High Dynamic Range for HD and Adaptive Bitrate Streaming

A Technical Paper prepared for SCTE/ISBE by

Sean T. McCarthy, Ph.D.
Independent Consultant
Sean McCarthy, Ph.D. Consulting
236 West Portal Avenue, #293
San Francisco, CA 94127
415-518-5287
sean.mccarthy@comcast.net

Table of Contents

List of Figures

List of Tables

Introduction

When High Dynamic Range (HDR) began to emerge a few years ago as a great new television experience, it was almost invariably thought of as only one part of Ultra HD/4k along with a wider range of colors and higher bit depths for video coding. Now though, many are wondering if HDR can be an independent contributor that makes for better TV experiences even without higher 4k resolutions. Indeed, several MVPDs, broadcasters and industry associations are moving towards enhanced programming and distribution that marries HDR with HD and sub-HD resolutions. The next-generation television standard, ATSC 3.0, for example, enables distribution of HDR and Wide Color Gamut (WCG) content at any resolution. Similarly, leading internet-based streaming services are leveraging multi-resolution adaptive bitrate (ABR) protocols to deliver HDR experiences.

The potential wrinkle is that HDR is still often tied to Ultra HD/4k/10-bit in the content creation studios and production houses. The HDR highlights that make HDR shine are often very small and localized. Thus, converting original Ultra HD HDR content to lower resolutions for HD and ABR streaming runs the risk of obliterating those visually potent HDR highlights. When the resulting video is compressed with High-Efficiency Video Coding (HEVC), HDR distortions could be made worse, particularly for aggressive streaming profiles.

In this paper, we use a recently developed method for measuring distortion in HDR video to quantify the impact of encoding original Ultra HD HDR content at HD and ABR resolutions. Our method is intrinsically independent of any particular HDR transfer function and may thus be applied to HDR content having Hybrid-Log Gamma (HLG), Perceptual Quantizer (PQ), or other transfer characteristic. Specifically, we show:

1) That each frame of HDR video can be represented as the sum of a Spatial Detail signal and a Foundation Image. The Spatial Detail signal contains the localized contrast variations associated with HDR features. The Foundation Image contains the more smoothly varying large-area contrast variations associated with textures and backgrounds.
2) The Spatial Detail signal is the main contributor to HDR distortions introduced by rescaling and compression. The Foundation Image contributes little to total HDR distortion.
3) Quantifying the correlation between the Spatial Detail signal of processed images and corresponding original images provides a systematic method to choose the best combinations of HEVC-encoded resolutions and bitrates to minimize overall HDR distortion.

Our key objectives in this presentation is to provide a practical method that can help ensure great HDR experiences at any resolution.

Content

1. Background

IP-based protocols and streaming are rapidly becoming powerful methods for distributing television to all screens from the big screens in living rooms to smaller-screen smartphones and tablets. At the same time, displays big and small are becoming much more capable of rendering the deep darks and bright highlights that make HDR[1,2] special.

Ideally, content producers and distributers would like HD variants of Ultra HD/4k HDR to differ as little as possible from the original full-resolution content. We would also like to create a consistent HDR experience across all screens, big and small. A challenge is that small screens tend to have lower resolution than big screens. They also tend to be used more often at the edge of lower-bandwidth wireless and more congested networks.

To create high-quality consistent HDR experiences, we need a method to quantify the differences between studio-approved full-resolution HDR content and corresponding HD and sub-HD variants that might be distributed over DOCSIS and adaptive streaming protocols that switch between encoded resolutions as network conditions vary.

Service providers choose specific combinations of encoded resolution and bitrate to enable best-quality IP and adaptive streaming. The set of combinations as called an adaptation set, or sometimes called an encoding ladder. In adaptive streaming, client players choose the specific resolution-bitrate pair available in the adaptation set based of the clients' available bandwidth and other performance criteria.

