1280×720 videos (*Big Ship, City, Night, Stockholm*) to evaluate the proposed algorithm. The first 300 frames of each sequence were used for encoding. In addition, we use the four 1080p video sequences captured directly in continuous shooting mode of a specific digital single lens reflex camera, as shown in Fig. 5, named *D*1, *D*2, *D*3, and *D*4. We employed H.264 and HEVC for the following experiments. JM ver. 19.0 [13] and HM ver. 16.10 [14] were used as H.264 and HEVC reference codes, respectively.

Table 1 shows some important parameters for video encoding configuration [15]. Here LD and RA indicate default low delay and random access, respectively. All the experiments were conducted on an Intel i7-3770@3.40Gz CPU with 12 GB RAM.

4.1 Parameter Set-up of the Proposed Algorithm

4.1.1 Determining Bandwidths of LPFs in LSD

In previous works such as Jang et al. [12], the σ corresponding to the highest frequency sub-band was

Fig. 5. First frames of four 1080p video sequences captured directly in the continuous shooting mode of a specific digital single lens reflex camera, denoted as follows (a) *D*1, (b) *D*2, (c) *D*3, (d) *D*4.

usually determined as less than 10. Therefore, this paper sets the maximum value of σ to 10 for visually lossless degradation. In the configuration of Fig. 3, we initially set N to 11. Then, σ_n is set to n for an integer $n < N$. In order to determine an appropriate value for M, we performed an experiment to evaluate the effect of each sub-band on visual quality. In this experiment, if the gain of a certain sub-band is set to 0.7, the gain of the other sub-bands is set to 1.

By lowering the gain of every sub-band starting from the highest frequency sub-band, we can find the initial M at which degradation of visual quality is first observed. We performed this experiment for all the test sequences. The examples in Fig. 6 show that the initial value of M can be 4. When G_5 is set to 0.7, we observed some artifacts around the edges, as shown in Fig. 6(f).

Now, the initial value of M is refined in terms of a quantitative metric, i.e., the bit-saving rate. The bit-rate saving ratio Δbitrate is defined by

$$\Delta bitrate = \frac{bitrate_{con} - bitrate_{pro}}{bitrate_{con}} \times 100 \qquad (12)$$

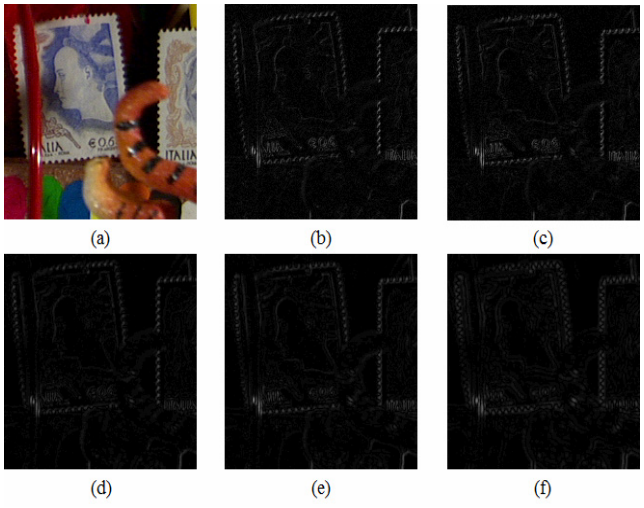

Fig. 6. Difference images between the original and processed images (a) original, (b) G_1 = 0.7, (c) G_2 = 0.7, (d) G_3 = 0.7, (e) G_4 = 0.7, (f) G_5 = 0.7. In (b) to (f), the gain of the other sub-bands is set to 1. In addition, histogram equalization was applied to (b) to (f) for easy comparison.

Table 2. Effect of each sub-band on bit-rate saving, i.e., Δ bitrate[%]. In each column, the gain of the other sub-bands are set to 1.

