SCALABLE CODING OF HIGH DYNAMIC RANGE VIDEO

Andrew Segall

Sharp Labs of America, 5750 NW Pacific Rim Blvd, Camas, WA 98607
asegall@sharplabs.com

ABSTRACT

A method for coding high dynamic range video sequences is considered. The technique is scalable, in that it facilitates the simultaneous transmission of standard and high dynamic range versions of the sequence in a single bit-stream. Furthermore, the approach is backwards compatible with the existing, state-of-the-art, AVC|H.264 video coding standard. Emphasis is placed on improved coding efficiency as well as managed computational complexity. Results illustrate the efficacy of the approach.

Index Terms – Video coding, high dynamic range, HDR, scalable

1. INTRODUCTION

The dynamic range of current generation displays is significantly less than the capabilities of a human observer. Modern displays typically render images with two to three orders of luminance magnitude and a maximum brightness of 500 cd/m2. By comparison, the overall luminance range of a human observer is approximately 14 orders of magnitude, varying from faint starlight (10^{-6} cd/m2) to bright sunlight (10^{8} cd/m2). The luminance range at any one time instant is approximately five orders of magnitude, which describes a contrast ratio of 100,000:1

Next generation display devices are narrowing the difference between the luminance range of the display and the luminance range of the human observer. For example, [1] reports a device with a maximum brightness of 2,700 cd/m2 and a dynamic range of 54,000:1. This is achieved by replacing the backlight of an LCD display with a DLP projector. Additionally, [1,2] report an alternative device that utilizes a grid of LEDs for the LCD backlight. The result has a dynamic range greater than 280,000:1 and a maximum brightness of 8,500 cd/m2.

This paper considers the problem of coding a high dynamic range video sequence for future display technology. A scalable scenario is the focus, as enlarging the dynamic range of a display is assumed to be an evolutionary process. In the next section, I provide an overview of the system. Section III better explains the individual coding tools. Finally, Section IV evaluates the performance of the proposed system.

2. SYSTEM OVERVIEW

A block diagram of the system appears in Figure 1. As can be seen from the Figure, the system employs a layered approach for scalability. The process begins by converting an original, high dynamic range image sequence into a lower dynamic range sequence that is suitable for legacy displays. This conversion, or mapping, process is commonly referred to as *tone mapping* [3], and its goal is to provide a representation of the larger dynamic range material in a smaller dynamic range image format. I stress that this process is not equivalent to discarding the least significant bits of the high dynamic range data.

The standard dynamic range sequence is then coded in a manner compliant with the AVC|H.264 video coding standard [4,5]. (Alternative coding systems are also viable.) The system first divides the image sequence into blocks. Then, each block is represented with a prediction mode and residual data. Prediction modes include intra-prediction methods that exploit spatial correlations with previously reconstructed blocks, as well as inter-prediction methods that exploit temporal correlations between the current block and previously reconstructed frames. The residual difference between the prediction and original frame is then transformed and quantized. These prediction modes and

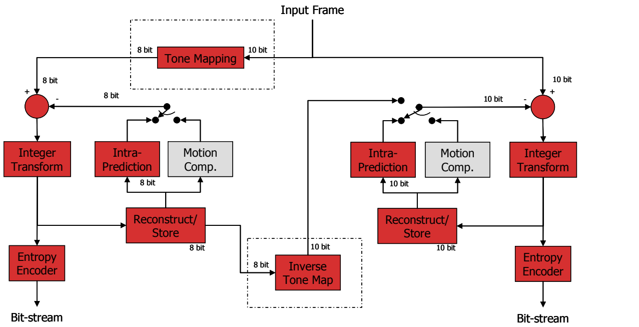

Figure 1 Block diagram of the system considered in the paper. A high dynamic range signal is converted to a lower dynamic range utilizing a tone mapping operation. The resulting image sequence is then coded with an H.264/AVC compliant codec. The high dynamic range sequence is also codec, and the decoded lower dynamic range sequence provides an additional, inter-layer prediction type for improving the coding efficiency. Bit-depths are provided for illustration only.

1-4244-1437-7/07/$20.00 ©2007 IEEE

quantized coefficients are subsequently transmitted to the decoder, where the process is reversed.

Coding of the high dynamic range image sequence depends on the standard dynamic range data. Here, the concept of *inter-layer prediction* is utilized to project information from the standard dynamic range data to the high dynamic range domain. Inter-layer prediction is further discussed in the next section. However, as shown in the Figure, the prediction tool provides an additional prediction method for coding. Thus, a block may be predicted using intra-frame, inter-frame or the additional inter-layer tools. As in the AVC|H.264 video coding standard, the prediction type is signaled to the decoder on a macro-block basis. Furthermore, the residual difference is transformed and quantized.

