COMPUTATIONALLY EFFICIENT TONE-MAPPING OF HIGH-BIT-DEPTH VIDEO IN THE YCBCR DOMAIN

Zicong Mai, Panos Nasiopoulos, and Rabab K. Ward

Electrical and Computer Engineering Department, University of British Columbia, Canada

ABSTRACT

High dynamic range (HDR) video content is able to provide superior picture quality. This is because the representation of HDR signals requires more bits than the 8-bit low dynamic range (LDR) video. Tone-mapping is the process that converts HDR to LDR signals. Most tone-mapping methods are derived only for the luminance component. This mapping function is then used in each of the R, G and B components to generate the LDR color image. This color tone mapping correction approach, however, cannot be directly applied to most videos since they are usually encoded in the YCbCr color space. This paper addresses this problem and proposes a tone-mapping method that is applied directly on the YCbCr signals. Experimental results show that the Cb and the Cr signals generated by our method are almost identical to those produced with the conventional pipeline up to round-off errors, with average PSNR at about 55 dB and average SSIM at 0.991. By avoiding all the round-off errors introduced in the conventional method, our approach provides a more accurate LDR picture. Moreover, the proposed solution has significantly lower complexity because it bypasses the processes such as color space transformation and up-sampling which are required by the conventional method.

Index Terms— Scalable video coding, tone-mapping, high dynamic range (HDR)

1. INTRODUCTION

Compared to existing 8-bit or low dynamic range (LDR) content, high dynamic range (HDR) videos provide life-like visual experience with significantly higher contrast and richer colors. To represent the larger amount of information, each color component of HDR video needs to be encoded with more bits, e.g., 12 or 16 bits.

Displaying HDR content also requires high-bit-depth output device while most existing displays can support only 8-bit, i.e. low dynamic range (LDR), signals. In order to view HDR content on existing imaging systems, tone-mapping techniques that convert HDR to 8-bit LDR signals are essential. A number of tone-mapping operators (TMOs) have been developed with a variety of designing principals [1, Ch. 6-8]. Tone-mapping is also incorporated in the bit-depth scalable video compression [2]. The majority TMOs are derived for only the luminance channel. To produce color representation, the present common approach is to tone-map the luminance channel first and then to map each of the RGB channels based on the tone-mapping of the luminance [3]. This process is called color correction.

The above color correction approach can be used in a straightforward manner when applied to HDR content represented in the RGB format. However, it cannot be directly applied on videos formatted in the YCbCr color space, which is the most commonly used format in practice. YCbCr is supported by the major image/video encoding standards, including JPEG 2000 and H.264/AVC. After the luminance component (Y) is tone-mapped, the conventional way to obtain the LDR color components (Cb and Cr) of a YCbCr video is to convert the signal from the YCbCr to the RGB space. Then the conventional color correction is performed in the RGB domain, and finally the tone-mapped color signal is transformed back to the YCbCr space. As the available chrominance components Cb and Cr are not the original components, as for most applications they are down-sampled after acquiring them, tone-mapping the YCbCr signals requires that they get up-sampled prior to the RGB conversion. This also adds to the computational load.

Not many studies have been done on color correction in the YCbCr color space for tone-mapping. Toda et al. studied the chromatic treatment of the Cb and Cr channels [4], but their work mainly addresses the creation of YCbCr HDR from LDR content. In this paper, we address the problem of mapping the high-bit-depth Cb and Cr components directly to their 8-bit counterparts, that is, without having to go through the intermediate processes of color space transformation, up-sampling and then down-sampling.

2. PROBLEM STATEMENT

For clarity of presentation, in general, capitalized symbols denote high-bit-depth signals while the lower case ones correspond to their 8-bit counterparts. A symbol with an apostrophe (′) denotes the gamma corrected version.