The key questions for service providers are the following. Which combinations of resolution and bitrate should be included in HDR adaptation sets to minimize HDR distortions and promote consistency across screens? Which combinations minimize the visibility of switching between the rungs of the encoding ladder? This paper provides methods to help answer these questions.

2. Test Sequences & Preparation

In this study, we used the Meridian[3] HDR test content as shown in Figure 1. Meridian is Ultra HD HDR 3840x2160 60 fps test content graded to 4000 cd/m^2 with transfer characteristics and color primaries specified by ITU-R BT.709[4]. (See the note below regarding BT.709 HDR transfer characteristics.) Meridian was professionally produced by Netflix and is publicly available for download[5] in Interoperable Master Format (IMF) (see SMPTE OV 2067[6]) for which the video essence is an MXF[7] file containing JPEG2000[8]-encoded 10-bit YCbCr 4:2:2 video data. To perform the calculations presented in this paper, we extracted data for individual frames using ffmpeg[9].

(Note on ITU-R BT.709 transfer characteristics for Meridian. HDR content is typically encoded using either HLG, an Opto-Electronic Transfer Function (OETF) specified in ITU-R BT.2100[10], or PQ, an Electro-Optical Transfer Function (EOTF) specified in SMPTE ST 2084[11]. Thus, it is perhaps surprising that Meridian makes use of ITU-R BT.709, an OETF originally intended for High-Definition TV (HDTV) and which was standardized before the emergence of HDR. For the purposes of this paper, it is not an issue. Our methodology is independent of HDR transfer characteristics, though the results of our method could be expected to provide tighter correlation with subjective video quality when applied to PQ- or HLG-encoded video.

It may be that BT.709 was used because Meridian was produced before HDR-support in IMF was standardized. The availability of Meridian was announced[3] in September 2016 and IMF did not include HDR support until shortly before that with the publication of SMPTE ST 2067-21[12] in July 2016.)

Figure 1 - Ultra HD HDR Test Sequence Used in this Study

Figure 2 - HD HDR Test Sequences Used in this Study

There is a dearth of contribution-quality 4k HDR content available for testing. Thus, we also used the HD-HDR WCG test sequences shown in Figure 2. These sequences were created by the "HdM-HDR-2014 Project"[13,14] to provide professional quality cinematic wide gamut HDR video for the evaluation of tone mapping operators and HDR displays. All clips are 1920x1080p24 and colour graded for ITU-R BT.2020[15] primaries & 0.005-4000 cd/m^2 luminance. We converted the original colour graded frames (RGB 48 bits per pixel TIFF files) to Perceptual Quantizer (PQ) YCbCr v210[16] format (4:2:2 10 bit) using the equations defined in ITU-R BT.2020, ITU-R BT.1886[17], and ITU-R BT.2100.

For each HdM-HDR-2014 test sequence, we created 50 variants having different encoded resolutions and bitrates. Of the 50 variants, 20 were raw uncompressed versions used to isolate the impact of different rescaling algorithms on HDR distortion. The remaining 30 variants for each test sequence were compressed using HEVC for each combination of encoded resolution (1920x1080, 1440x1080, 1280x720, 960x540, 720x540, and 640x360) and bitrate (10000, 3000, 1000, 300, and 100 kbps).

All rescaling was performed in Matlab[18] using the *imresize* function. All compression was performed using command-line x265[19] (main10 profile).

3. Decomposing an HDR Frame into Spatial Detail and Foundation Images

Our approach to measuring HDR distortion begins with decomposing each HDR frame (A) into a Spatial Detail signal (B) and a Foundation Image (C), as illustrated in Figure 3. The pixel-by-pixel sum of the Spatial Detail signal and Foundation Image is equal to the original image.

Figure 3 - Decomposition of an HDR frame into Spatial Detail and Foundation Image

The method of calculating the Spatial Detail signal is described in the Appendix and detailed more thoroughly in previous publications[20-23]. In brief, the Spatial Detail signal can be thought of as the condensed essence of the original image. It isolates the features and details that are unique to the original while minimizing statistically expectable characteristics that the original image shares with images as a statistical class.

