Video pre-processing with JND-based Gaussian filtering of superpixels

Lei Ding, Ge Li*, Ronggang Wang, Wenmin Wang
School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University

ABSTRACT

In this paper, an innovative method of HEVC video pre-processing is proposed. The method applies a simple linear iterative clustering (SLIC), which adapts a k-means clustering to group pixels into perceptually meaningful atomic regions of superpixels. By calculating the average of weighted average of luminance differences around each pixel in the superpixel, a suitable parameter of Gaussian filter for the superpixel is determined. Experimental results show that bit rate can be reduced up to 29% without loss in visual quality.

Keywords: HEVC, JND, superpixel, video pre-processing, Gaussian filtering

1. INTRODUCTION

The recently developed HEVC [1], high efficiency video coding standard, is becoming more and more popular. Its improved compression performance relative to the existing standard is in the range of 50% bit rate reduction.The human eyes perceive images through the human visual system (HVS), which provides a possibility to get a higher video compression ratio. Extensive research has been conducted to improve the performance of encoder conformable with standards.

Video pre-processing [2] can improve the subjective quality of a reconstructed video or reduce the bit rate in the generation of a compressed bit stream. The usual video pre-processing adopted video data spatial filtering, temporal filtering and image sharpening [3]. By applying a visual perception threshold (PTHD) with just noticeable distortion (JND), one can achieve a compression gain up to 10% to 15% by exploiting video data perceptual redundancy [4-6].The Gaussian filter has been widely used for de-noising in image processing. By applying Gaussian filtering, the new value of pixel (x, y) is the weighted average of the pixels around itself. Gaussian filtering makes the image smoother, which means the deviation of the pixels in a coding block will become smaller, and thus will reduce the bit rate in the process of motion estimation, transformation, scaling and quantization. Smooth areas can withstand strong filtering without being noticed, while edge areas or textured areas will be blurred, thus these areas should be filtered slightly or not at all.

The superpixel is an area with similar texture, contour, colour, etc. Superpixel algorithms group pixels into perceptually meaningful atomic regions. They capture image redundancy and provide a convenient primitive from which to compute image features. Algorithms for generating superpixels can be broadly categorized as either graph-based or gradient ascent methods. Graph-based approaches to superpixel generation treat each pixel as a node in a graph. Edge weights between two nodes are proportional to the similarity between neighbouring pixels. The superpixels are created by minimizing a cost function defined over the graph. Whereas the gradient-ascent-based methods start from a rough initial clustering of pixels, iteratively refine the cluster until some convergence criterion is met to form superpixels. SLIC [7], a state-of-the-art method for generating superpixels based on K-means clustering, has been shown to outperform existing superpixel methods.

Human eyes cannot perceive any changes below the JND threshold of around a pixel due to their underlying spatial/temporal masking properties [8]. The sensitivity of distortion by human eyes can vary significantly in different areas of a frame, upon which the JND model is set. The major factors that contribute to the JND model are spatial contrast sensitivity function, luminance adaption [9-10] etc. These factors reflect the texture, edge and boundary of the frame, and the frame can be filtered according to these factors.

Based on the related works above, we present a new approach of incorporating SLIC and JND for video pre-processing in order to reduce the bit rate without loss of visual quality. Subjective and objective evaluation is carried out to verify the effectiveness of the proposed approach.

Visual Information Processing and Communication VI, edited by Amir Said, Onur G. Guleryuz,
Robert L. Stevenson, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9410, 941004
 · doi: 10.1117/12.2083818

2. SLIC SUPERPIXEL SEGMENTATION

Simple linear iterative clustering (SLIC) is an adaptation of k-means for superpixel generation. It has two major merits. The first is that the number of distances calculated in the optimization can be reduced by refining the search space to an area proportional to the superpixel size, and thus reduces the complexity function to be linear. The second is that the weighted distance measurement combines colour and spatial proximity and provides control over the size and compactness of the superpixels. The above merits make SLIC run faster, and more memory efficient than existing methods.

