Improving Compression Performance in Bit Depth SVC with a Prediction Filter

S. Aravind Kumar

Anand Institute of Higher Technology, Kazhipattur, Tamil Nadu, India

***Abstract*-** High dynamic range (HDR) is a technique that allows a great dynamic range of luminance between the lightest and darkest area of an image. For video compression, the HDR sequence is reconstructed by inverse tone-mapping a compressed low dynamic range (LDR) version of the original HDR content. In this paper, we show that the appropriate choice of a Tone-mapping operator (TMO) can significantly improve the reconstructed HDR quality. It is used to compress a large range of pixel luminance in to smaller range that is suitable for display on devices with limited dynamic range. we formulate a numerical optimization problem to find the tone-curve that minimizes the expected mean square error (MSE) in the reconstructed HDR sequence. We also develop a simplified model that reduces the computational complexity of the optimization problem to a closed-form solution. It is also shown that the LDR image quality resulting from the proposed methods matches that produced by perceptually-based TMOs

***Index Terms*-** Bit-depth scalable, High dynamic range,video compression, tone-mapping

I. Introduction

Natural scene contain far more visible information that can be captured by the majority of digital imagery and video devices .This is because traditional display devices can only support a limited dynamic range and color gamut. A classic photographic task is the mapping of the potentially high dynamic range of real world luminances to the low dynamic range of the photographic print. This tone reproduction problem is also faced by computer graphics practitioners who map digital images to a low dynamic range print or screen. For video compression, these advances in display technology have motivated the use of extended gamut color spaces. These include xvYCC (x.v.Color) for home theater and the Digital Cinema Initiative color space for digital theater applications. Yet, even these extended color spaces are too limited for the amount of contrast that can be perceived by the human eye. High dynamic range (HDR) video encoding goes beyond the typical color space restrictions and attempts to encode all colors that are visible and distinguishable to the human eye [3], and is not restricted by the color gamut of the display technology used. The main motivation is to create a video format that would be future-proof, independent of a display technology, and limited only by the performance of the human visual system (HVS). HDR images preserve colorimetric or photometric pixel values (such as CIE XYZ) within the visible color gamut and allows for intra-frame contrast exceeding 5–6 orders of magnitude (10^6:1), without introducing contouring, banding or posterization artifacts caused by excessive quantization.

Backward-compatibility can be achieved if the HDR video stream contains 1) a backward-compatible 8-bit video layer which could be directly displayed on existing devices, and 2) additional information which along with this 8-bit layer can yield a good quality reconstructed version of the original HDR content. Such a stream can also contain a residual layer to further improve the quality of the HDR reconstruction. Fig. 1 illustrates the general coding structure used to provide a backward compatible HDR video bit stream.

Several proposals have been suggested to allow the above-mentioned HDR backward-compatibility function within the scalable extension of the H.264/AVC video coding standard [7]–[15]. The contrast of the original HDR content is first quantized into the 8-bit range using a tone mapping operator (TMO) to produce an LDR representation. The LDR sequence is then compressed using a standard video encoder (H.264/AVC). A larger dynamic range video can then be reconstructed by decoding the LDR layer and applying the inverse of the tone-mapping operator to reconstruct the HDR representation. The shape of the TMO can be encoded using supplemental enhancement information (SEI) messages. Finally, a HDR residual signal can also be extracted and encoded in the bit stream as an enhancement layer.

It address the problem of finding an optimal tone-curve for such a backward-compatible encoding scheme. To compute the tone-curve, we propose a method that minimizes the difference in the video quality between the original and the reconstructed HDR video.

