HIGH DYNAMIC RANGE IMAGE AND VIDEO COMPRESSION - FIDELITY MATCHING HUMAN VISUAL PERFORMANCE

Rafał Mantiuk, Grzegorz Krawczyk, Karol Myszkowski, Hans-Peter Seidel

MPI Informatik, Saarbrücken, Germany

ABSTRACT

Vast majority of digital images and video material stored today can capture only a fraction of visual information visible to the human eye and does not offer sufficient quality to fully exploit capabilities of new display devices. High dynamic range (HDR) image and video formats encode the full visible range of luminance and color gamut, thus offering ultimate fidelity, limited only by the capabilities of the human eye and not by any existing technology. In this paper we demonstrate how existing image and video compression standards can be extended to encode HDR content efficiently. This is achieved by a custom color space for encoding HDR pixel values that is derived from the visual performance data. We also demonstrate how HDR image and video compression can be designed so that it is backward compatible with existing formats.

***Index Terms*—** HDR, high dynamic range, video coding, MPEG, color space, backward-compatible coding, scene-referred, output-referred

1. INTRODUCTION

Only a fraction of visual information visible to the human eye is stored using common image and video compression formats, such as MPEG or JPEG. Since these formats have been originally designed for CRT displays, the quality they offer is often not sufficient to fully exploit capabilities of new display devices. While the latest displays offer contrast of 10,000:1 (dynamic contrast, in production, 2006) or even 50,000:1 (single image contrast, prototypes, 2006) and brightness exceeding 500 cd/m^2 or 3,000 cd/m^2 for prototypes, the digital image and video content is still in the era of CRT displays with the contrast of 1:200 and the maximum brightness of 100 cd/m^2. The 8-bit coding of luminance (luma) values cannot encode the full range of luminance that can be found in the real world and seen by the human eye. This results in image and video with clipped specular highlight, missing bright light sources and flattened shadows. The latest wide-color-gamut displays also cannot take full advantage of their capabilities as the color spaces used for video and image coding store only a portion of the visible color gamut. To fully benefit from new display technologies and to improve compatibility of digital content between different devices, video and image formats need to be extended to encode higher dynamic range and wider color gamut.

High dynamic range (HDR) image and video formats encode the full visible range of luminance and color gamut, thus offering ultimate fidelity, limited only by the capabilities of the human eye and not by any existing technology. In this paper we outline the major benefits of HDR representation and show how it differs from traditional encodings. In Section 3 we demonstrate how existing image and video compression standards can be extended to encode HDR content efficiently. This is achieved by a custom color space for encoding HDR pixel values that is derived from the visual performance data. In Section 4 we show how HDR image and video compression can be designed so that it is backward compatible with the existing formats.

2. HDR IMAGE AND VIDEO COMPRESSION

In the following paragraphs we will explain how HDR image and video representation differs from standard representations found in the popular JPEG and MPEG formats.

Bit-depth and dynamic range. HDR images and video can encode higher contrast (*dynamic range*). However, higher *bit-depth* encoding does not necessary imply higher contrast. Although numerous standard image and video formats offer extended bit-depth encodings, such as 14-bit luma coding in Fidelity Range Extensions (FRExt), part of ISO/IEC 14496-10 Advanced Video Coding [1], these are intended to offer higher precision (fidelity), but not necessary to encode higher contrast. Therefore FRExt MPEG file can encode subtle luminance differences, mostly invisible to the human eye and below the camera noise level, but due to the design of its transfer functions and color spaces, the same high-end format cannot encode many high-contrast real-world scenes containing very bright as well as shaded parts. In contrast, HDR image and video formats are designed to encode all luminance levels that can be found in the real world, which may vary from as low as 10^{-5} cd/m^2 which is roughly the luminance of a moonless sky, to 10^{10} cd/m^2, which exceeds the luminance of the sun. HDR formats can usually store contrast $1 : 10^{15}$ or dynamic range of 15 $\log_{10}$ units, while standard formats usually do not exceed the dynamic range of 3 $\log_{10}$ units.

Output-referred and scene-referred encoding. Majority of image and video formats are *output-referred*, which means that their content describes what should be displayed on an output-device rather than what was captured in the actual scene. An example of output-referred approach is the sRGB color space, a de-facto standard for digital images, which was designed to approximate the performance of a typical CRT display. While an image encoded using the sRGB color space is well specified in terms of a 100-cd/m^2 display (maximum luminance of a typical CRT display), it is not clear how the same image can be displayed on a 1000-cd/m^2 display. Output-referred formats suffer from compatibility issues and require solving difficult gamut-mapping problems when the same image is to be presented on devices of significantly different color gamut and maximum contrast. HDR formats are *scene-referred*, which means that they encode colors of the original scene, rather than their renderings on any particular display. Therefore, HDR image always delivers enough information for a display to render it optimally. A display device is responsible for applying appropriate rendering algorithm (so called tone-mapping) that would limit image contrast and color gamut to its capabilities.