3. CODING TOOLS

A. Inter-layer Prediction

Inter-layer prediction is realized using a gain plus offset operation. This has the major advantage of being computational simple. Moreover, it allows simple control by an encoder for adapting the inter-layer prediction process to the decoded, lower dynamic range content. The gain and scaling parameters are transmitted in the enhancement bit-stream and provide control of the inter-layer prediction mechanism. The system transmits the parameters on a block-by-block basis, and efficiently encodes the data with a context adaptive entropy codec.

While any number of gain and offset realizations might be reasonable, here I consider the following operation.

$$HDR(x,y)=\sum_{\forall i} a_i * LDR(x,y) << i + Offset(x,y), \tag{1}$$

where *HDR* and *LDR* are, respectively, the high dynamic range and low dynamic range version of the image sequence, *x* and *y* denote the spatial location within the image frame, a_i is a binary indicator that belongs to the set {-1,0,1}, and *i*={0,1,2,3}, and *Offset*(*x*,*y*) is an offset value. Transmission of Offset(x,y) can be a component of the inter-layer prediction mode information; alternatively, it is equivalent to the DC value of the residual data.

The introduction of an offset parameter allows for the standard dynamic range content to be shifted within the luma space of the high dynamic range sequence. Unfortunately, this shifting is complicated by the fact that most color spaces utilized for video coding are not iso-luminant. For example, a video codec typically transmits data in the YCbCr color space with code word mappings defined in ITU-R BT.709 [6]. This requires either a color transform in the decoding loop or additional offset parameters in the bit-stream.

Incorporating a color transform into the inter-layer prediction step is undesirable due to the additional complexity overhead. Instead, I further modify the inter-prediction process to account for the luma and chroma relationship. To better understand this, let us start by considering performing inter-layer prediction following the color transform

$$\begin{aligned} Y_{LDR} &= Y_{LDR} \\ b &= \frac{Cb_{LDR}}{Y_{LDR}+Cr_{LDR}+Cb_{LDR}}, \\ y &= \frac{Y_{LDR}}{Y_{LDR}+Cr_{LDR}+Cb_{LDR}} \end{aligned} \tag{2}$$

where Y_{LDR} denotes the luma code words in the low dynamic range image, Cb_{LDR} and Cr_{LDR} denote the chroma components of the low dynamic range image, and *b* and *y* are color difference channels as defined. Please note that this is similar in spirit to an *xy*Y representation. The important difference is the use of non-linear luma and chroma codewords instead of linear luminance and chrominance values.

Inter-layer prediction in this alternative color space is then performed on the Y_{LDR} component, so that

$$Y_{HDR} = \alpha Y_{LDR} + Offset,$$

with α and *Offset* representing the general scale and offset procedure. Following that operation, the inverse color transform would be applied and the predicted data recovered as

$$\begin{aligned} Y_{HDR} &= Y_{HDR} \\ Cb_{HDR} &= \frac{bY_{HDR}}{y}, \\ Cr_{HDR} &= \frac{(1-b-y)Y_{HDR}}{y} \end{aligned}$$

where Cb_{HDR} and Cr_{HDR} denote the chroma components of the high dynamic range image.

I have previously mentioned that the above color transformation step could be utilized for inter-layer prediction, and that this is undesirable due to the additional complexity within the decoder. Additionally, I would like to draw addition to another problem: the Y, Cb and Cr components are not always represented at the same resolution or located at the same location on the sample grid. Specifically, many video coding scenarios employ the

4:2:0 or 4:2:2 color samplings that down-sample the chroma components relative to the luma information. In most cases, the position of the chroma sample values is not aligned with the luma sample position. (This is true in the H.264|AVC standard for example.)

Simplifying Eqs (2)-(4) suggests an alternative method for addressing chroma within the inter-layer prediction stage and addressing the above difficulties. Specifically, the components Cb*HDR* and Cr*HDR* can be expressed as

$$\begin{aligned} Cb_{HDR} &= \alpha Cb_{LDR} + Offset \cdot \frac{Cb_{LDR}}{Y_{LDR}} \\ Cr_{HDR} &= \alpha Cr_{LDR} + Offset \cdot \frac{Cr_{LDR}}{Y_{LDR}} \end{aligned} , \qquad (5)$$

where *Offset* is the offset parameter transmitted in the bit-stream and applied to the luma component. As can be seen from the Eq. (5), the offset parameter for each chroma component is a function of the luma Offset and the luma and chroma pixel values.