The only parameter need to be determined is K, the desired number of approximately equally sized superpixels. In this paper, the size of superpixels is proposed to be equal to the size of a macroblock (16 x 16). The reasoning behind is as follows: If the size is too small, blocking artefacts arise after filtering; otherwise, too many pixels share the same filter, and as a result, major features may be filtered out, and the whole image will become obscure.

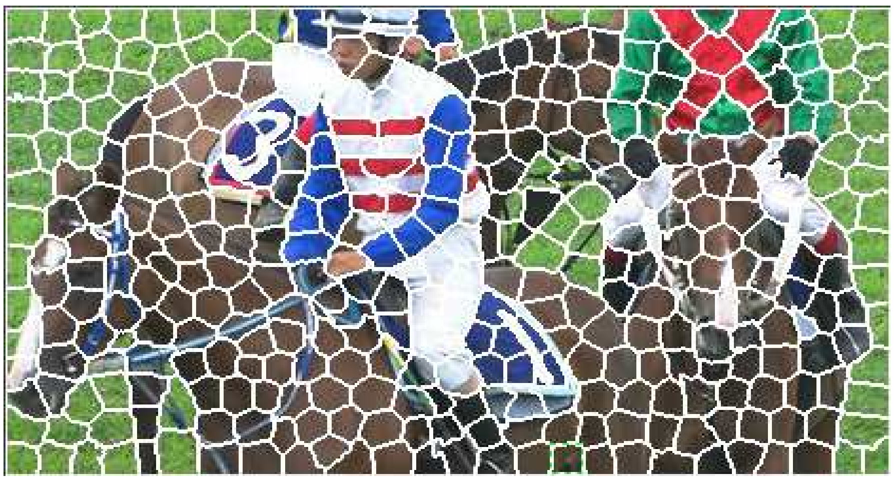


Figure 1. “Racehorse” image segmented into superpixels of size 256 pixels with SLIC

As shown in Figure 1, image is segmented into superpixels that conform to a grid, of which the size is about 256 pixels. Each pixel in the superpixel shares similar luminance, contour, texture etc.

3. JND FACTOR EXTRACTION

The visibility threshold of JND in normal images could be a very complicated function of the factors mentioned above. In the following equations, the factor that reflects the gradients and the texture around the pixel is shown. For the pixel at (x, y), $G(x,y)$denotes the maximal weighted average of gradients around the pixel [11],

$$G(x,y) = \max_{k=1,2,3,4}\{|grad_k(x,y)|\} \tag{1}$$

$$grad_k(x,y) = \frac{1}{16}\sum_{i=1}^{5}\sum_{j=1}^{5} I(x-3+i, y-3+j)\cdot g_k\,(i,j) \tag{2}$$

where $g_k(i,j)$ are four high-pass filters for texture detection as shown in Figure 2.

g_1

0	0	0	0	0
1	3	8	3	1
0	0	0	0	0
-1	-3	-8	-3	-1
0	0	0	0	0

g_2

0	0	1	0	0
0	8	3	0	0
1	3	0	-3	-1
0	0	-3	-8	0
0	0	-1	0	0

g_3

0	0	1	0	0
0	0	3	8	0
-1	-3	0	3	1
0	-8	-3	0	0
0	0	-1	0	0

g_4

0	1	0	-1	0
0	3	0	-3	0
0	8	0	-8	0
0	3	0	-3	0
0	1	0	-1	0

Figure 2. Filters for calculating the weighted average of luminance changes in four directions

As shown in Figure 3, the reconstructed image (a) has almost the same boundaries with the original "Racehorse" image. Apparently, they share the same contour for the feature. In the reconstructed image, three typical areas—A, B and C are selected to show the connection between the value of the maximal weighted average of gradients around the pixel and

(a)

30	36	14	13	11	22	16
9	11	5	10	17	28	9
19	26	20	17	25	36	26
10	12	21	10	12	12	17
16	18	27	19	6	10	6
16	14	13	17	10	12	9
20	28	15	11	6	7	7