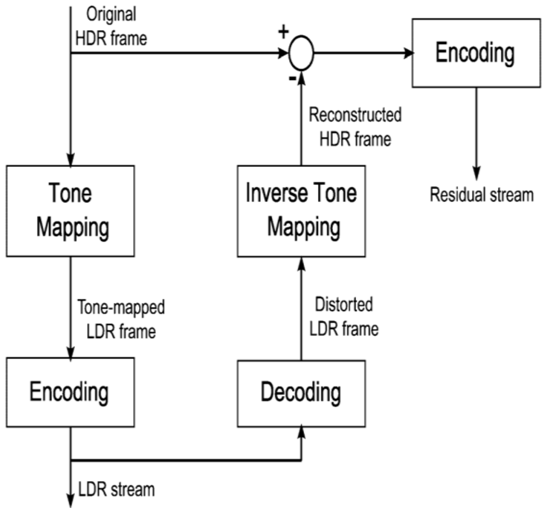

Fig: 1 General structure of the scalable approach used for backward-compatible HDR video encoding. The base layer encodes an 8-bit LDR representation of the HDR input. The enhancement layer encodes the difference (residual) between the inverse tone-mapped base layer and the original HDR source.

The remainder of this paper is organized as follows: an overview of related work is presented in Section II. In Section III, the proposed tone-mapping approach that considers tone-mapping together with compression is discussed in detail. Section IV demonstrates and analyzes the performance of the proposed methods. Finally, we draw our conclusions in Section V.

II. RELATED WORK

Backward compatible HDR video encoding has received significant interest recently. A color space of encoding HDR content based on the luminance threshold sensitivity of the human visual system. They concluded that 10–12 bit luma encoding is sufficient to encode the full range of visible and physically plausible luminance levels. A tone-mapping curve was encoded together with the tone-mapped and residual video sequences. The residual video sequence was additionally filtered to remove the information that is not visible to the human eye.

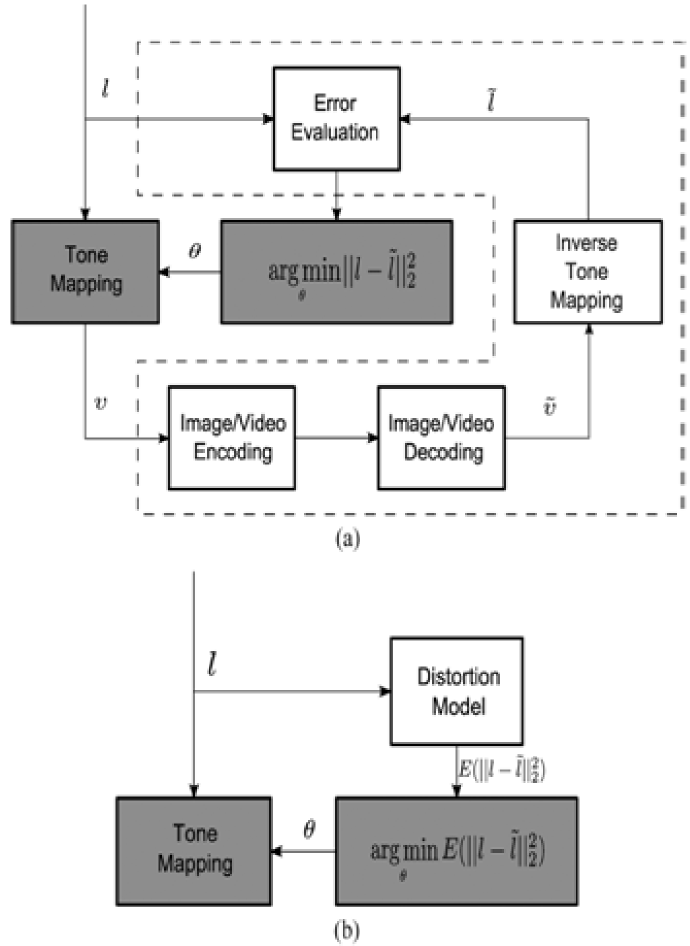

Figure 2: System overview of the proposed tone-mapping method. (a) Demonstrates the ideal scenario where the actual H.264/AVC encoding is employed. (b) Shows the practical scenario which is addressed by this paper.