The expression in Eq. (5) is still plagued by the practical problems of different resolutions for luma and chroma sample grids. Addressing this problem can be accomplished by upsampling or down-sampling the signal components to alternative grids. In the system discussed here though, the expectation operator is utilized to calculate the chroma offset parameter for each block. This is motivated by the fact that the luma Offset parameter also varies on a block by block basis. Additionally, it can be implemented with low computational cost, as the DC value is available in the baselayer bit-stream.

The inter-layer prediction process for estimating the higher dynamic range signal therefore becomes

$$\begin{aligned} Y_{HDR} &= \alpha Y_{LDR} + Offset \\ Cb_{HDR} &= \alpha Cb_{LDR} + Offset \cdot \frac{Cb_{LDR,DC}}{Y_{LDR,DC}} , \qquad (6) \\ Cr_{HDR} &= \alpha Cr_{LDR} + Offset \cdot \frac{Cr_{LDR,DC}}{Y_{LDR,DC}} \end{aligned}$$

where $Y_{LDR,DC}$, $Cb_{LDR,DC}$ and $Cr_{LDR,DC}$ are the DC portion of the luma and chroma components in the LDR image block, respectively. Please note that this DC value is readily extracted from the transmitted bit-stream.

B. Residual Coding

Residual coding describes the task of sending the difference between the original high dynamic range data and the lower dynamic range sequence. This refines the estimate for the high dynamic range sequence within the decoder and provides a mechanism for transmitting portions of the high dynamic range sequence that are completely discarded by the tone mapping operation. Coding the residual information utilizes the traditional residual coding techniques of the H.264/AVC standard. Namely, the residual difference is transformed with either an integer 4x4 or 8x8 transform.

C. Motion Compensation

To further reduce the complexity of the decoding operation, a single loop decoding structure is employed. This is similar to the method standardized in the scalable extensions to H.264/AVC [7,8]. The single loop design reconstructs a high dynamic range image sequence without requiring motion compensation in the lower dynamic range sequence. Inter-layer prediction is performed as before, but it is applied to the residual difference signal only. In addition to this prediction tool, motion vectors and reference frame indices are also projected from the lower dynamic range sequence to the higher dynamic range.

4. SIMULATIONS

To measure the performance of the proposed system, high dynamic range material is encoded with both a two-layer and single layer video codec. The input material is prepared as follows. First, the original high dynamic range video sequence is converted to the xyY color space and tone mapped to the high dynamic range output device. The tone mapping operation consists of a clipping process, and the resulting image is subsequently returned to the RGB color space and encoded with the BT.709 codeword construction process. The resulting image is quantized to 10-bits.

Generation of the lower dynamic range sequence is derived from the initial, tone mapped image. The process begins with the image frames generated for the higher dynamic range output device. The data is then mapped to the lower dynamic range device by applying a sigmoid operator to the luminance values. For these experiments, the process is defined as

$$Y_{LDR} = \frac{2}{1 + \exp\left(-Y_{HDR}/T\right)} - 1 , \qquad (7)$$

where *T* is a threshold equal to the mean of Y_{HDR}. This is a simple but reasonable tone mapping procedure, and a visual example of the process appears in Figure 2.

The proposed system processes the resulting low dynamic range and high dynamic range images. In order to provide reproducible results, images from a public high dynamic range database are utilized [9]. (Please note that this results in intra-only coding.) Results are reported in Table 1 with a Qp value of 26 for the baselayer. As can be seen, the proposed system provides a significant performance improvement compared to simulcast. An average bit-rate reduction of greater than 28% is measured, with maximum and minimum values of 49% and 10%, respectively. Moreover, the proposed system performs within 12% of the single layer solution, on average. This illustrates the potential of the system, as high dynamic range image sequences can be transmitted with a backwards compatible, lower dynamic range image sequence with only a 12% increase in bit-rate.

REFERENCES

[1] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, A. Vorozcovs, “High Dynamic Range Display Systems”, *ACM Transactions on Graphics (Special Issue: Proceedings of SIGGRAPH’04)*, vol.23, no.3., August 2004.

[2] H. Seetzen, L. Whitehead and G. Ward, “High Dynamic Range Display Using Low and High Resolution Modulators," *The Society for Information Display International Symposium*, Baltimore, MD, May 2003.