25	33	37	39	36	11	31
33	37	39	36	11	31	33
35	31	25	22	2	42	45
27	21	17	13	18	49	28
15	15	15	5	32	47	9
12	19	16	9	44	36	11
21	21	12	19	47	23	11

0	0	0	0	0	0	0
1	1	1	1	0	0	0
2	3	3	2	1	0	1
3	4	4	4	3	2	1
3	2	2	4	6	6	6
3	2	1	2	5	6	6
3	3	2	3	1	2	4

A B C

Figure 3. (a) Reconstructed image of original "Racehorse" image with the maximum weighted average of luminance difference around each pixel,(b) pixel value of area A, B and C of the reconstructed image

the smoothness and visibility threshold. Higher value of weighted average of luminance difference ($G(x,y)$) results in a brighter pixel, and vice versa. As shown in Figure 3 (b), area A, which is located in the grass field of the original image, has lower $G(x,y)$ value than that of Area B, located between the leg of the jockey and the horse, where there is complicated texture and edge. Area C, located at the horseback, has the lowest values less or equal to 10.

It can also be deduced that lower $G(x,y)$ value represents less texture and more smoothness while the higher $G(x,y)$ value means more texture, boundary and surface crease. It can also be deduced that areas with lower $G(x,y)$ values may withstand greater distortion caused by filtering without been discernible, and vice versa.

4. JND-BASED GAUSSIAN FILTERING FOR SUPERPIXELS

Traditionally, Gaussian filtering for the frame with the same parameter leads to blurring effects and significant loss of subjective quality at the same time. Under the circumstances, a novel approach is adopted, with which different areas of the frame can be filtered adaptively. Since each pixel within a superpixel shares the similar texture, smoothness and surface crease, which are correlated with JND factors, therefore different superpixel may require different filters as determined by $G(x,y)$.

As a result, superpixels within the frame are filtered individually with different parameters. A simple formula is given to implement a filter for a superpixel. Let η denote the average of $G(x, y)$ for pixels in the superpixel. A standard deviation for a $k \times k$ Gaussian filter on the superpixel is defined as follows,

$$\sigma = f_s(\eta) = \begin{cases} b - d \cdot \eta, & \eta < \phi \\ c, & \phi \leq \eta \leq \varphi \\ 0 & \end{cases} \tag{3}$$

Here b = 8.5, k = 3, ϕ = 15, φ = 30, c = 1 and d = 0.5 based on our work. σ is linearly negatively correlated with η when η is smaller than ϕ, whereas σ will be set at value c when η is smaller than φ and greater than ϕ. Once η is above the threshold φ, the superpixel will not be filtered. Also, certain pixels, whose values of maximal weighted average of luminance around themselves are bigger than the threshold (set at 15 here), will not be filtered as well. As to the boundary between superpixels, standard deviation will be determined by $G(x, y)$ of the superpixels it belongs to. If the gap of $G(x, y)$ is smaller than a set value, then η will be average of $G(x, y)$ for superpixels sharing the same boundary. Otherwise, the boundary will remain the same. As shown in Figure 4, the pre-processed image obtains a better quality than that of direct Gaussian filtering.

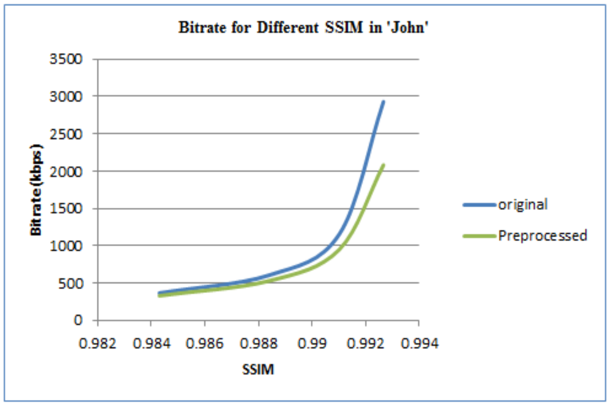
Figure 4. Pre-processed image by means of proposed method and Gaussian filtering

5. EXPERIMENTAL RESULTS

The proposed scheme is implemented based on the HEVC HM 13.0 reference encoder with RDOQ enabled. Sequences of 832x480p, 1280x720p and 1920x1080p with all the frames are used in the experiment. Under the same QP and different bit rates, the subjective quality of the encoded video is compared between the proposed method and the original reference encoder.