Code values and colorimetry. Standard images are represented using 8, 10 or 12-bit integer numbers, so called *code values*. A triple of code values non-linearly relates to the actual colorimetric values of *luminance* and *chrominance*. This relation, although crucial for the final appearance of images, is often neglected when designing and optimizing compression algorithms. HDR images are originally represented as a triple of floating point numbers that linearly relate to luminance and chrominance. Such precise colorimetric description gives better opportunities to facilitate visual perception modeling and optimize imaging algorithms considering the limitations of the human visual system.

Lightness and brightness. Standard images and video operate in the domain of *lightness*, which is defined relative to a reference white (usually the highest code value), while HDR images better describe brightness, since they encode absolute luminance values. Lightness is well defined for paper prints, where the reference white is the brightest achievable color. Reference white for a display in a dim room is difficult to determine and depends on a scene. This is because the brightest colors on the display can be self-luminous (highlights, light sources), thus brighter than the reference white. Therefore, the assumption that the largest code value (e.g. a triple of 255) represents white is false for most modern displays. HDR formats are better suited for modern displays, where the knowledge about desired brightness reproduction is more important than the knowledge about image lightness.

Having outlined the major differences between standard and high dynamic range image and video representations, we will proceed with the description of a color space for efficient encoding of HDR pixels.

3. COLOR SPACE FOR HDR PIXELS

Although the most natural representation of HDR images is a triple of floating point numbers, floating point representation of pixel values does not lead to the best image or video compression ratios and adds complexity to compression algorithms. Moreover, since the existing image and video formats, such as MPEG-4 or JPEG2000, can encode only integer numbers, HDR pixels must be represented as integers in order to be encoded using these formats. Therefore, it is highly desirable to convert HDR pixels from a triple of 32-bit floating point values, to integer numbers. Such integer encoding of luminance should take into account the limitations of the human visual perception and the fact that the eye can see only limited numbers of luminance levels and colors. This section gives an overview of the color space that can efficiently represent HDR pixel values using only integer numbers and the minimal number of bits. More information on this color space can be found in [2].

To offer the best trade-off between compression efficiency and visual quality without imposing any assumptions on the display technology, we propose that the color space used for compression meets the following requirements:

1. The color space can encode the full color gamut and the full range of luminance that is visible to the human eye. This way the human eye, instead of the current imaging technology, defines the limits of such encoding.

2. A unit distance in the color space correlates with the Just Noticeable Difference (JND). This offers a nearly uniform distribution of distortions across an image and simplifies control over distortions for lossy compression algorithms.

3. Only positive integer values are used to encode luminance and color. Integer representation simplifies and improves image and video compression.

4. A half-unit distance in the color space is below 1 JND. If this condition is met, the quantization errors due to rounding to integer numbers are not visible.

5. The correlation between color channels should be minimal. If color channels are correlated, the same information is encoded twice, which worsens the compression performance.

6. There is a direct relation between the encoded integer values and the photometrically calibrated XYZ color values.

There are several color spaces that already meet some of the above requirements, but there is no color space that accommodates them all. For example, the Euclidean distance in the $CIE\ L^*u^*v^*$ color space correlates with the JND (Requirement 2), but this color space does not generalize to the full range of visible luminance levels, ranging from scotopic light levels, to very bright photopic conditions. Several perceptually uniform quantization strategies have been proposed [3, 4], including the grayscale standard display function from the DICOM standard [5]. However, none of these take into account a broad dynamic range and diversified luminance conditions imposed by Requirement 1.

To meet the requirements from 1 to 4, we derive a transfer function for luminance encoding from the contrast detection models, such as threshold versus intensity functions or contrast sensitivity functions (CSF). The transfer function converts physical luminance given in cd/m^2 into luma, a 12-bit integer code value, suitable for image and video compression. Since detailed derivation can be found in [2], we include below only the final formula for luminance to luma conversion (Requirement 6):

$$l(y) = \begin{cases} a \cdot y & \text{if } y < y_l \\ b \cdot y^c + d & \text{if } y_l \leq y < y_h \\ e \cdot \log(y) + f & \text{if } y \geq y_h \end{cases} \quad (1)$$

and for inverse conversion, from luma to luminance:

$$y(l) = \begin{cases} a' \cdot l & \text{if } l < l_l \\ b'(l + d')^{c'} & \text{if } l_l \leq l < l_h \\ e' \cdot \exp(f' \cdot l) & \text{if } l \geq l_h \end{cases} \quad (2)$$