[3] E. Reinhard, G. Ward, S. Pattanaik and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*, Elesevier, 2006.

[4] ITU-T Recommendation H.264 | ISO/IEC 14496-10, *Advanced Video Coding for Generic Audiovisual Services*, March 2005.

[5] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the H.264/AVC Video Coding Standard”, *IEEE Transactions and Circuits for Video Technology*, vol.13, no.7, pp.560-576, July 2003.

[6] International Telecommunication Union, “Parameter Values for the HDTV standards for production and international programme exchange”, ITU-R BT.709-5, April, 2002.

[7] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz and M. Wien, “Joint Draft 7 of SVC amendment (revision 2)”, JVT-T201, Klagenfurt, Austria, July 15-21, 2006.

[8] H. Schwarz, D. Marpe and T. Wiegand, “Overview of the Scalable Video Coding Standard”, *IEEE Transactions and Circuits for Video Technology*, to appear.

[9] G. Ward, “High Dynamic Range Image Examples,” Dec. 14, 2003; http://www.anywhere.com/gward/hdrenc/pages/originals.html.

Table I

Sequence	Single Layer (HDR Only)	Simulcast (HDR&LDR)	Proposed (HDR&LDR)
Atrium Night	2242.8	4435.92	2271
BigFogMap	3910.32	7476.48	4084.68
Church1	8563.92	13143.12	9876
Dani Synagogue	3914.88	6165.72	4381.32
Desk	10639.44	16494.24	14828.04
Display1000	4476.6	6029.88	4413.12
Memorial	5142.28	9449.76	5595.72
Mt. Tam West	3425.16	4537.68	3441.24
Spheron Price Western	5083.44	8049	6313.44
Still Life	3460.68	5365.2	3790.32
Sunset1	5098.68	6800.4	5378.64
Tree	8918.88	14602.92	11161.44

Performance of the proposed system. Twelve publicly available video sequences are encoded. Single layer results are calculated by encoding the high dynamic range sequence only, while simulcast results correspond to encoding the low and high dynamic range sequences independently. The proposed, scalable method is then considered. Results are reported in kbps and show that the proposed system provides an average bit-rate reduction of 28% compared to simulcast. Moreover, the scalable solution requires an average bit-rate increase of 12% compared to single layer coding.

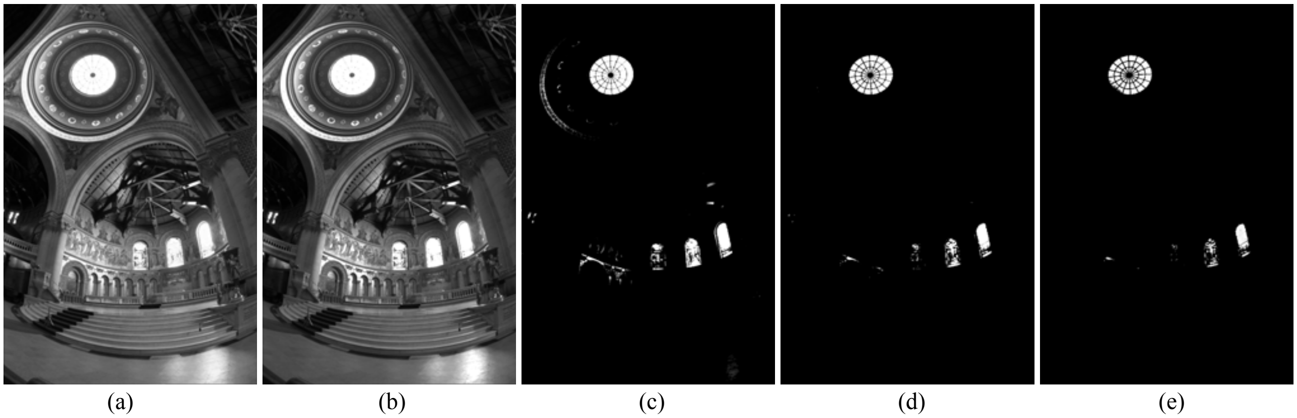

Figure 2 Illustrative example of the input to the scalable video codec (a) low dynamic range image, (b)-(e) high dynamic range image. The high dynamic range image can not be printed or viewed with a conventional monitor, so the image is divided into intensity ranges for illustration here. Images correspond to the intensity range: (b) [0,255]; (c) [256, 511]; (d) [512,767], and (e) [768,1023]. Note that the low dynamic range and lowest intensity range of the high dynamic range sequence are similar but not identical. The degree of similarity is content dependent in practice.