The double stimulus impairment scale (DSIS) test method described in ITU-R BT.500 subjective evaluation standard [12] is applied for this subjective evaluation. The reference sequence is shown in Fig.5, followed by the sequence encoded with the pre-processing. A 65'' Skyworth LCD display (65E810U) is used. The viewing distance is about 3 times the image height. Five observers with expertise in image / video processing are asked to give a score from 1 (very annoying) to 5 (imperceptible). The mean opinion score (MOS) is computed as the mean score of all observers. As shown in Table 1, all of the sequences have MOS of 5 or close to 5, and an overall average as 4.97, implying that the perceptual quality is the same as the reference sequences.

Also, the structural similarity (SSIM) index is applied to evaluate the quality. SSIM value is calculated for each encoded sequence by original encoder and the proposed method with the initial uncompressed sequence as reference. As a result, the curves with bitrate for different SSIM in 'John' and 'Kimo' by original encoded sequence and proposed method are drawn in Fig. 5. The result shows a bit rate gain.



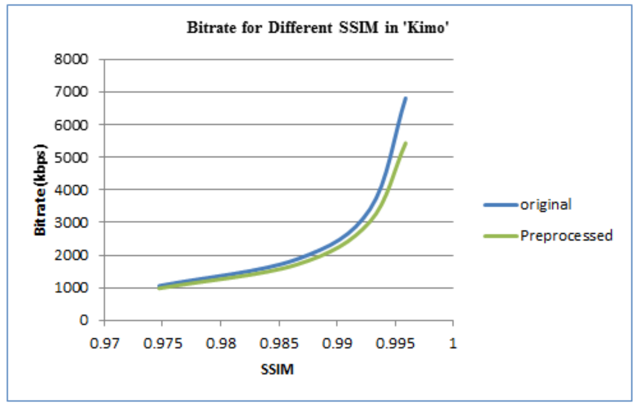


Figure 5. Bitrate for different SSIM in sequence 'John' and 'Kimo' encoded by original reference encoder and pre-processed method

Table 1 shows the bit rate reduction and MOS when the proposed scheme is used. On the average, the bit rate is reduced by 9.3% for all sequences. Fig. 5 shows the average bit rate change at different QPs. In general, a higher resolution and a smaller QP can result in a higher reduction of bit rate, up to 29% without a visual loss. To some extent, Gaussian filtering for superpixel misses the relevance of temporal domain, which causes the bit reduction effect of proposed video pre-processing to diminish. This also occurs as QP increases, since more coefficients are quantized to zero.