The primary goal of tone-mapping is to produce the best quality low-dynamic range rendering of an HDR scene that is visually close to the visual high contrast signal.

III. PROBLEM STATEMENT AND PROPOSED SOLUTION

In this section, we present the challenges of obtaining a good quality reconstructed HDR representation in a backward-compatible HDR video encoding system and describe in detail the approach we propose towards overcoming these challenges. The performance of a backward-compatible HDR video and image encoding system depends on the coding efficiency of the LDR base layer and the HDR enhancement layer.

A. Tone-Mapping Curve

The global tone-mapping curve is a function that maps HDR luminance values to either the display's luminance range, or directly to LDR pixel values. The tone-mapping curve is usually continuous and non-decreasing. The two most common shapes for the tone curves are the sigmoidal ("S-shaped") or a compressive power function with an exponent 1 (gamma correction).

The tone-mapping curve can then be uniquely specified by a set of slopes.

$$S_k = \frac{V_{k+1} - V_k}{\delta} \qquad (1)$$

which forms a vector of tone-mapping parameters . Using this parameterization, the forward tone-mapping function is defined as

$$V(l) = (l-l_k).S_k + V_k \qquad (2)$$

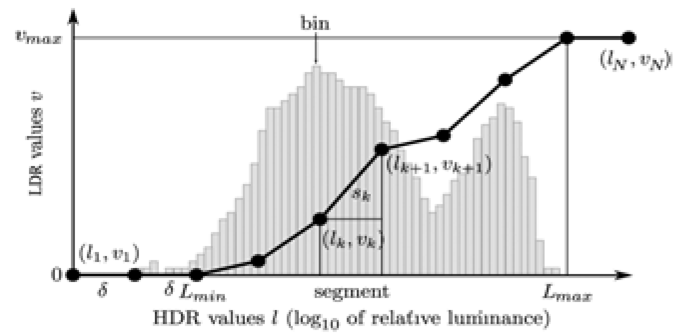

Figure 3: Parameterization of a tone-mapping curve and the notation. The bar-plot in the background represents an image histogram used to compute p(l).

where is the LDR pixel value, is the segment corresponding to HDR value , that is

$l_k \leq l < l_{k+1}$. The inverse mapping function is then

$$\bar{l}(v;S_k) = \begin{cases} \sum_{l \in S_0}^{v-v_k} l.pL(l) \end{cases} \quad \text{for } S_k<0, S_k=0$$

$$\text{where } S_k \in \{ S_1 \ldots S_N \}. \qquad (3)$$

When the slope is zero ($S_{k=0}$), $\bar{l}(v;S_k)$ is assigned an expected HDR pixel value for the entire range S_0 in which the slope is equal zero. $P_L(l)$ is the probability of HDR pixel value .

B. Statistical Distortion Model

As mentioned earlier in Section III, accurately computing the distorted HDR values would be too computationally demanding. Instead, we estimate the error $\left\|\bar{l} - l\right\|_2^2$ assuming that the compression distortions follow a known probability distribution P_C. Under this assumption, the expected value of the error $\left\|\bar{l} - l\right\|_2^2$ is

$$E\left[\left\|\bar{l} - l\right\|_2^2\right] = \sum_{l=l_{min}}^{l_{max}} \sum_{\bar{v}=0}^{v_{max}} \left(\bar{l}(v;s_k) - l\right)^2 \cdot pc(v(l) - \bar{v} / v(l)) \cdot p_L(l) \qquad (4)$$

Where $pc(v-\bar{v}/v)$ is the probability that the encoding error equals $v-\bar{v}$. Note that (2) and (3) show that both and v and l are uniquely determined by the values of and , respectively. It simplify the expression above by removing the dependency of on .Consequently, the continuously relaxed objective function is written as

$$\varepsilon(S_K) = \sum_{l=l_{min}}^{l_{max}} \sum_{\bar{v}=0}^{v_{max}} \left(\bar{l}(v;s_k) - l\right)^2 . pc(v-\bar{v}).p_L(l) \qquad (5)$$

The only unknown variable is the probability distribution of the compression error $pc(v-\bar{v})$, which can be estimated for any lossy compression scheme. In Appendix A we model such distribution for the H.264/AVC I-frame coding.2 However, we will show in Section III-D that the distribution of the compression scheme error is not necessary to calculate a good approximation of the encoding error.