Images of natural and other complex scenes have an interesting statistical property: They have spatial-frequency magnitude spectra that tend to fall off with increasing spatial frequency in proportion to the inverse of spatial frequency[24]. The magnitude spectra of individual images can vary significantly; but, as an ensemble-average statistical expectation, it can be said that "the magnitude spectra of images of natural and other complex scenes fall off as one-over-spatial-frequency."

The Spatial Detail signal is effectively the result of de-emphasizing the statistically expectable one-over-frequency characteristic. As such, the Spatial Detail signal emphasizes the unique unexpectable details in an image.

The Foundation Image is obtained by simply subtracting the Spatial Detail image from the original HDR image. As such, the Foundation Image may be thought of as a special kind of low-pass filtered version of the original HDR image in which the unique spatial details are selectively attenuated. Although perhaps difficult to appreciate on the printed page, the Foundation Image gives the visual sensation of being out of focus.

One way to think of the Foundation Image is that it represents the overall contrast and luminance range of the HDR image, whereas the Spatial Detail signal can be thought of as representing localized contrast and luminance variations.

The concept of decomposing an image into coarse and fine components is not new. Indeed, it is the basis for scalable video encoding schemes such a Scalable High Efficiency Video Coding[25] (SHVC) that is included in ATSC 3.0[26]. Wavelet-based video compression, such as JEG2000, also relies on decomposition.

We use decomposition for a different reason in this paper. SHVC, for example, provides for a base layer that that can be decoded and displayed. SHVC also provides enhancement layers that can be added to the base layer to improve quality and detail. The amount of enhancement that can be achieved depends on bandwidth availability and the capabilities of the decoder and display. SHVC is thus a scheme to optimize

the quality and detail of a displayed video for various network and user-device situations. JPEG2000 combines layers of differing detail to achieve target bitrates and visual quality.

We do not intend the Foundation Image to ever be displayed, nor should it be thought of as the lowest-resolution variant in an encoding ladder. Rather, our decomposition into a Spatial Detail signal and Foundation Image is strictly a mathematical way of concentrating most of the distortions introduced in processing HDR video into a more concise signal than the original HDR video. The resulting concentration of visual detail provides a useful way to quantify HDR distortion, as described below.

4. Spatial Detail Signal as a Guide to Features and Textures

The histogram of the Spatial Detail signal is shown in Figure 4. A special property of the Spatial Detail signal is that the shape of its histogram, which resembles a two-sided exponential distribution, is preserved across images. For example, the Spatial Detail signals for the different HdM-HDR-2014 test sequence all have Spatial Detail histograms that resemble two-sided exponential distributions. Like scale-invariance, the shape of the histogram appears to be an expectable statistic of natural and complex images.