Test sequence	G_1 = 0.7	G_2 = 0.7	G_3 = 0.7	G_4 = 0.7
Cactus	18.87	16.23	4.30	3.31
Toys and calendar	15.43	13.30	3.19	2.66
Traffic	7.95	6.62	3.97	3.97
Flag shoot	3.47	3.23	1.44	1.92
Park scene	10.46	8.54	3.43	3.32
BQ terrace	26.38	16.35	3.34	2.36
Big ship	17.41	14.18	4.12	3.62
City	20.42	15.38	6.40	5.60
Night	12.60	10.27	4.32	4.17
Stockholm	32.11	21.32	6.27	5.18
D1	14.67	12.67	2.00	1.33
D2	10.42	8.92	3.85	3.85
D3	9.07	8.52	4.67	5.68
D4	8.78	7.97	3.58	4.16
Average	14.86	11.68	3.92	3.65

where $bitrate_{pro}$ and $bitrate_{con}$ stand for the bit-rates of the proposed algorithm and the conventional encoding scheme without the proposed algorithm, respectively. Table 2 shows the bit-saving ratios when the quantization parameter (QP) value is 22.

We see that R1 and R2 have a similar bit-saving effect, and R3 and R4 also show a similar trend. Therefore, the similar sub-bands are merged. Thus, the value of M is adjusted to 2 by re-assigning σ_1 and σ_2 as 2 and 4, respectively. With this parameter set-up, the pre-processing is accomplished according to Eqs. (8) and (9).

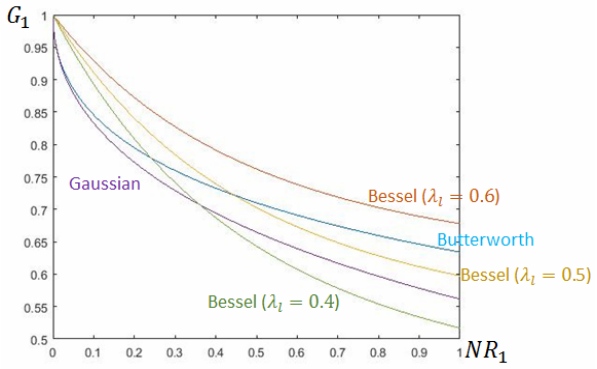

Fig. 7. G_1 graph according to NR_1.

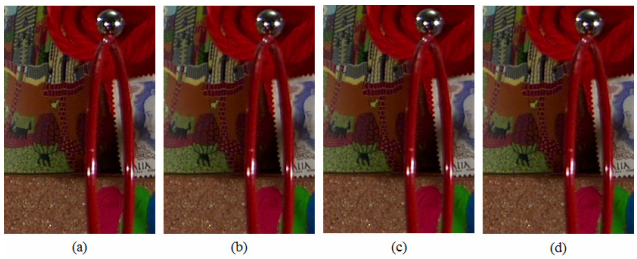

Fig. 8. Comparison of subjective visual quality for different gain models (a) original, (b) Gaussian, (Δ bitrate=18.09%), (c) Butterworth (Δ bitrate=16.72%), and (d) Bessel (Δ bitrate=12.93%).

4.1.2 Determining a Gain Model

In the gain model of Eq. (9), we employed a Gaussian function. In this subsection, we compare the Gaussian model with two different models: one from the Butterworth LPF, and the other from the Bessel LPF. Based on the *n*-th order Butterworth LPF, the Butterworth gain is defined by

$$G_n = \frac{1}{1+(NR_n/\lambda)^{\overline{\sigma_n}}} \tag{14}$$

In addition, based on the transfer function of the third-order Bessel LPF, the Bessel gain model is defined by

$$G_n = (1-\lambda_l)\times\left[\frac{15}{\left(\frac{NR_n}{\overline{\sigma_n}}\right)^3+6\cdot\left(\frac{NR_n}{\overline{\sigma_n}}\right)^2+15\cdot\left(\frac{NR_n}{\overline{\sigma_n}}\right)^1+15}\right]+\lambda_l \tag{15}$$

For example, G_1 can be represented according to different gain models, as in Fig. 7. In Fig. 7, λ of Eq. (14) is set to 3, and three different λ_l values (0.4, 0.5, and 0.6) in Eq. (15) are evaluated. The other parameters in Eqs. (14) and (15) are equivalent to those of Section 4.1.1. Next, we measured the bit-saving effect of different gain models on the H.264 platform. Table 3 shows the result for the

Table 3. Bit-rate saving effect of the different gain models.

