Lossy Compression of High Dynamic Range Images and Video

Rafał Mantiuk, Karol Myszkowski, and Hans-Peter Seidel

MPI Informatik, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany;

ABSTRACT

Most common image and video formats have been designed to work with existing output devices, like LCD or CRT monitors. As display technology makes progress, these formats no longer represent the data that new devices can display. Therefore a shift towards higher precision image and video formats is imminent.

To overcome limitations of common image and video formats, such as JPEG, PNG or MPEG, we propose a novel color space, which can accommodate an extended dynamic range and guarantees the precision that is below the visibility threshold. The proposed color space, which is derived from contrast detection data, can represent the full range of luminance values and the complete color gamut that is visible to the human eye. We show that only minor changes are required to the existing encoding algorithms to accommodate the new color space and therefore greatly enhance information content of the visual data. We demonstrate this with two compression algorithms for High Dynamic Range (HDR) visual data: for static images and for video. We argue that the proposed HDR representation is a simple and universal way to encode visual data independent of the display or capture technology.

Keywords: color space, transfer function, luminance quantization, perceptually uniform color space, high dynamic range, HDR, image compression, video compression, contrast detection

1. INTRODUCTION

The recent advances in digital camera and display technologies make standard 8-bit per color channel representation of visual data insufficient. This is mostly due to the extended dynamic range of new capture and display devices: high dynamic range cameras can capture dynamic range over 150dB (compared to 65dB for a typical camera) and new HDR displays can show contrast ratio of 30,000:1 (compared to 400:1 for a typical LCD). Furthermore, these devices can cover much wider range of absolute luminance levels, ranging from 0.1 cd/m^2 to $3,000$ cd/m^2 for a HDR display. Since the typical color spaces, such as YCrCb , sRGB or L*u*v* cannot encode the full luminance range of HDR data, a new representation of the visual data that can accommodate the extended dynamic range is needed.

High dynamic range (HDR) imaging is a very attractive way of capturing real world appearance, since it assumes the preservation of complete and accurate luminance (or spectral radiance) values that can be found in a scene. Each pixel is represented as a triple of floating point values, which can range from 10^{-5} to 10^{10}. Such a huge range of values is dictated by both real world luminance levels and the capabilities of the human visual system (HVS), which can adapt to a broad range of luminance levels, ranging from scotopic (10^{-5} – 10 cd/m^2) to photopic ($10 - 10^8$ cd/m^2) conditions. Obviously, floating point representation results in huge memory and storage requirements and is impractical for storage and transmission of images and video. In this paper we propose an efficient encoding scheme for HDR pixels that preserves all the visual information visible to the human eye.

The paper is organized as follows: A brief overview of the existing image formats that can encode a higher dynamic range is given in Section 2. We outline differences between typical low-dynamic range (LDR) and HDR formats in Section 3. In Section 4 we derive a new color space that can efficiently encode HDR pixels. In Section 5 we describe two lossy HDR compression algorithms that utilize the proposed color space. We discuss similarities and differences between low dynamic range and HDR encoding in Section 6. Finally, we conclude the paper in Section 7.

Further author information: (Send correspondence to R.M.)
R.M.: E-mail: mantiuk@mpi-sb.mpg.de, Telephone: +49 681 9325-427

2. HDR IMAGE FORMATS

Most of the existing HDR image formats offer only lossless compression, which results in huge files sizes. This is one of the factors preventing those formats from gaining widespread acceptance and use. Another reason is lack of standards, which comes from little interest from the image and video format community in encoding higher dynamic ranges.

The existing formats that are capable of encoding higher dynamic range can be classified into three groups:

- Formats originally designed for high dynamic range images. The quantities they store are usually floating points values of a linear radiance or luminance factor*. There are several lossless formats, such as Radiance's RGBE,[1] logLuv TIFF[2] and OpenEXR,[3] and a lossy and backward compatible JPEG HDR format.[4] An excellent overview of these formats can be found in a recently published book.[5]
- Formats designed to store a higher dynamic range because of their application. This group includes: Digital Picture Exchange *DPX* format used in the movie industry to store scanned negatives, *DICOM* format for medical images, and a variety of so called *RAW* formats used in digital cameras. All these formats use more than 8 bits to store luminance, but they are not capable of storing such an extended dynamic range as the HDR formats.
- Formats that store larger number of bits but are not necessary intended for HDR images. Twelve or more bits can be stored in JPEG-2000, MPEG-4 (ISO/IEC 14496-2 or ISO/IEC 14496-10) and TIFF files. All these formats can easily encode HDR if they take advantage of a color space that can encode HDR, such as the one proposed in this paper.