Fig.1 demonstrates the whole process in the conventional approach for tone-mapping a high-bit-depth YCbCr video frame. The color correction should be

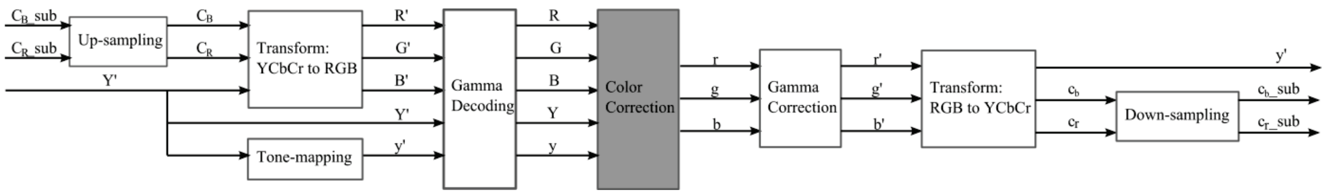


Fig. 1 Pipeline of the conventional method for tone-mapping the Cb and the Cr components.

conducted on the linearized *R*, *G* and *B* signals. The linearized R, G and B signals are the originally acquired signals and are proportional to the light intensity of the scene captured by a camera. Video signals however are usually stored using the non-linear (gamma-corrected) signals for compression purpose. Therefore, prior to color correction, a gamma decoding process needs to be applied prior to color correction so that the gamma-corrected *R'*, *G'* and *B'* signals are converted to their linearized counterparts *R*, *G* and *B*. To obtain the *R'*, *G'* and *B'* signals, a transform from the YCbCr to the *R'G'B'* is required. As mentioned above, the available Cb and Cr are usually sub-sampled from the original chromatic components Cb and Cr. These are denoted by C_B_sub and C_R_sub while the luminance is not sub-sampled. The input C_B_sub and C_R_sub signals need to be up-sampled before the color transform can be applied. The color correction process tone–maps the RGB and luminance Y signals and generates 8-bit *r*, *g*, and *b* signals in the linearized form. The Gamma correction process and the color transform are then performed to produce the full-resolution low dynamic range signal in the YCbCr domain. Finally, the 8-bit Cb and Cr components are sub-sampled.

As can be observed, the above process involves many intermediate stages in order to achieve the tone-mapping of the Cb and Cr components. Our proposed approach aims at generating the tone-mapped YCbCr signal with the least possible computational complexity. In our solution, a high-bit-depth YCbCr video sequence is directly tone-mapped to an 8-bit YCbCr stream (see Fig.2), thus all the intermediate processes, including up-sampling, color transforms, gamma decoding and correction, and down-sampling are bypassed. Our approach also benefits from the fact that it avoids all the round-off errors introduced by the intermediate steps of the conventional method, and, in turn, produces a more accurate LDR representation.

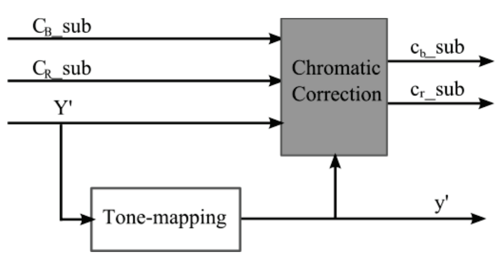


Fig. 2 Framework of the proposed solution.

3. PROPOSED SOLUTION

The proposed solution maps the Cb and the Cr components of the high-bit-depth HDR signal directly to its 8-bit LDR counterpart, given the mapping of the luminance channel. We firstly provide a detailed derivation of the proposed method. The parameter values of our model for two different ITU standards are derived in the second part of the section.

3.1. Derivation

It is challenging to obtain a nice closed-form solution due to the color transforms and the non-linear processes in the conventional method. If one simply cascades all the functions in the blocks of the full pipeline of Fig. 1, a complicated function will result. The computation of this function will be comparable to that of the conventional approach and no advantage is obtained.

Below, we derive an efficient closed-form mapping formula for the Cb component. The derivation can easily be adapted for the Cr component.