QP	Gaussian	Butterworth	Bessel ($\lambda_1 = 0.4$)
22	18.09	16.72	12.93
27	13.58	12.53	11.27
32	9.91	9.02	8.85
37	7.77	7.33	7.28

Table 4. Viewing environment.

Viewing factor	Setting
Display	Samsung UN46F8000AF
Type, size	LED, 46 in.
Resolution	1920 × 1080
Number of subjects	10
Viewing distance	2H

Cactus sequence. We can see that the Gaussian model of Eq. (9) outperforms the other models.

In addition, Fig. 8 compares the subjective visual quality when QP = 22 in Table 3. They all show visual quality similar to the original. Thus, we chose the Gaussian gain model as the best option.

4.2 Subjective Evaluation

In order to evaluate subjective visual quality of the proposed algorithm, we employed a well-known adjectival categorical judgment (ACJ) test [16]. In the ACJ test, a test image and its reference image are simultaneously shown to subjects. Ten male subjects aged 25–30 were selected for this experiment. The subjects choose one of seven scores: much worse: -3, worse: -2, slightly worse: -1, the same: 0, slightly better: 1, better: 2, and much better: 3. H.264 results with the proposed algorithm and without the proposed algorithm correspond to the test image and reference images, respectively. The viewing conditions for this experiment are given Table 4. Table 5 shows the ACJ test results when QP was set to 22. Note that all the parameters were determined as in Section 4.1.1. The average score for all the test images was only -0.14. This indicates that the viewers found it difficult to discriminate between the test images and their reference images.

Fig. 9 compares the decoded images for the 10th frame of the *Cactus* sequence. Here, QP was set to 22. Regardless of the CODEC type, the proposed algorithm did not reduce the visual quality of the decoded images.

For the *D*1 sequence, it is difficult to distinguish the test images from their reference images, as observed in Fig. 10.

4.3 Objective Evaluation

For quantitative evaluation, we adopted a famous object-distortion metric called the multiple scale–structural similarity (MS-SSIM) index [17]. Table 6 shows the MS-SSIM and bit-rate according to the CODEC type. The proposed algorithm (pro) is compared with the conventional approach (con) on each CODEC platform.

Table 5. ACJ score.

Test sequence	Score
Cactus	-0.1
Toys and Calendar	-0.1
Traffic	0
Flag Shoot	-0.2
Park Scene	-0.4
BQ Terrace	0.3
Big Ship	-0.3
City	0
Night	0.1
Stockholm	-0.1
D1	-0.5
D2	0.2
D3	-0.2
D4	-0.7
Average	-0.14

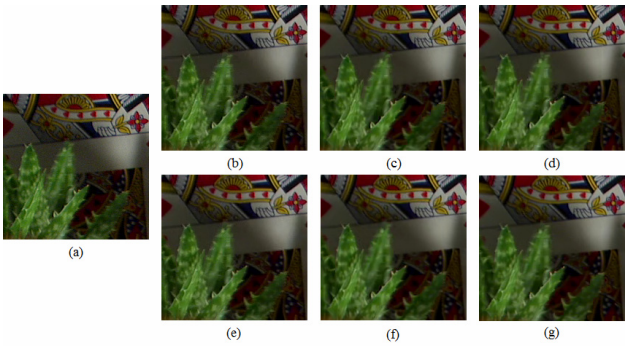

Fig. 9. Decoded images of the 10th frame of the *Cactus* sequence (a) part of the original, (b) H.264, (c) HEVC low delay (LD), (d) HEVC random access (RA), (e) H.264 w/ the proposed algorithm, (f) HEVC (LD) w/ the proposed algorithm, (g) HEVC (RA) w/ the proposed algorithm.