In this paper we do not propose a new image or video format, but rather show how existing formats, that have gained widespread use, can be extended to store HDR data.

3. DEVICE- AND SCENE-REFERRED REPRESENTATION OF IMAGES

Commonly used image formats (JPEG, PNG, TIFF, etc.) contain data that is tailored to particular display devices: cameras, CRT or LCD monitors. For example, two JPEG images shown using two different LCD monitors may be significantly different due to dissimilar image processing, color filters, gamma correction etc. Obviously, such representation of images vaguely relates to the actual photometric properties of the scene it depicts, but it is dependent on a display device. Therefore those formats can be considered as device-referred (also known as output-referred), since they are tightly coupled with the capabilities and characteristic of a particular imaging device.

ICC color profiles can be used to convert visual data from one device-referred format to another. Such profiles define the colorimetric properties of a device for which the image is intended for. Problems arise if the two devices have different color gamuts or dynamic ranges, in which case a conversion from one format to another usually involves the loss of some visual information. The problem is even more difficult when an image captured with an HDR camera is converted to the color space of a low-dynamic range monitor (see a multitude of tone mapping algorithms). Obviously, the ICC profiles cannot be easily used to facilitate interchange of data between LDR and HDR devices.

Scene-referred representation of images offers a much simpler solution to this problem. As suggested by Greg Ward, the scene-referred image should encode the actual photometric characteristic of a scene it depicts.[5] Conversion from such common representation, which directly corresponds to physical luminance or spectral radiance values, to a format suitable for a particular device is the responsibility of that device. HDR file formats are examples of scene-referred encoding, as they usually represent either luminance or spectral radiance, rather than gamma corrected "pixel values".

*Luminance factor, $\tilde{y}$, is a relative measure of luminance, which is different from the actual luminance by a constant coefficient k ($y = k \cdot \tilde{y}$). The luminance factor is usually the result of multi-exposure techniques for HDR image capturing.

4. SCENE-REFERRED COLOR REPRESENTATION FOR IMAGE AND VIDEO ENCODING

Choice of the color space used for image or video compression has a great impact on the compression performance and capabilities of the encoding format. 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 has the following properties:

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 more 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 $L^*u^*v^*$ color space correlates with the JND (Property 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,[6,7] including the grayscale standard display function from the DICOM standard.[8] However, none of these take into account broad dynamic range and diversified luminance conditions as required by Property 1.

4.1. Luminance and Luma

We begin the derivation of the color space that incorporates all of the above listed properties with the luminance channel. Real-world physical luminance, given in cd/m^2, should be converted into integer numbers (Property 3 and 6), so that the error due to rounding to the nearest integer is not visible (Property 4). Additionally, it is desirable that the integer values representing luminance closely correspond to the sensory response of the HVS (Property 2). For example, intensity of sound is usually measured using non-linear decibel (dB) units since such a measure well corresponds to the perceived loudness of sound. We would like to find a similar measure of luminance for all possible light conditions. Our derivation is similar to other methods that model sensory output for a physical signal based on its threshold characteristic, such as transducer functions,[9,10] the grayscale standard display function,[8] or the capacity function in tone mapping.[11] Such luminance conversion is also called a *transfer function* in image compression literature.

Let us assume that the function $t(y_{adapt})$ gives a conservative estimate of the smallest difference of luminance that is visible to the human eye (the detection threshold) at a particular adaptation level, y_{adapt}. We are looking for a function $l \rightarrow y : y(l)$ that converts sensory units l (e.g. response of the photoreceptor), which we will call *luma*[†], into physical luminance y. Because the luma values l will be encoded as integer numbers

[†]Luma is a new word proposed by the NTSC in 1953 to prevent confusion between the Y' component of a color signal and the traditional meaning of luminance. While luminance is the weighted sum of the linear RGB components of a color video signal, proportional to intensity, luma is the weighted sum of the non-linear $R'G'B'$ components after gamma correction has been applied, and thus is not the same as either intensity or luminance. Source: http://www.wikipedia.org/