The Cb component reflects the difference between the luminance and the blue channels and can be written as:

$$\begin{aligned} C_B &= M_B(B'-Y') + T \\ c_b &= m_b(b'-y') + t \end{aligned} \tag{1}$$

where C_B and c_b denote the Cb component of the high-bit-depth and the 8-bit signals. *T* and *t* are the offsets for HDR and LDR signals respectively. Let *D* denote the maximum intensity value of the high bit-depth signal, e.g., 1023 for a 10-bit signal. The values of *T* and *t* are assigned to be $(D+1)/2$ and 128. The constants M_B and m_b are defined as:

$$M_B = \frac{1}{2\cdot(1-K_B)}\cdot\frac{Q_B}{D} \;\; , \; m_b = \frac{1}{2\cdot(1-k_b)}\cdot\frac{q_b}{255} \tag{2}$$

where the values of K_B and k_b differ in different standards. For example, $K_B = k_b = 0.114$ for ITU-R BT.601 and $K_B = k_b = 0.0722$ for ITU-R BT.709. The factors Q_B and q_b are used to scale the signals to a proper intensity level: Q_B and q_b can be as large as $(D - 1)$ and 254, respectively. These values may vary in different standards as well.

In tone-mapping, one of the most commonly used approaches for color correction is given by [5], [6]:

$$b = \left(\frac{B}{Y}\right)^s y \tag{3}$$

where s is used to adjust the color saturation of the resulting LDR content. Equation (3) here relates to the correction of the blue channel, but similar relationships can be applied to the red and the green channels.

Note that the values of B, b, Y and y refer to the values in the linearized (not gamma-corrected) color space. However, the input and the output in our tone-mapping pipeline refer to variables that have been or should be gamma-corrected (Fig. 1). The following formulae can be used to convert the linearized signals to gamma-corrected signals and vice versa. Let Γ and γ be the values used in the gamma correction for high-bit-depth and 8-bit signals. Equations (4) are used for the luminance channel and equations (5) can be applied to the color channels:

$$\frac{Y'}{D} = \left(\frac{Y}{D}\right)^{\frac{1}{\Gamma}} \quad , \quad \frac{y'}{255} = \left(\frac{y}{255}\right)^{\frac{1}{\gamma}} \tag{4}$$

$$\frac{B'}{D} = \left(\frac{B}{D}\right)^{\frac{1}{\Gamma}} \quad , \quad \frac{b'}{255} = \left(\frac{b}{255}\right)^{\frac{1}{\gamma}} \tag{5}$$

The values of D and 255 are used to normalize the signals to 0-1, which is necessary to keep the output range the same as the input range (0-1). The gamma values for LDR (γ) and HDR (Γ) are allowed to be different.

Equations (1) can be re-organized as:

$$B' = \frac{C_B - T}{M_B} + Y' \ , \ b' = \frac{c_b - t}{m_b} + y' \tag{6}$$

We rewrite equations (4) and (5) in terms of B and b, and substitute them in (3). After simplification we obtain:

$$b' = \left(\frac{B'}{Y'}\right)^{\frac{\Gamma \cdot s}{\gamma}} \cdot y' \tag{7}$$

We substitute the expressions of B' and b' obtained by equations (6) into equation (7). After re-organizing, the closed-form formula that maps the Cb component from high bit depth to 8 bits can be written as:

$$c_b = m_b\left[\left(\frac{C_B - T}{M_B \cdot Y'} + 1\right)^{\frac{\Gamma \cdot s}{\gamma}} - 1\right] \cdot y' + t \tag{8}$$

Note that the above equation is not defined when $Y' = 0$. For this reason, we assign 128 to c_b for this special case since $Y' = 0$ means neutral color information can be found in the c_b channel. Thus, the complete mapping solution for the c_b channel is given by:

$$c_b = \begin{cases} m_b\left[\left(\frac{C_B - T}{M_B \cdot Y'} + 1\right)^{\frac{\Gamma \cdot s}{\gamma}} - 1\right] \cdot y' + t & \text{for } Y' \neq 0, \\ 128 & \text{for } Y' = 0. \end{cases} \tag{9}$$

Similarly, for the c_r component we have:

$$c_r = \begin{cases} m_r\left[\left(\frac{C_R - T}{M_R \cdot Y'} + 1\right)^{\frac{\Gamma \cdot s}{\gamma}} - 1\right] \cdot y' + t & \text{for } Y' \neq 0, \\ 128 & \text{for } Y' = 0. \end{cases} \tag{10}$$

where C_R and c_r denote the Cr component of the high-bit-depth and the 8-bit signals. The constants M_R and m_r have similar definition as M_B and m_b, and their values may vary for different standards or specifications.