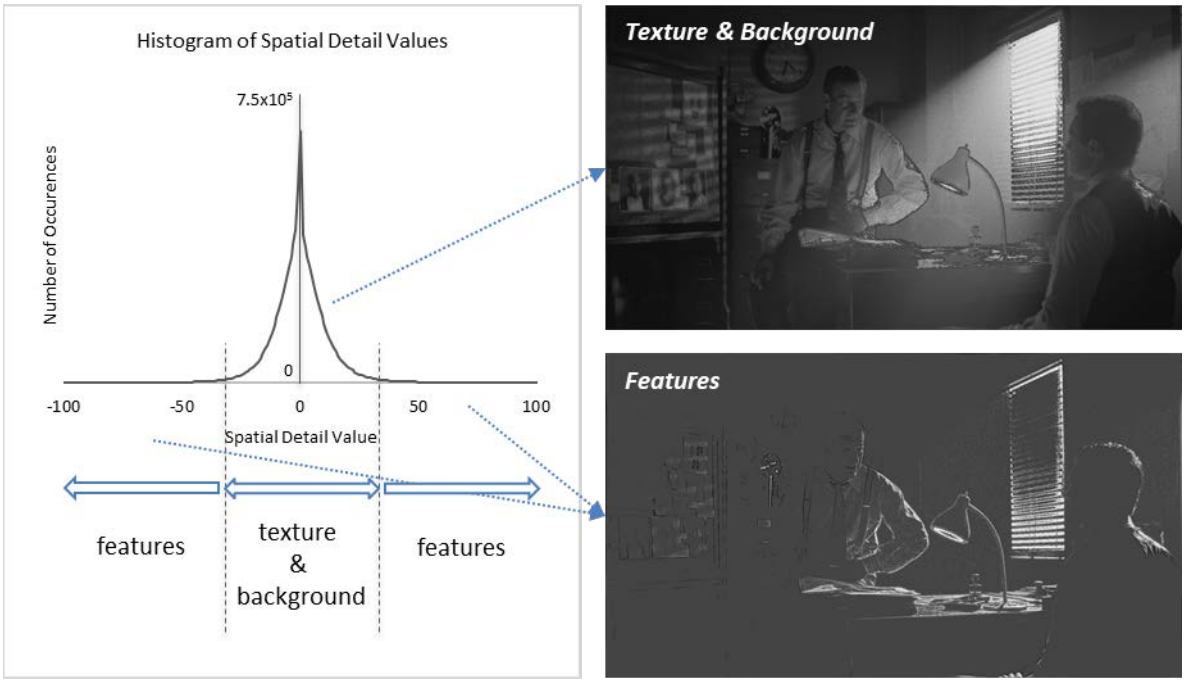


Figure 4 - Spatial Detail Signal as a Guide to Features, Textures, and Background

Another special property of the Spatial Detail signal is that its values provide a convenient guide to the parts of an image that would tend to be called features and the parts that would tend to be called textures and background. Large absolute values of the Spatial Detail signal tend to correlate with features. Small absolute values tend to correlate with textures and background. The images in Figure 4 were created by creating a masking image by applying a threshold to the absolute value of the Spatial Detail signal, and then applying the mask to the original luma values of the HDR frame data. The “Features” image are the original luma values for every pixel at which the absolute values of the Spatial Detail signal exceeded the applied threshold. The “Textures & Background” image is the compliment of the “Features” image.

5. Spatial Detail Distortion is Most of the Total HDR Distortion (I)

An advantage of decomposing each HDR frame into a Spatial Detail signal and Foundation Image can be appreciated by examining the relative contribution of each to total HDR distortion.

Mean-Squared Error (MSE) and Peak-Signal-to-Noise Ratio (PSNR) are common and equivalent metrics of video distortion[27,28]. (PSNR is proportional to the logarithm of MSE.) MSE is the average over all pixels of the squared difference between an original image and a corresponding encoded variant.