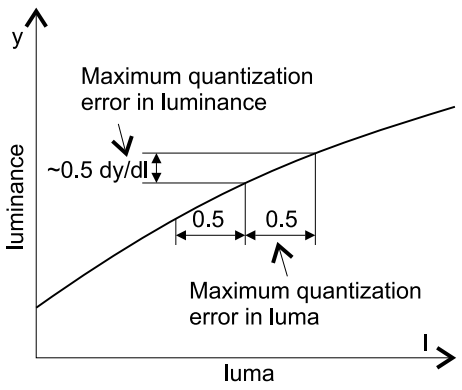

Figure 1. Maximum quantization error in sensory values, l, must be expressed in luminance, y, before it can be compared with the detection threshold, $t(y_{adapt})$.

(Property 3), we have to make sure that rounding to integers does not introduce visible distortions (Property 4). The maximum quantization error due to rounding of luma values, l, is ± 0.5. Since the detection thresholds are given in luminance, we have to convert this rounding error from luma, l, to luminance, y. This can be done by the Taylor series expansion of the function $y(l)$:

$$y(l+0.5) - y(l) \approx 0.5 \cdot \frac{dy}{dl} \tag{1}$$

This step is illustrated in Figure 1. We then make sure that the maximum rounding error is below or equal the detection threshold, $t(y_{adapt})$:

$$0.5 \cdot \frac{dy}{dl} < t(y_{adapt}) \tag{2}$$

To simplify our problem, we assume that the eye is adapted to the luminance of a single pixel, $y_{adapt} = y$. Although such an assumption is not true in real-world situations, it gives a conservative estimate of the detection threshold: the detection threshold is higher when the eye is not fully adapted. We can rewrite the above inequality as the following equality:

$$\frac{dy}{dl} = 2 \cdot \frac{t(y)}{k} \tag{3}$$

where k is a constant greater than 1. The larger the value of k, the more conservative the encoding (the lower is the detection threshold), but also the more bits are needed to encode l. An important consequence of rewriting Inequality 2 as Equality 3 is that the magnitude of a differential increase in l now directly relates to the sensory threshold, therefore the equation meets Property 2. The above equation can be solved in either of two ways:

- by solving a differential equation:

$$\frac{dy}{dl} = 2 \cdot \frac{t(y(l))}{k} \tag{4}$$

- or an integral:

$$\frac{dl}{dy} = 0.5 \cdot \frac{k}{t(y)} \;\Rightarrow\; l(y) = 0.5 \int \frac{k}{t(y)} dy \tag{5}$$

The solution of Equation 5 gives a function $y \rightarrow l : l(y)$, which converts physical luminance y into sensory units l, and the solution of Equation 4 gives the inverse function $l \rightarrow y : y(l)$. Note that y from Equation 3 has been replaced in Equation 4 with $y(l)$ to make the right side of the equation the function of l.

Finally, we must decide on the boundary conditions and find the value of the constant k. The boundary conditions will define the range of physical luminance that should be represented by the sensory units l. A reasonable range of luminance is within 10^{-5} cd/m^2 and 10^{10} cd/m^2, which can capture the luminance of both

a moonless sky ($3 \cdot 10^{-5}$ cd/m^2) and the surface of the sun ($2 \cdot 10^9$ cd/m^2). Therefore we can write the boundary conditions:

$$\begin{aligned} y(0) &= 10^{-5}\ cd/m^2 \\ y(l_{max}) &= 10^{10}\ cd/m^2 \end{aligned} \tag{6}$$

where l_{max} is the maximum value of l we want to encode and is usually equal $l_{max} = 2^{bits} - 1$. This gives us two point boundary problem, which can be solved using the *shooting method*[12‡]. The solution will give us the value of k. If the value of k is greater than 1, the sensory units, l, can represent luminance with sufficient precision, and that we have chosen an adequate number of bits.