The constants are given in the table below:

$a = 17.554$	$e = 209.16$	$a' = 0.056968$	$e' = 32.994$
$b = 826.81$	$f = -731.28$	$b' = 7.3014e-30$	$f' = 0.0047811$
$c = 0.10013$	$y_l = 5.6046$	$c' = 9.9872$	$l_l = 98.381$
$d = -884.17$	$y_h = 10469$	$d' = 884.17$	$l_h = 1204.7$

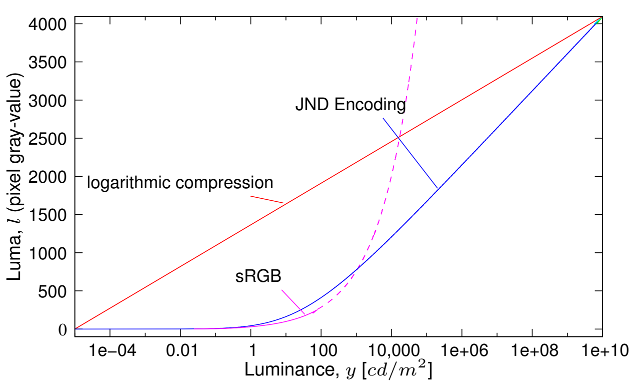

Fig. 1. Functions mapping physical luminance y to encoded luma values l. JND Encoding – perceptual encoding of luminance; sRGB – nonlinearity (gamma correction) used for the sRGB color space; logarithmic compression – logarithm of luminance, rescaled to 12-bit integer range.

Function $l(y)$ (Equation 1) is plotted in Figure 1 and labelled **JND encoding**. Note that both the formula and the shape of the JND encoding is very similar to the nonlinearity (gamma correction) used in the sRGB color space. Both JND encoding and sRGB nonlinearity follow similar curve on the plot, but the JND encoding is more conservative (a steeper curve means that a luminance range is projected on a larger number of discrete luma values thus lowering quantization errors). However, the sRGB non-linearity results in a too steep function for luminance above 100 cd/m^2, which requires too many bits to encode higher luminance values. Therefore the proposed transfer function can be regarded as an extension of gamma correction to HDR pixel values. The derived formulas guarantee that the same difference of values l, regardless whether in bright or in dark region, approximately corresponds to the same visible difference. Neither luminance nor the logarithm of luminance has this property, since the response of the human visual system to luminance is complex and non-linear. The values of l lay in the range from 0 to 4095 (12 bit integer) for the corresponding luminance values from 10^{-5} to 10^{10} cd/m^2, which is the range of luminance that the human eye can effectively see (although the values above 10^6 would mostly be useful for representing the luminance of bright light sources).

Color is encoded in the proposed color space for HDR pixels using CIE 1976 Uniform Chromacity Scales (UCS) $u'v'$, which are multiplied by 410 to rescale the values to 8-bit integer numbers. Unlike CIE u^*v^*, CIE $u'v'$ show almost no correlation with luminance (Requirement 5) and are approximately perceptually uniform (Requirement 2).

The color space described in this section can be directly used for many existing image and video compression formats, such as JPEG-2000 and MPEG-4. Both these formats can encode luminance with 12 or more bits, which make them fully capable of representing HDR pixel values. As a proof of concept we extended an MPEG-4 compression algorithm to use the proposed color space [6]. We demonstrate that HDR content can be played in real-time on an advanced HDR displays or on standard displays after applying tone mapping. We can also simulate perceptual effects that are normally only evoked when observing scenes of large contrast and luminance range, such as night vision, an optically accurate motion blur and visual glare. More information on the project can be found in [6] and [7].

4. BACKWARD-COMPATIBLE HDR VIDEO COMPRESSION

Since the standard low-dynamic range (LDR) file formats for images and video, such as JPEG or MPEG, have become widely adapted standards supported by almost all software and hardware equipment dealing with digital imaging, it cannot be expected that these formats will be immediately replaced with their HDR counterparts. To facilitate transition from the traditional to HDR imaging, there is a need for backward compatible HDR formats, that would be fully compatible with existing LDR formats and at the same time would support enhanced dynamic range and color gamut. Encoding movies in an HDR format is specially attractive for DVD distribution, where the original content, captured with high-end cinematographic cameras, offers extended dynamic range and wider color gamut. The proposed backward-compatible compression algorithm encodes in a single video bit-stream both output-referred content (after color-grading) intended for LDR displays, as well as the original high-quality content, suitable for a large range of display devices.