C. Optimization Problem

The optimum tone curve can be found by minimizing the $\varepsilon(S_K)$ function with respect to the segment slopes

$$S_K \ \arg\min \varepsilon(S_K) \qquad (6)$$

where $s_1 \ldots s_N$ Subject to :

$$s_{min} \leq s_K \leq s_{max} \quad \text{for } k=1\ldots N$$

$$\sum_{k=1}^{N} s_K.\delta = v_{max}. \qquad (7)$$

The first constraint restricts slopes to the allowable range, while the second ensures that the tone curve spans exactly the range of pixel values from 0 to v_{max}.The minimum slope s_{min} ensures that the tone-mapping function is strictly increasing and thus invertible and $\bar{l}(v;s_K)$ can be computed. we can write

$\bar{l}(v+1;s_k)-\bar{l}(v;s_k) > \log_{10}(1.01)$ (8)

So that $s_K = (\log_{10}(1.01))^{-1}$ (9)

D. Closed-Form Solution

The distortion model in (5) gives a good estimate of compression errors, but poses two problems for practical implementation in an HDR compression scheme: 1) it requires the knowledge of the encoding distortion distribution pc, and 2) the optimization problem can only be solved numerically using slow iterative. l in the distortion model (5) using the inverse mapping function in (3), this gives

$$\varepsilon(S_k) \approx \sum_{l=l_{\min}}^{l\max} \sum_{\bar{v}=0}^{v_{\max}} pc(v-\bar{v}).$$

$$p_L(l).(\frac{v-\bar{v}}{S_k})^2 \quad .(10)$$

After reorganizing we get

$$\varepsilon(S_k) \approx \sum_{l=l_{\min}}^{l\max} \sum_{\bar{v}=0}^{v_{\max}} pc(v-\bar{v}).$$

$$= \sum_{l=l_{\min}}^{l_{\max}} \frac{pL(l)}{S_k^2}.\ Var(v-\bar{v}) \quad . (11)$$

Since the variance of $(v-\bar{v})$ does not depend on the slopes , it does not affect the location of the global minimum of $\varepsilon(S_k)$. The constrained optimization problem defined in (6) can now be re-written as follows:

$$\arg\min_{s1......sN} \sum_{k=1}^{N} \frac{p_k}{s_k^2}$$

Subject to $\sum_{k=1}^{N} s_k = \frac{v_{\max}}{\delta}$ (12)

Where $p_k = \sum_{l=l_k}^{l+1} p_L(l)$ and l_k and l_{k+1} define the lower and the upper bounds of a segment, respectively. This problem can be solved analytically by calculating the first order Karush-Kuhn-Tucker (KKT) optimality conditions of the corresponding Lagrangian, which results in the following system of equations:

$$\begin{cases} \frac{-2p_1}{s_1^3}+\lambda=0 \\ \frac{-2p_2}{s_2^3}+\lambda=0 \\ . \\ . \\ \frac{-2p_N}{s_N^3}+\lambda=0 \end{cases}$$

$$\sum_{k=1}^{N} s_K - \frac{v_{\max}}{\delta} = 0 \quad (13)$$

where λ is the Lagrange multiplier. The solution to the above system of equations results in the slopes s_k given by

$$s_k = \frac{v_{\max} \cdot p_K^{1/3}}{\delta.\sum_{k=1}^{N} p_k^{1/3}}. \quad (14)$$