So far we have not made any assumptions about the actual shape of the detection threshold function $t(y_{adapt})$. We can start with a simplistic case, where this function equals 1% of the Weber fraction[§], that is $t(y_{adapt}) = 0.01\ y_{adapt}$. This is a very imprecise, but unfortunately still commonly used assumption in computer vision and image compression, which is also referred as the Weber-Fechner law[¶]. We make further simplification and consider the case where $k = 1$, where the maximum quantization error is exactly equal $t(y_{adapt})$, rather than being greater than the threshold. From Equation 5 and our assumptions, we get:

$$l(y) = 0.5 \int \frac{1}{0.01y} dy = 50 \cdot \ln|y| + c \tag{7}$$

From the lower boundary condition (Equation 6), we have $c = -50 \cdot \ln|10^{-5}| = 575.65$. This way we derive a logarithmic compression function, which is commonly used for processing HDR images. Additionally, the derived function has the useful property that the unit difference corresponds to 1% contrast. We insert the upper boundary condition into Equation 7, we get $l(10^{10}) = 1726.9$, which means that we need at least 11 bits to represent the full visible range of luminance with a 1% step. Although such precision is usually regarded sufficient for video displayed on CRT displays, skilled observers are reported to notice contrast as low as 0.25%. Moreover, the contrast detection threshold is decreased with increased luminance of adaptation. Since new LCD and plasma displays are much brighter than their CRT counterparts, they eye is adapted to higher luminance levels when viewing such displays. Therefore, it is not certain whether 1% contrast is still a conservative assumption. To accurately predict visibility of distortions under a broad range of viewing conditions, more accurate models of detection threshold should be employed.

The detection threshold of the HVS is usually modelled in psychophysics with either a threshold versus intensity function (t.v.i.) or a more complex Contrast Sensitivity Function (CSF). The difference between them is that the t.v.i. function is measured for a fixed pattern, such as a circular patch on a uniform background, and the CSF is measured for a sinusoidal patterns or Gabor patches of different spatial frequencies. In our analysis we consider the most popular models of t.v.i. and CSF, which include:

- Ferwerda's t.v.i.,[14] which is commonly used in computer graphics;
- The t.v.i. model suggested by Bodmann[15] based on Blackwell's data[16] for 20–30 year old observers and adopted by the CIE standard[17];
- Barten's CSF model[10] adopted by the DICOM standard[8];
- and Meeteren's CSF model,[18] improved by Kodak and used in the Visual Difference Predictor (VDP).[19]

While t.v.i. functions can be used directly to replace the function $t(y_{adapt})$, some assumptions must be made before the thresholds can be found from a CSF. Sensitivity modelled by a CSF can depend on the stimuli size, viewing conditions, spatial and temporal frequency, eccentricity and orientation. To make a conservative choice,

‡Briefly, a shooting method is an iterative procedure that performs a binary search for the k value until the differential equation meets the boundary conditions.

§Weber fraction is usually defined as $W = (y_{max} - y_{min})/y_{min}$.

¶It was shown over 40 years ago that the Weber-Fechner law does not match the experimental data for luminance.[13] The discrepancy between the law and the real measurements is even higher for high dynamic range images.

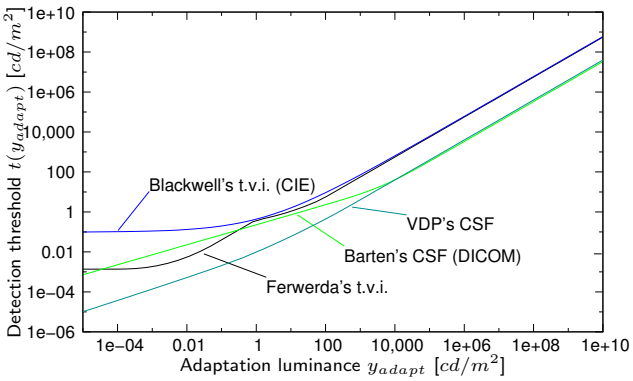

Figure 2. Comparison of the detection threshold models based on different CSF and t.v.i. functions.

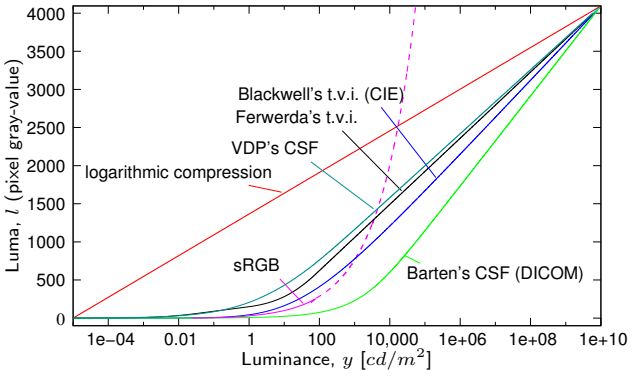

Figure 3. Luminance to luma mappings, derived from different threshold models. A logarithmic function and the sRGB color space are included for comparison.