If the original high-bit-depth signal is in the 4:4:4 format (all components in full resolution), the resolution of Y' and y' do not need to be changed when applying equations (9) and (10). If the input is in the 4:2:0 format (the Cb and the Cr components are down-sampled by a factor of 2), we should sub-sample Y' and y' before using the proposed solution to map the chromatic components.

3.2. Derivation of parameters for different standards

Below, we derive the parameters M_b, m_b, M_R, m_r, T and t for equations (9) and (10) for the ITU-R BT.601 and ITU-R BT.709 standards. Assuming the input is a 10-bit signal, the value of D is equal to 2^{10}-1 = 1023.

According to ITU-R BT.601, $K_B = k_b = 0.114$ and $q_b = 224$. Let us use the maximum value for Q_B, which is 1022. Using equations (2) we get $M_B = 0.5638$, and $m_b = 0.4957$. Likewise, we can compute M_R and m_r for the Cr component: $M_R = 0.7126$ and $m_r = 0.6266$. The values of T and t can be easily determined to be $T = (1023 + 1)/2 = 512$ *and* t is 128.

Regarding ITU-R BT.709, we can perform similar computations and obtain: $M_B = 0.5384$, $m_b = 0.4734$, $M_R = 0.6344$, $m_r = 0.5578$, $T = 512$, and $t = 128$.

For both cases, the parameter s controls the saturation of the output signal. The larger the value of s, the more saturated is the resulting image/video. Its moderate value is 0.8. The gamma value Γ depends on how the high-bit-depth video is encoded; 2.2 is one of the most commonly used values. The gamma value γ is assigned based on how the 8-bit output should be gamma corrected. The two widely-used values for γ are 1.8 and 2.2

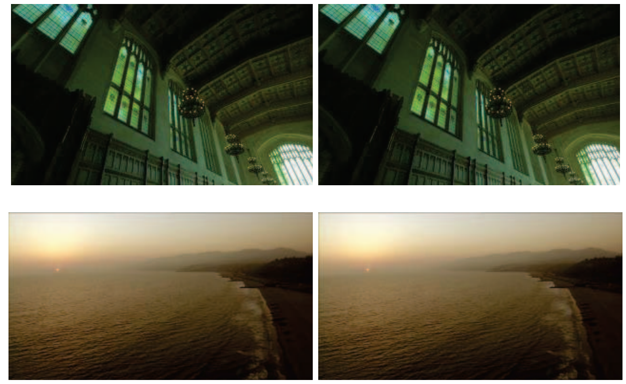
4. RESULTS AND DISCUSSION

In this section, we demonstrate the performance of the proposed approach and compare it with that of the conventional method shown in Figure 2. We choose to compare our method against the conventional approach since in the case of tone-mapping there is no original or ground-truth tone-mapped version for other approaches to be evaluated against. In fact, our proposed method may yield better visual results since it avoids round-off errors that occur in the conventional pipeline. The similarity between the resulting video frames generated with the proposed and the conventional approaches is a good indicator of how successfully our method is in yielding results that resemble those obtained by the conventional approach.

Our chromatic correction approach is independent of any existing tone-mapping operator that is used to generate the 8-bit luminance component. Without loss of generosity, in the experiment we use a tone-mapping method designed

Table 1. Average performance of the proposed approach

	Min	Max	Average
PSNR for Cb (dB)	52.8	58.4	55.1
PSNR for Cr (dB)	52.5	56.7	54.6
SSIM for Cb	0.985	0.995	0.991
SSIM for Cr	0.984	0.995	0.991



Conventional Proposed

Fig. 3 Tone-mapped images using the conventional approach (left column) and the proposed method (right column) for chromatic correction.

for backward compatible HDR video compression [7]. We set s = 0.8 and $\Gamma = \gamma = 2.2$. The constant assigned in (9) and (10) are based on the ITU-R BT.709 standard. Five video sequences, each 60 frames long, were used in the test.