Note that the expression derived in (14) does not consider the upper bound constraint imposed on s_k in (7). Let be the set of the index of a segment with a slope that exceeds the upper bound. We overcome the upper bound violation using the following adjustment:

$$S_k = \left\{ \frac{\left(v_{\max -} \sum_{i\in X} \delta_{\max^{\delta}}\right)}{\delta.\sum_{j\notin X}^{N} p_j^{1/3}} .p_k^{1/3} \right. \quad \text{for } S_k \in I,$$

(15)

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section we first validate the proposed methods: optimization using the statistical model proposed in Section III-B and the closed-form solution based on a simplified model derived in Section III-D. Then, our models are further analyzed based on the generated tone curve and the distortion of the reconstructed HDR content. The performance of our models is also evaluated by comparing it with existing tone-mapping methods

.

A. Model Validation

In this section, we validate that the statistical model of Section III-B results in a tone curve that truly reflects the ground-truth results. Ground-truth results are achieved using the ideal scheme illustrated in Fig. 2(a), where the actual H.264/AVC encoder and decoder are employed to find the truly optimal piecewise linear tone curve

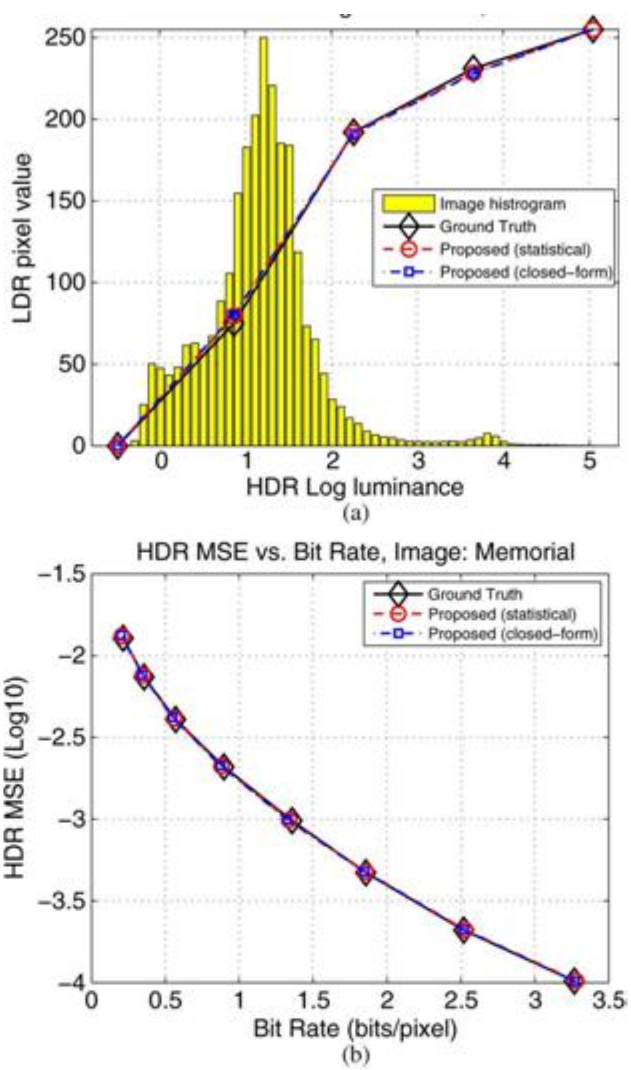

Figure 4: Validation of the proposed models by comparison with the ground-truth solution. The top figure, (a), shows the tone curves computed using the statistical model, the closed-form solution and the ground-truth optimization for the image "Memorial". The x axis denotes the HDR luminance in the log-10 scale, and y axis is the LDR pixel value. (b) demonstrates the result of HDR MSE (in log10 scale) versus bit rate (bits/pixel). The lower theMSE value, the better the image quality

.