we assume the worst case scenario and always choose the point on the CSF where sensitivity is the highest. Since sensitivity is modelled as an inverse of Weber's fraction, we get:

$$t(y_{adapt}) = \frac{y_{adapt}}{\max_{\rho} CSF(\rho, y_{adapt})} \tag{8}$$

assuming a simplified CSF, which is a function of spatial frequency ρ and luminance of adaptation y_{adapt}.

For comparison, the $t(y_{adapt})$ functions based on the above listed t.v.i. and CSF models are plotted in Figure 2. Note that all functions follow a similar shape, but they are also shifted along t-axis between each other. This comes from the difference in measuring methods and also from the differences in the peak sensitivity between individuals. In general, the CSF models show lower thresholds than the t.v.i. models.

Using each of the four detection threshold models, we found the coefficient k by solving the two point boundary problem, as described above, for the visible range of luminance and for 12-bit encoding. The resulting curves ($l(y)$ functions) are plotted in Figure 3. The constant k was above 1 for all functions (the smallest $k = 1.4481$ was found for VDP's CSF), therefore rounding the values of those functions to integer numbers does not introduce errors above the detection threshold. All four curves based on the t.v.i. or CSF data have slightly different shapes, resulting in different sensitivity for different luminance ranges. Additionally, Figure 3 contains two additional curves depicting the nonlinearity (gamma correction) used in the sRGB standard[20] and a logarithmic compression. The sRGB nonlinearity is plotted as a continuous line up to 80 cd/m^2, which is the display white luminance level assumed by the standard. The dashed line illustrates how the sRGB nonlinearity accelerates for high luminance levels, making it practically unsuitable for HDR data. The sRGB color space has not been designed to encode luminance levels above a few hundreds cd/m^2. Also, the logarithmic compression curve has been fit into 12-bit luma range. This curve is significantly different than the other functions, which model the human perception more accurately. The 12-bit logarithmic encoding resulted in a relative quantization error about 0.42%.

At this point, we removed both the curve derived from the Ferwerda's t.v.i. and the curve based on the Barten's CSF from further consideration. The Ferwerda's t.v.i. is based on data from very few subjects and is measured for cone and rod vision separately, therefore it is less plausible than the other curves. The curve derived from Barten's CSF results in too coarse quantization for luminance below $1,000$ cd/m^2 and too conservative quantization for luminance above that point (a steeper curve means that a luminance range is projected on a larger number of discrete sensory values, l, thus lowering quantization errors). Since we would like the quantization error to be at least as conservative as the quantization of the sRGB color space, this curve is not suitable for our application. The remaining two curves are equally suitable for encoding HDR and the choice between them may depend on the application. VDP's CSF is more conservative for low luminance. The curve derived from the CIE data is close to the gamma correction used in the sRGB color space, which gives better compatibility with low-dynamic range images, for which sRGB is de facto a standard.

We use a numerical method to derive the functions shown in Figure 3. However, for many applications, it is desirable to have an analytical formula, which could facilitate conversion between HDR luminance and 12-bit luma. We propose an analytical model that is both simple and resembles similar formulas used for the same purpose but for low dynamic range. We define a conversion from luminance to luma as:

$$l(y) = \begin{cases} a \cdot y & \text{if } y < y_l \\ b \cdot y^c + d & \text{if } y_l \le y < y_h \\ e \cdot \log(y) + f & \text{if } y \ge y_h \end{cases} \tag{9}$$

The above model is similar to the sRGB non-linearity, which also consists of linear and power function segments. The difference is that the above model additionally includes a logarithmic segment for high luminance.