Table 1 shows the peak signal to noise ratio (PSNR) and structural similarity (SSIM) index between the proposed method and the full correction pipeline for 300 video frames. For both quality measures, larger values suggest higher correlation between the two mapping approaches. We observe that the resulting PSNR values range from 52.8 dB to 58.4 dB for the Cb component and range from 52.5 dB to 56.7 dB for the Cr component. The average PSNR is 55.1 dB and 54.6 dB for the Cb and Cr, respectively. These PSNR values are considered to be very high. The SSIM quality metric is agreed to be more faithful to the human visual system and reaches the maximum value of 1 only if the two compared frames are exactly the same. It can be observed from the table that the minimum SSIM value is as high as 0.984 and the average SSIM value is 0.991 for both Cb and Cr, which indicates that the two signals are believed to be perceptually equivalent. The high PSNR and SSIM values indicate that the Cb/Cr components generated using the proposed method and the full conventional pipeline are virtually identical.

Fig. 3 demonstrates the visual similarity between the mapping schemes. The video frames that have been tone-mapped by our efficient method (right) appear to be visually the same to those produced by the conventional method (left). The high similarity stems from the fact that the two schemes are theoretically the same. The difference is caused only by round-off errors introduced in the intermediate steps through the conventional pipeline. Again, the difference does not suggest that our method generates worse results. The fact that our method bypasses round-off errors in its implementation may promise a better tone-mapped signal.

5. CONCLUSIONS

This paper addresses the problem of mapping high-bit-depth Cb and Cr components directly to their 8-bit counterparts. We derive a closed-form mapping solution for the two chrominance components. This solution has significantly lower complexity since it avoids processes such as color space transformation and up-sampling which are required by the conventional method. We choose to compare our method against the conventional approach since in the case of tone-mapping there is no original or ground-truth tone-mapped version for other approaches to be evaluated against. The results show that the Cb and the Cr signals generated by our method the conventional approach are very similar, with average PSNR of about 55 dB and average SSIM at 0.991. The resulting color content also demonstrates the perceptual equivalence between our fast mapping method and the conventional approach. The high correction between the two methods stems from the fact that our method is derived closely based on the conventional pipeline and manages to by-pass all of its intermediate processes. By avoiding all the round-off errors introduced by the intermediate steps of the conventional method, our approach will actually provide higher fidelity.

6. REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, "*High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*," Morgan Kaufmann, 2005.

[2] M. Winken, D. Marpe, H. Schwarz, and T. Wiegand, "Bit-depth scalable video coding," *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 1, pp. I–5–8, 2007.

[3] R. Mantiuk., R. Mantiuk, A. Tomaszewska, and W. Heidrich, "Color correction for tone mapping," *Computer Graphics Forum (Proc. of EUROGRAPHICS)* vol. 28, no. 2, pp.193-202, 2009.

[4] M. Toda, M. Tsukada, A. Inoue, and T. Suzuki, "High Dynamic Range Rendering for YUV Images with a Constraint on Perceptual Chroma Preservation," *16th IEEE International Conference on Image Processing (ICIP)*, pp. 1817 – 1820, 2009.

[5] A. Akyuz and E. Reinhard, "Color Appearance in High-Dynamic-Range Imaging," Journal of Electronic Imaging, vol. 13, no. 1, pp. 126–138, 2004.

[6] S. N. Pattanaik, J. Tumblin, H. Yee, and D. P. Greenberg, "Time Dependent Visual Adaptation for Fast Realistic Image Display," 27th Ann. Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH2000), pp. 47–54, 2000.

[7] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, W. Heidrich, "Optimizing a Tone Curve for Backward-Compatible High Dynamic Range Image/Video Compression," *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 6, pp. 1558-1571, June 2011.