B. Dependence of the Tone Curves on QP

Next, we verify that the proposed statistical model can be well approximated by the closed-form solution which produces a tone curve that is independent of QP.

C. Further Analysis of the Closed-Form Solution

The tone curve resulting from the closed-form solution given by (14) can be generalized as follows:

$$s_k = \frac{v_{\max} \cdot p_k^{1/t}}{\delta . \sum_{k=1}^{N} p_k^{1/t}} \qquad (16)$$

In our closed-form solution, is set to be equal to 3. Note that when t=1, (16) is identical to the histogram equalization operation.

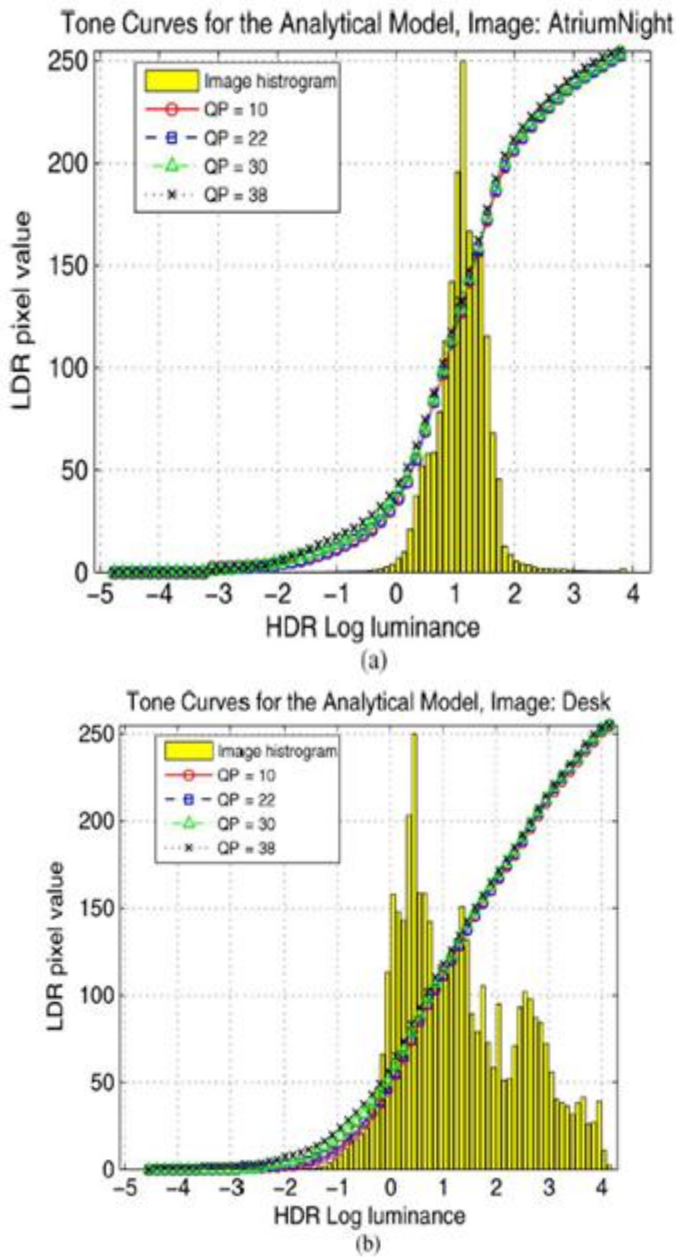

Figure 5: Tone curves generated using the statistical model with different QP values for the images “AtriumNight” and “Desk”. The notation of the axis is the same as Fig. 4(a). The smaller the value of QP, the better the compression quality 87 and 88 segments are used for “AtriumNight” and “Desk” respectively.