To fit the model to the numerical solution of $l(y)$ for both the CIE and VDP's detection models, we use the Levenberg-Marquardt nonlinear regression. Additionally, we enforce C^1 continuity in y_l and y_h in order to achieve a smooth function. We get the best fit to the data for the constants listed in the table below:

Model	a	b	c	d	e	f	y_l	y_h
CIE t.v.i.	17.554	826.81	0.10013	-884.17	209.16	-731.28	5.6046	10469
VDP's CSF	769.18	449.12	0.16999	-232.25	181.7	-90.160	0.061843	164.10

An inverse mapping, from luma to luminance, can be found using the formula:

$$y(l) = \begin{cases} a' \cdot l & \text{if } l < l_l \\ b'(l + d')^{c'} & \text{if } l_l \le l < l_h \\ e' \cdot \exp(f' \cdot l) & \text{if } l \ge l_h \end{cases} \tag{10}$$

where the coefficients are given in the table below:

Model	a'	b'	c'	d'	e'	f'	l_l	l_h
CIE t.v.i.	0.056968	7.3014e-30	9.9872	884.17	32.994	0.0047811	98.381	1204.7
VDP's CSF	0.0013001	2.4969e-16	5.8825	232.25	1.6425	0.0055036	47.568	836.59

It is important to note that the model from Equation 9 is only an approximation of the accurate mapping function, derived by a numerical or analytical solution of Equations 4 or 5. The applications that require high accuracy of the predicted quantization errors should use the accurate solution rather than the approximate model. Although it is possible to design a more accurate model, it would be too complex to be practical. It is also important to note that a pre-computed lookup table for luma to luminance mapping can often give much better performance than an analytical formula that involves computationally expensive power and logarithmic functions. However, as we recognize that the lack of simple formulas often discourage the application of a method, we propose this simplified model as a better alternative to the logarithmic compression,

4.2. Chrominance and Chroma

Having derived the luminance component of the color space for HDR, we now focus on encoding chrominance as two 8-bit chroma channels. Using eight bits per channel to encode color is motivated by existing image formats, which often offer twelve or more bits for luminance channel, but rarely encode chrominance with higher precision then eight bits per channel.

Although an obvious choice for image and video compression would be a variant of YC_rC_b color space, we rejected it because of its limited color gamut. HDR frames should preserve the full visible color gamut (recall Property 1 from Section 4), even though it cannot be displayed on the existing displays. We have experimented with several color spaces, including a variant of RGB with an extended gamut (more saturated primaries), but finally we achieved the best results with the CIE 1976 Uniform Chromacity Scales (u', v'). Similarly as in [Ward Larson 1998[2]], we compute the values for chrominance channels using the equations:

$$\begin{aligned} u' &= \frac{4X}{X+15Y+3Z} \\ v' &= \frac{9Y}{X+15Y+3Z} \end{aligned} \tag{11}$$

Then we encode u' and v' using 8-bits:

$$u_{8bit} = u' \cdot 410$$
$$v_{8bit} = v' \cdot 410 \tag{12}$$

Note that we use u' and v' chromatices rather than u^* and v^* of the $L^*u^*v^*$ color space. Although u^* and v^* give better perceptual uniformity and predict loss of color sensitivity at low light (Property 2), they are strongly correlated with luminance. Such correlation is undesired in image or video compression (Property 5). Besides, u^* and v^* could reach high values for high luminance, which would be difficult to encode using only eight bits.

The remaining question is whether u_{8bit} and v_{8bit} lead to visible quantization errors and thus contouring artifacts (Property 4). It has been reported that skilled observers can see differences in u', v' of only about 0.002 (0.82 for u_{8bit} and v_{8bit}) (see [Hunt 1995[21]], p. 154), which is still below the maximum quantization error $u_{8bit} \pm 0.5$ and $v_{8bit} \pm 0.5$. For validation, we displayed a chromacity diagram for quantized u_{8bit} and v_{8bit} for several luminance levels on a calibrated monitor. We could see contouring artifacts for blue and purple colors for the highest luminance levels, which would suggest that 8-bit encoding does not give sufficient precision. This is alleviated by either limiting the color gamut or using perceptually more uniform color space (the u', v' chromacity diagram is only approximately uniform and the ratio between the smallest and the largest color difference can exceed four to one). However, such artifacts are not expected to be noticeable in complex images.

5. LOSSY HDR COMPRESSION OF STATIC IMAGES AND VIDEO

To validate the proposed color space, we implemented two lossy HDR compression algorithms: one for static images, the other for video.