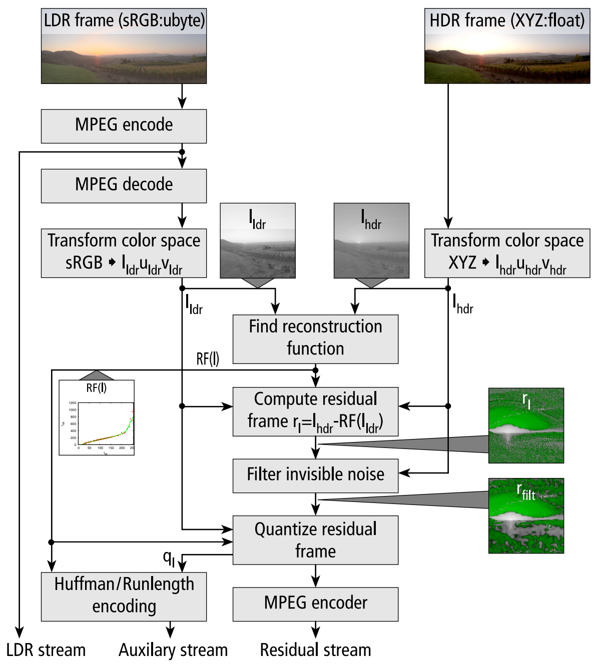

Fig. 2. A data flow of the backward compatible HDR MPEG encoding.

The complete data flow of the proposed backward compatible HDR video compression algorithm is shown in Figure 2. The encoder takes two sequences of HDR and LDR frames as input. The LDR frames, intended for LDR devices, usually contain a tone mapped or gamut mapped version of the HDR frames. The LDR frames are compressed using a standard MPEG encoder (see **MPEG encode** in Figure 2) to produce a backward compatible LDR stream. The LDR frames are then decoded to obtain a distorted (due to lossy compression) LDR sequence, which is later used as a reference for the HDR frames (**MPEG decode**). Both the LDR and HDR frames are then converted to compatible color spaces, which minimize differences between LDR and HDR colors. The reconstruction function (**Find reconstruction function**) reduces the correlation between LDR and HDR pixels by giving the best prediction of HDR pixels based on the values of LDR pixels. The residual frame is introduced to store a difference between the original HDR values and the values predicted by the reconstruction function. The reconstruction function is selected to make the values of the residual frame as small as possible. To improve compression, invisible luminance and chrominance variations are removed from the residual frame (**Filter invisible noise**). Such filtering simulates the visual processing that is performed by the retina to predict the contrast detection threshold at which the eye does not see any differences. The contrast magnitudes that are below this threshold are set to zero. Finally, the pixel values of a residual frame are quantized (**Quantize residual frame**) and compressed using a standard MPEG encoder into a residual stream. Both the reconstruction function and the quantization factors are compressed using lossless arithmetic encoding and stored in an auxiliary stream.

This section is intended to give only an overview of the compression algorithm. For details and results, the reader should refer to [8] and to the project web page [9].

5. CONCLUSIONS

We argue that image and video compression standards should encode the full range of luminance and color gamut visible to the human eye, including the colors that cannot be displayed on typical displays. Such encoding makes the content device independent and suitable for the future generations of display devices. To achieve good compression ratio, we derive a perceptually-motivated HDR color space, optimized for the human visual performance. Such color space can be used with existing image and video compression algorithms. To minimize compatibility issues while offering smooth transition from traditional to HDR content, we propose a backward compatible HDR video compression algorithm.

6. REFERENCES

[1] T. Wiegand, GJ Sullivan, G. Bjntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 13, no. 7, pp. 560–576, 2003.

[2] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Lossy compression of high dynamic range images and video,” in *Proc. of HVEI XI*. 2006, vol. 6057, p. 60570V, SPIE.

[3] M.I. Sezan, K.L. Yip, and S. Daly, “Uniform perceptual quantization: Applications to digital radiography,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 17, no. 4, pp. 622–634, 1987.

[4] J. Lubin and A.P. Pica, “A non-uniform quantizer matched to the human visual performance,” *Society of Information Display Int. Symposium Technical Digest of Papers*, , no. 22, pp. 619–622, 1991.

[5] DICOM PS 3-2004, “Part 14: Grayscale standard display function,” in *Digital Imaging and Communications in Medicine (DICOM)*. 2004.

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

[7] WWW:, “HDR MPEG-4,” `http://www.mpi-inf.mpg.de/resources/hdrvideo/index.html`.

[8] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, “Backward compatible high dynamic range mpeg video compression,” *ACM Trans. on Graphics*, vol. 25, no. 3, 2006.

[9] WWW:, “Backward-compatible HDR MPEG,” `http://www.mpii.mpg.de/resources/hdr/hdrmpeg/`.