Photographic TMO

Adaptive logarithmic TMO

Display Adaptive TMO

Figure 6: Rate-distortion curves, tone curves and tone-mapped images for the image “AtriumNight”. The first row demonstrates the resulting tone-curves with different TMOs, followed by the results for MSE and SSIM versus bit rates; the second row shows tone-mapped LDR images using the proposed statistical model and the closed-form solution. The third row shows the tone-mapped images using the existing tone-mapping methods. All the tone-mapped images shown are compressed. The compression quantization parameters used for “AtriumNight” is 10. The number of segments used for the histogram is 87.

Figs. 6, 7 and 8 display the tone curve, rate-distortion curves and tone-mapped LDR images for three images. Additional results for more images are included in the supplementary material. The LDR images shown in these figures demonstrate that the images tone-mapped using our method also provide good quality. To further demonstrate the quality of the LDR images generated by the proposed models, Fig. 11 shows the distortion maps of the LDR images compared with their original HDR counterparts.

Photographic TMO

Adaptive logarithmic TMO

Display Adaptive TMO

Image: Coby; LDR images shown are compressed with QP = 10

Fig.7. Rate-distortion curves, tone curves and tone-mapped images for the image “Coby”. The notation is the same as Fig.6. The compression quantization parameters used for “Coby” is 22. The number of segments used for the histogram is 36.

Statistical (proposed)

Close-form (proposed)

Photographic

Logarithmic

Adaptive Display

Figure 8. Distortion maps of the LDR images relative to the original HDR images. The LDR images evaluated have not been compressed. In each of the distortion maps, three colors denote three different types of distortions: green for loss of visible contrast; blue for amplification of invisible contrast; red for reversal of visible contrast. The higher intensity of a color correlates with higher distortion of that type.

V. CONCLUSION

In this paper, we showed that the appropriate choice of a tone-mapping operator (TMO) can significantly improve the reconstructed HDR quality. We developed a statistical model that approximates the distortion resulting from the combined processes of tone-mapping and compression. Using this model, we formulated a constrained optimization problem that finds the tone-curve which minimizes the expected HDR MSE. The resulting optimization problem, however, suffers from high computational complexity. Therefore, we presented a few simplifying assumptions that allowed us to reduce the optimization problem to an analytically tractable form with a closed-form solution. The closed-form solution is computationally efficient and has a performance compatible to our developed statistical model. Moreover, the closed-form solution does not require the knowledge of QP, which makes it suitable for cases where the compression strength is unknown. Although our models are designed to minimize HDR MSE, the extensive performance evaluations show that the proposed methods provide excellent performance in terms of SSIM and the LDR image quality, in addition to an outstanding performance in MSE.

REFERENCES

[1] H. Seetzen, W. Heidrich, W. Stuerzlinger,G.Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, “High dynamic range display systems,” *ACMTrans. Graph. (Proc. SIGGRAPH)*, vol. 23, no. 3, pp. 757–765, 2004.

[2] IEC61966-2-4, Colour Measurement and Management—Part 2–4: Colour Management—Extended-Gamut YCC Colour Space for Video Applications— xvYCC 2006.

[3] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, “Perception-motivated high dynamic range video encoding,” *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 23, no. 3, pp. 730–738, 2004.

[4] G. Ward, “Real pixels,” *Graphics Gems II*, pp. 80–83, 1991.

[5] R. Bogart, F. Kainz, and D. Hess, “Openexr image file format,” in *Proc. ACM SIGGRAPH 2003, Sketches Applicat.*, 2003.

[6] G. W. Larson, “Logluv encoding for full-gamut, high-dynamic range images,” *J. Graph. Tools*, vol. 3, no. 1, pp. 15–31, 1998.

[7] Y. Gao and Y.Wu, “Applications and Requirement for Color Bit Depth Scalability,” Oct. 2006, Tech. Rep., ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-U049.

AUTHORS

S. Aravind Kumar,
M.E CSE, Anna University of technology Chennai, Anand Institute of Higher Technology, Kazhipattur, Tamil Nadu, India.
saravind123@gmail.com