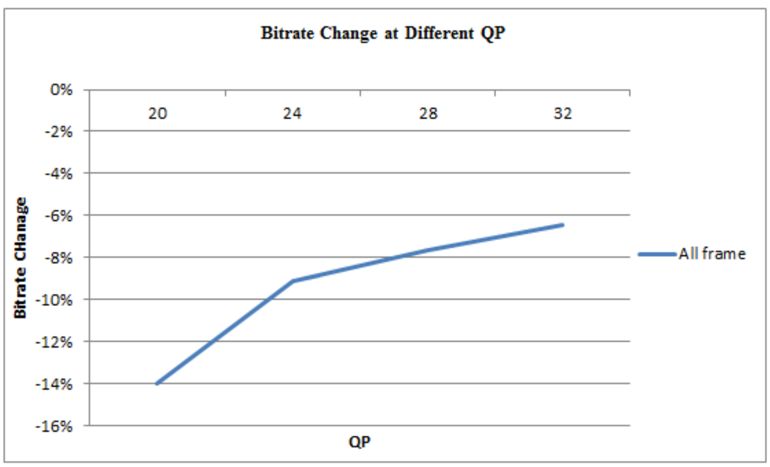


Figure 6. Bit rate changes at different QPs for all frames

6. CONCLUSIONS

In this paper, we proposed a method that exploits the SLIC superpixel and JND in video pre-processing to improve the compression efficiency of HEVC encoder without causing visual loss. The whole frame is handled by means of JND-based Gaussian filtering for superpixels generated by SLIC. This approach reduces the average bit rate by 9.3% without a loss of visual quality. The saving of bit rate is more significant when a video has higher resolution or is encoded at a lower QP.

7. ACKNOWLEDGMENTS

This work was partly supported by a grant of National Science Foundation of China 61370115, Shenzhen Peacock Plan and Shenzhen fundamental research project JCYJ20120614150301623.

Table 1. Bit rate reduction of pre-processed sequences and their MOS.

	Sequence	QP	Bit rate change	MOS
832x480p sequences	BasketballDrill	20	-8.70%	5
		24	-7.90%	5
		28	-6.50%	5
		32	-4.70%	5
	BQMall	20	1.70%	4.6
		24	0.00%	5
		28	-2.40%	5
		32	-3.40%	5
	RaceHorses	20	-9.50%	5
		24	-6.80%	5
		28	-5.40%	5
		32	-4.40%	4.8
1280x720p sequences	Johnny	20	-29%	5
		24	-17.90%	5
		28	-12.40%	5
		32	-9.80%	5
	FourPeople	20	-13.30%	5
		24	-6.00%	5
		28	-5.10%	5
		32	-4.70%	4.8
	KristenAndSara	20	-19.80%	5
		24	-11.10%	5
		28	-8.50%	5
		32	-6.40%	4.8
1920x1080p sequences	Kimono	20	-20.20%	5
		24	-11%	5
		28	-9%	5
		32	-7%	5
	ParkScene	20	-13.2	5
		24	-12%	5
		28	-12%	5
		32	-11%	5

REFERENCES

[1] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Circuits and Systems for Video Technology. Papers 22, 1649-1668 (2012).

[2] K. J. Shen, "Video preprocessing techniques for real-time video compression-noise level estimation techniques," IEEE Circuits and Systems. Papers 2, 778-781 (1999).

[3] S. J. Huang, "Adaptive noise reduction and image sharpening for digital video compression," IEEE International Conference on Computational Cybernetics and Simulation. Papers 4, 3142–3147 (1997).

[4] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao "Motion-compensated residue preprocessing in video coding based on justice-noticeable-distortion profile," IEEE transactions on Circuits and Systems for Video Technology. Papers 15, 742-752 (2005).

[5] J. Geol, D. Chan, and P. Mandi, "Pre-processing for MPEG compression using adaptive spatial filtering," IEEE transactions on Consumer Electronics. Papers 41, 687-691 (1995).

[6] I. Hontish and L. J. Karam "Adaptive image coding with perceptual distortion control," IEEE transactions on Image Processing. Papers 11, 213-222 (2002).

[7] R. Achanta, A. Shaji, K. Smith, A. Lucchi, and P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," IEEE Pattern Analysis and Machine Intelligence. Papers 34, 2274-2282 (2012).

[8] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong, and S. S. Yao, "Just noticeable distortion model and its applications in video coding," Signal Processing: Image Communication. Papers 20, 662-680 (2005).

[9] Y. Jia, W. Lin, and A. A. Kassim, "Estimating just-noticeable distortion for video," IEEE Circuits and Systems for Video Technology. Papers 16, 820-829 (2006).

[10] X. Yang, W. Lin, Z. Lu, E. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," IEEE Circuits and Systems for Video Technology. Papers 15, 742-752 (2005).

[11] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile," IEEE Circuits and Systems for Video Technology. Papers 5, 467-476 (1996).

[12] Methodology for the subjective assessment of the quality of television pictures, ITU-R Recommendation BT.500-11 (2002).