The algorithm for encoding static HDR images is mostly based on the JPEG image encoding with a few extensions added to accommodate HDR data. Instead of YCrCb we use the color space derived in Section 4. Since luminance in this color space is encoded with 12 bits, both DCT transformation and variable-length coding are extended to support larger values. The results show that our DCT-based image compression for HDR images is both efficient and fast. The algorithm is specifically developed to be included in the Open Source OpenEXR library (www.openexr.org) as a freely available and efficient lossy compression format for HDR images. OpenEXR is a HDR image format developed by Industrial Light and Magic for the purpose of a special effect production. The format is currently promoted as a special-effect industry standard and is already supported by many software packages.

MPEG-4 video compression can be extended to encode HDR information as well.[22] The HDR video compression employs similar color space as proposed in this paper, but uses an 11-bit encoding of luminance, derived from Ferwerda's t.v.i. function. The resulting video streams contain complete photometric information and therefore can be used to show images on an advanced HDR displays. Full dynamic range content is also suitable for applying post-processing effects, which require physical luminance, rather than gamma-corrected pixel values. The additional complexity and bit-rate required to encode HDR are moderate, therefore our encoding scheme is an attractive extension to the existing video encoding standards.

6. DISCUSSION

The luminance encoding proposed in this paper can be considered an extension of typical gamma correction for the full range of luminance values visible to the human eye. Obviously, "gamma correction" is not the correct term for the proposed nonlinearity since we do not correct voltage of cathode ray tubes. Nevertheless, it is worth pointing out that both the gamma correction and the proposed nonlinearity are consistent in the luminance range from about 1 to 500 cd/m^2 (i.e. the luminance range in which typical CRT and LCD displays operate). In this range both nonlinearities are modelled as a power function with the exponent being less than one (see Equation 9).

Interestingly, there is also an analogy between the derived $l(y)$ function and the response of a typical film negative. The film response, as shown in Figure 4, consists of five segments: D_{min} (minimum density), the toe, the straight-line segment, the shoulder, and D_{max} (maximum density). Such a characteristic is also known as the

Figure 4. A response curve for a typical negative film. See text for details.

D-logE curve. If we compare the film response from Figure 4 with the $l(y)$ function from Figure 3, we notice that our visual system has a minimum response (below 0.01 cd/m^2) followed by the segment of gradually increasing slope, which corresponds to the toe in the film response. The visual system shows a logarithmic response above $1,000 - 10,000$ cd/m^2, similar to the straight-line segment (on log-linear plot) for a film. The difference is that the visual system, unlike a film, does not saturate for high luminance when it is adapted to these luminance levels. In addition, the eye can perceive simultaneously much larger dynamic range of luminance than a film can capture.

One difficulty that arises from our color encoding is that the source HDR images must be calibrated in absolute units of cd/m^2 (also absolute XYZ values, not normalized to 0-100 range). This is necessary since the performance of the HVS is significantly affected by the absolute luminance levels. For instance, the detection thresholds are significantly higher for low light conditions. The major source of this problem are the existing HDR capture techniques, such as multi-exposure methods, which give an accurate measurement of relative luminance (luminance factor), but give no information on absolute luminance levels. The conversion from relative to absolute luminance units is however very simple and requires multiplication of all XYZ color coordinates by a single constant. Such a constant needs to be measured only once for a camera. The measurement can be done by capturing a scene containing a uniform light source of known illuminance or a surface of measured luminance. If such a measurement is not possible, an approximate calibration of an image to absolute units, by assuming typical luminance levels of some objects (e.g. the sky or a daylight illuminated wall), is usually sufficient.

Although we strongly support scene-referred encoding of image and video we also see some problems related to this approach. A substantial part of the visual material created today is not an exact replica of the real world, but rather stems from human or computer-created or enhanced images, which are only intended to look like the real scenes. For instance, night scenes in movies are often shot at daylight and then post-processed to give them a nocturnal look. How should such scenes be encoded if they intend to represent low light conditions but are displayed at much higher luminance levels? In such cases, scene-referred encoding of images may not be appropriate and images should represent the intended appearance of a scene. Nevertheless, such images should be stored in an HDR "appearance-referred" format, which would encode the optimal luminance levels at which particular scene should be displayed. If a display device is not capable of displaying such an image, it would apply a tone mapping algorithm[5] to deliver the best image for its capabilities.

The luminance encoding proposed in this paper is limited by the capabilities of the HVS, therefore it is not suitable for those applications where precision larger than that of the HVS is required. This could include remote sensing and some cases of medical imaging. However, since most of imaging applications have the human eye as the target "consumer", the proposed encoding gives in all those cases the best trade-off between the precision and the efficiency of encoding.