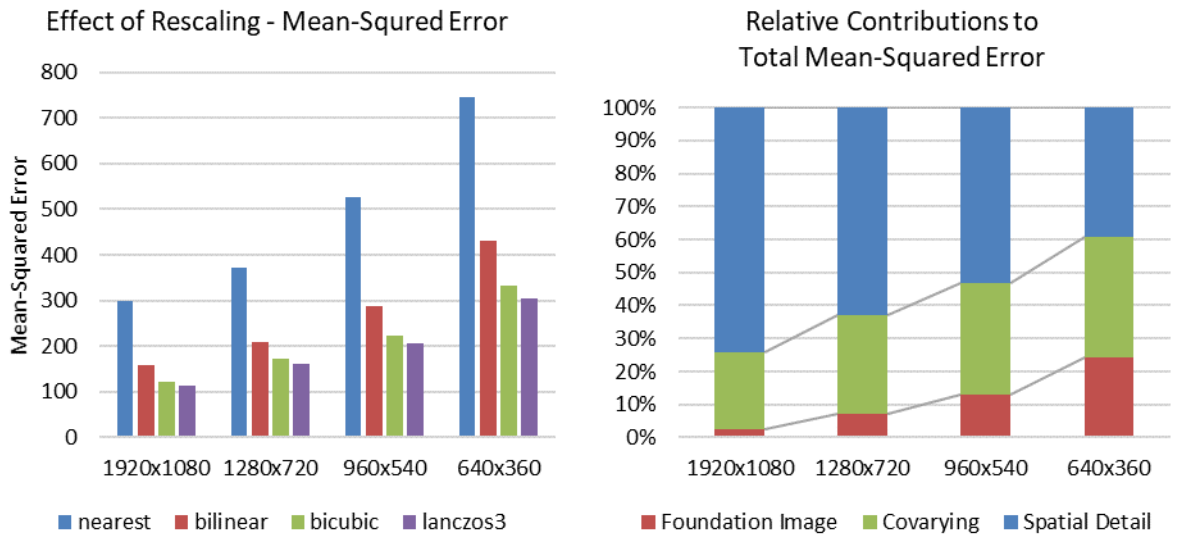


Figure 5 - Relative Contribution of the Spatial Detail Signal and Foundation Image to Total MSE

MSE values calculated for four encoded resolutions and four candidate rescaling algorithms are shown in Figure 5. The candidate rescaling algorithms are nearest-neighbour interpolation, bilinear interpolation, bicubic interpolation, and lanczos3 resampling[29]. The MSE values shown in Figure 5 are for "Meridian". Lower MSE values indicate that the rescaled variant is less distorted from the original in terms of squared-error (and thus PSNR). The data show that lanczos3 resampling provides the lowest MSE values for all resolutions and should thus be considered the best choice in constructing adaptive streaming variants. If other considerations such as processing demands are significant, bicubic interpolation can deliver nearly as good results.

Total MSE is the mathematical sum of the sum of the Spatial Detail MSE by itself, the Foundation Image MSE by itself, and a contribution from the covariance of Spatial Detail signal and the Foundation Image.

The data in Figure 5*B* show that the Foundation Image MSE is a small fraction of the total MSE. The Foundation Image contribution to total MSE increases with progressively more aggressive downscaling, which indicates that rescaling progressively distorts the underlying smoothly-varying luminance of the encoded video. Yet, even for the 4-fold downscaling from the original 3840x2160 to 960x540, the error associated with the Foundation Image is only 10-15% of the total error. Most of the total error is attributable to distortion of the Spatial Detail signal (approximately 75% for 1920x1080, ~65% for 1280x720, and ~ 55% for 960x540. For 640x360 – a 6-fold downscaling -- no single portion of MSE is

the majority though the Spatial Detail contribution remains largest. (The data shown in Figure 3B are for lanczos rescaling.)

6. Spatial Detail Distortion is Most of the Total HDR Distortion (II)

Figure 6 and Figure 7show another way of visualizing distortions introduced by processing HDR content. The data in Figure 6*A* show the original luma code values (horizontal axis) plotted against average encoded luma code values (vertical axis). The data are for the 3840x2160 Meridian HDR test content rescaled at four sub-4k resolutions (1920x1080, 1280x720, 960x540, and 680x320 using lanczos3 resampling). The rescaled luma values display an almost perfect linear correlation with the original values: The encoded values fall along a straight line with a slope of 1. The exceptions, in this example, are for bright regions having code values above ~900 (dashed square) and dark regions having code values less than ~75 (dashed circle with arrow) within the valid 10-bit luma range of 64 to 940. (For some other test content, not shown, the elevation of dark luma values is more significant). The data in the bright range of Figure 6*A* is shown close-up in Figure 6*B*. For the Meridian Ultra HD HDR test content, rescaling results in a progressively decreasing luma values with increasingly aggressive downscaling.