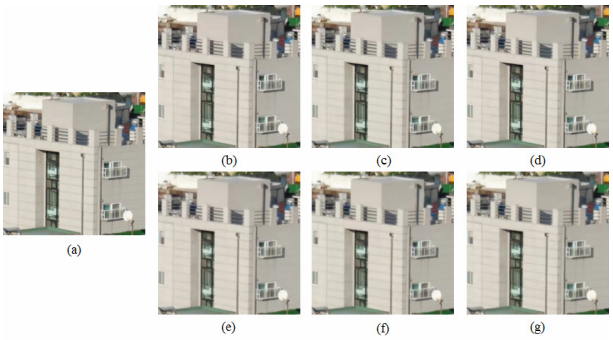

Fig. 10. Decoded images of the 10th frame of the *D*1 sequence (a) part of the original, (b) H.264, (c) HEVC (LD), (d) HEVC (RA), (e) H.264 w/ the proposed algorithm, (f) HEVC (LD) w/ the proposed algorithm, (g) HEVC (RA) w/ the proposed algorithm.

The average values for four QP values of the MS-SSIM and the bit-rate are provided in Table 6. In addition, they are averaged for the entire frame in each sequence. For

Table 6. Comparison in terms of MS-SSIM and bitrate.

Test sequence	H.264					HEVC (LD)					HEVC (RA)				
	MS-SSIM		Bitrate (kbps)		Bitrate (%)	MS-SSIM		Bitrate (kbps)		Bitrate (%)	MS-SSIM		Bitrate (kbps)		Bitrate (%)
	Con	Pro	Con	Pro		Con	Pro	Con	Pro		Con	Pro	Con	Pro	
Cactus	0.9795	0.9779	10,467	8,848	12.34	0.9752	0.9738	3,748	3,273	10.24	0.9781	0.9768	4,996	4,486	9.12
Toys and Calendar	0.9796	0.9788	6,560	5,706	9.30	0.9776	0.9770	2,071	1,841	7.72	0.9803	0.9796	3,180	2,924	6.35
Traffic	0.9840	0.9822	6,370	5,657	10.81	0.9829	0.9812	2,887	2,592	10.04	0.9867	0.9851	5,154	4,732	8.35
Flag Shoot	0.9840	0.9840	3,532	3,436	2.08	0.9829	0.9828	1,369	1,349	1.24	0.9844	0.9843	1,605	1,581	1.23
Park Scene	0.9755	0.9738	8,464	7,663	8.79	0.9709	0.9692	3,720	3,373	8.89	0.9762	0.9745	4,874	4,465	8.49
BQ Terrace	0.9823	0.9796	18,228	13,500	25.60	0.9825	0.9804	5,463	3,814	25.05	0.9842	0.9824	7,423	5,969	16.91
Big Ship	0.9737	0.9704	3,371	2,697	18.50	0.9715	0.9686	1,327	1,076	17.41	0.9785	0.9758	1,959	1,705	12.78
City	0.9819	0.9778	5,327	3,972	22.05	0.9779	0.9738	1,885	1,437	19.89	0.9829	0.9791	2,813	2,411	13.25
Night	0.9859	0.9842	6,399	5,411	14.59	0.9807	0.9789	2,556	2,173	13.06	0.9825	0.9808	3,031	2,703	9.93
Stockholm	0.9771	0.9749	5,934	4,578	19.20	0.9742	0.9723	1,752	1,343	18.06	0.9778	0.9761	2,706	2,310	12.15
D1	0.9862	0.9833	11,884	9,432	19.39	0.9857	0.9830	4,410	3,644	15.87	0.9881	0.9855	6,855	6,006	12.51
D2	0.9802	0.9783	8,411	7,378	10.74	0.9787	0.9770	4,011	3,565	10.44	0.9828	0.9810	5,903	5,376	8.89
D3	0.9684	0.9678	5,670	5,307	4.74	0.9641	0.9636	1,904	1,804	4.26	0.9645	0.9640	2,133	2,015	4.72
D4	0.9744	0.9740	5,465	5,025	7.05	0.9971	0.9709	1,884	1,760	6.33	0.9712	0.9709	1,995	1,864	6.75
Average	0.9795	0.9776	7,577	6,329	13.2	0.9769	0.9752	2,785	2,360	12.04	0.9799	0.9783	3,902	3,468	9.39

ease of comparison, Δ bitrate of Eq. (12) is provided in Table 6.

In terms of MS-SSIM, the proposed algorithm shows slightly lower values, with differences of 0.0018 for H.264, 0.0017 for HEVC LD, and 0.0016 for HEVC RA, in comparison with before-use or no pre-processing. On the other hand, the proposed algorithm provides outstanding bit-rate reduction compared with conventional video CODECs. For example, the proposed algorithm provides an additional bit-saving of 13.2% on the H.264 platform.