7. CONCLUSIONS

To enrich visual information stored in image or video files, we postulate scene-referred encoding in favor of device-referred representation commonly used today. HDR images are an example of such scene-referred encoding, which unlike plain images can represent the whole visual information visible to the human eye. We show that HDR

scene-referred images and video can be efficiently encoded. We derive a color space for efficient encoding of HDR data from the detection thresholds of the HVS. To test and demonstrate efficiency of our approach, we implement complete HDR JPEG and MPEG-4 encoders.

REFERENCES

1. G. Ward, "Real pixels," *Graphics Gems II* , pp. 80–83, 1991.
2. G. Ward Larson, "Logluv encoding for full-gamut, high-dynamic range images," *Journal of Graphics Tools* **3**(1), pp. 815–30, 1998.
3. R. Bogart, F. Kainz, and D. Hess, "OpenEXR image file format," in *ACM SIGGRAPH 2003, Sketches & Applications*, 2003.
4. G. Ward and M. Simmons, "Subband encoding of high dynamic range imagery," in *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, pp. 83–90, 2004.
5. E. Reinhard, G. Ward, S. Pattanaik, and P. Debever, *High Dynamic Range Imaging. Data Acquisition, Manipulation, and Display*, Academic Press, 2005.
6. M. Sezan, K. Yip, and S. Daly, "Uniform perceptual quantization: Applications to digital radiography," *IEEE Transactions on Systems, Man, and Cybernetics* **17**(4), pp. 622–634, 1987.
7. J. Lubin and A. Pica, "A non-uniform quantizer matched to the human visual performance," *Society of Information Display Int. Symposium Technical Digest of Papers* (22), pp. 619–622, 1991.
8. DICOM PS 3-2004, "Part 14: Grayscale standard display function," in *Digital Imaging and Communications in Medicine (DICOM)*, National Electrical Manufacturers Association, 2004.
9. H. Wilson, "A transducer function for threshold and suprathreshold human vision," *Biological Cybernetics* **38**, pp. 171–178, 1980.
10. P. G. Barten, *Contrast sensitivity of the human eye and its effects on image quality*, SPIE – The International Society for Optical Engineering, P.O. Box 10 Bellingham Washington 98227-0010, 1999. ISBN 0-8194-3496-5.
11. M. Ashikhmin, "A tone mapping algorithm for high contrast images," in *Rendering Techniques 2002: 13th Eurographics Workshop on Rendering*, pp. 145–156, 2002.
12. W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C*, Cambridge Univ. Press, 1993.
13. S. Stevens and J. Stevens, "Brightness function: parametric effects of adaptation and contrast," *Journal of the Optical Society of America* **50**, p. 1139A, Nov. 1960.
14. J. Ferwerda, S. Pattanaik, P. Shirley, and D. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of SIGGRAPH 96*, *Computer Graphics Proceedings, Annual Conference Series*, pp. 249–258, Aug. 1996.
15. H. Bodmann, "Visibility assessment in lighting engineering," *Journal of the Illuminating Engineering Society* **2**(4), pp. 437–443, 1973.
16. O. Blackwell and H. Blackwell, "Visual performance data for 156 normal observers of various ages," *Journal of the Illuminating Engineering Society* **1**(1), pp. 3–13, 1971.
17. CIE, *An Analytical Model for Describing the Influence of Lighting Parameters Upon Visual Performance*, vol. 1. Technical Foundations, CIE 19/2.1, International Organization for Standardization, 1981.
18. A. Van Meeteren and J. J. Vos, "Resolution and contrast sensitivity at low luminances," *Vision Research* **12**, pp. 825–833, 1972.
19. S. Daly, "The Visible Differences Predictor: An algorithm for the assessment of image fidelity," in *Digital Image and Human Vision*, A. Watson, ed., pp. 179–206, Cambridge, MA: MIT Press, 1993.
20. IEC 61966-2-1:1999, *Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB*, International Electrotechnical Commission, 1999.
21. R. Hunt, *The Reproduction of Colour in Photography, Printing and Television: 5th Edition*, Fountain Press, 1995.
22. R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," *ACM Transactions on Graphics (Proc. of SIGGRAPH 2004)* **23**(3), pp. 730–738, 2004.