Finally, we examined the computational complexity of the proposed algorithm. We implemented the proposed algorithm in C on an Intel Core i7-4790 CPU @3.60 Hz with 16 GB RAM. For H.264 encoding, the additional required computational time is only 0.2%, on average.

5. Conclusions

This paper proposes an LSD-based pre-processing technique for effective perceptual video coding, which can suppress unnecessary high-frequency components without degrading coding efficiency. First, the proposed algorithm decomposes an input image into the proper number of sub-bands via LSD. Then, particular sub-bands having low visual sensitivity are suppressed by assigning a small gain to those sub-bands. Experimental results show that if the proposed algorithm is adopted for pre-processing in an HEVC encoder, it can provide significant bit-savings of approximately 12% and 9.4% in LD and RA modes, respectively, with negligible computational complexity.

Acknowledgement

This research was partly supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF- 2016R1A2B4007353).

References

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, July 2003. Article (CrossRef Link)

[2] G. J. Sullivan, J. R. Ohm, W. J. Han, T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012. Article (CrossRef Link)

[3] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, pp. 1668-1683, Dec. 2012. Article (CrossRef Link)

[4] H. R. Wu and K. R. Rao, *Digital Video Image Quality and Perceptual Coding*, Boca Raton, FL: CRC Press, Nov. 2005. Article (CrossRef Link)

[5] C. W. Tang, C. H. Chen, Y. H. Yu, and C. J. Tsai, "Visual Sensitivity Guided Bit Allocation for Video Coding," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 11-18, Feb. 2006. Article (CrossRef Link)

[6] 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. Article (CrossRef Link)
[7] J. Kim, S. H. Bae, and M. 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. Article (CrossRef Link)
[8] M. Naccari and F. Pereira, "Advanced H.264/AVC-Based Perceptual Video Coding: Architecture, Tools, and Assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 6, pp. 766-782, Jun. 2011. Article (CrossRef Link)
[9] R. Vanam, L. J. Kerofsky, and Y.A. Reznik, "Perceptual Pre-Processing Filter for Adaptive Video on Demand Content Delivery," *Proc. IEEE Int. Conf. on Image Processing*, pp. 2537-2541, 2014. Article (CrossRef Link)
[10] H. Oh and W. Kim, "Video Processing for Human Perceptual Visual Quality-Oriented Video Coding," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1526-1535, Apr. 2013. Article (CrossRef Link)
[11] Z. Wei and K. N. Ngan, "Spatio-Temporal Just Noticeable Distortion Profile for Gary Scale Image/Video in DCT Domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337-346, Mar. 2009. Article (CrossRef Link)
[12] 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. Article (CrossRef Link)
[13] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, Joint Model Reference Software, version 19.0. Article (CrossRef Link)
[14] HM Reference Software 16.10. [Online]. Available: https://hevc.hhi.fraunhofer.de/trac/hevc/browser/tags/HM-16.10, accessed May 31, 2016. Article (CrossRef Link)
[15] K. Suehring and X. Li, "JVET Common Test Conditions and Software Reference Configurations," JVET-B1010, 2016. Article (CrossRef Link)
[16] ITU, "Methodology for the Subjective Assessment of the Quality of Television Pictures," Geneva, Switzerland, ITU-R BT.500-11, 2002. Article (CrossRef Link)
[17] 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. Article (CrossRef Link)

Kwang Yeon Choi received his BSc in electronic engineering from Inha University, Incheon, Korea in 2015. Currently, he is pursuing an MSc in electronic engineering from Inha University. His research interests include image processing, contrast enhancement, and video coding.

Byung Chel Song received his BS, MS, and PhD in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1994, 1996, and 2001, respectively. From 2001 to 2008, he was a senior engineer in the Digital Media R&D Center, Samsung Electronics Co., Ltd., Suwon, Korea. In March 2008, he joined the Department of Electronic Engineering, Inha University, Incheon, Korea, and currently is an associate professor. His research interests are in the general areas of image processing, computer vision, and multimedia system design.