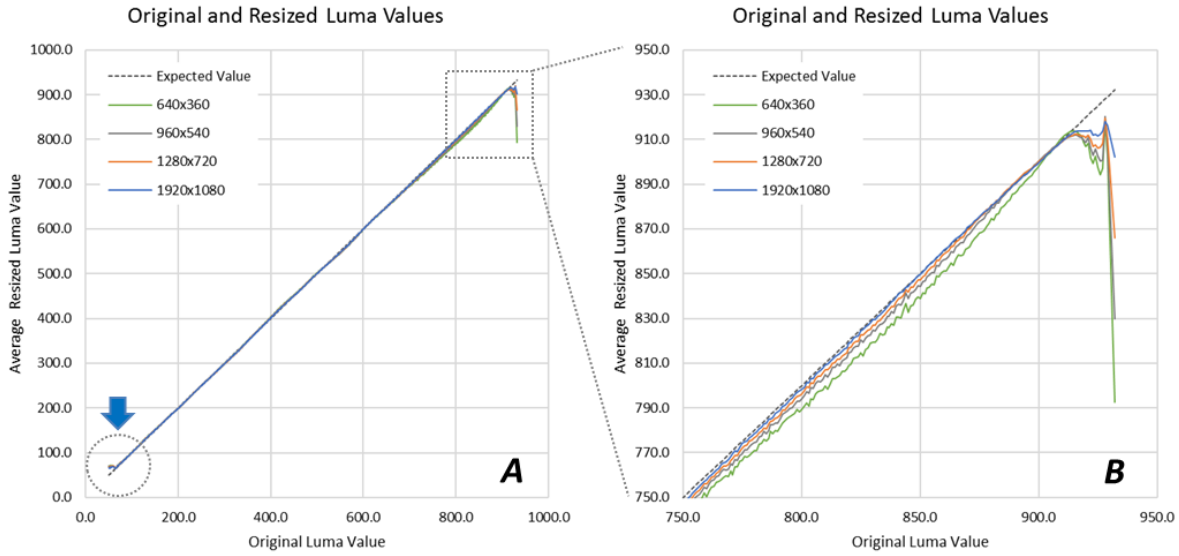


Figure 6 - Correlation Between Original and Rescaled Luma Values

The distortions of very dark and very bright regions appear to be a result of spatial averaging during resizing. For a finite range of allowed luma values, values near the limits of the range would tend to be averaged with neighbour values nearer to the center of the range, particularly for small isolated bright and dark regions. Thus, extreme values would tend to be pulled away from the upper and lower limits towards more central values during the resizing process. The data shown in Figure 6 indicate that there is room to develop HDR-sensitive resizing algorithms that are better than even laczos3 resampling for use in multi-resolution HDR video services.

(A note on calculating average code values – Each average code values in the rescaled and/or encoded frame was calculated by finding all the pixel locations in the original frame that have a specific code value, for example, 312 out the possible range of 64 to 940 for 10-bit encoding. The encoded code values

for the corresponding pixel locations tend to have a distribution of values because of scaling and/or compression. We use the average over the distribution. Thus, the data in Figure 6, and other similar figures in this paper, provide a view of the correlation between expected values (original value) and the corresponding average code value in the processed image.)

The data plotted in Figure 7A & B provide more details on the nature of the luma distortion evident in Figure 6. In Figure 7A, there is negligible apparent distortion of the Foundation Image component of the rescaled image compared to the Foundation Image component of the original. (Data for all rescaled resolutions are plotted in Figure 7A. The data lay on top of each other thus giving the appearance of a single line.) On the other hand, the Spatial Detail signal of the rescaled image (Figure 7B) is significantly different from the Spatial Detail signal of the original. If there were no Spatial Detail distortion, the data would lay along the dashed line having unity slope. Thus, the conclusion is that luma distortion evident in Figure 6 is mainly a result of distortion of the Spatial Detail signal with hardly any contribution from the Foundation Image component.

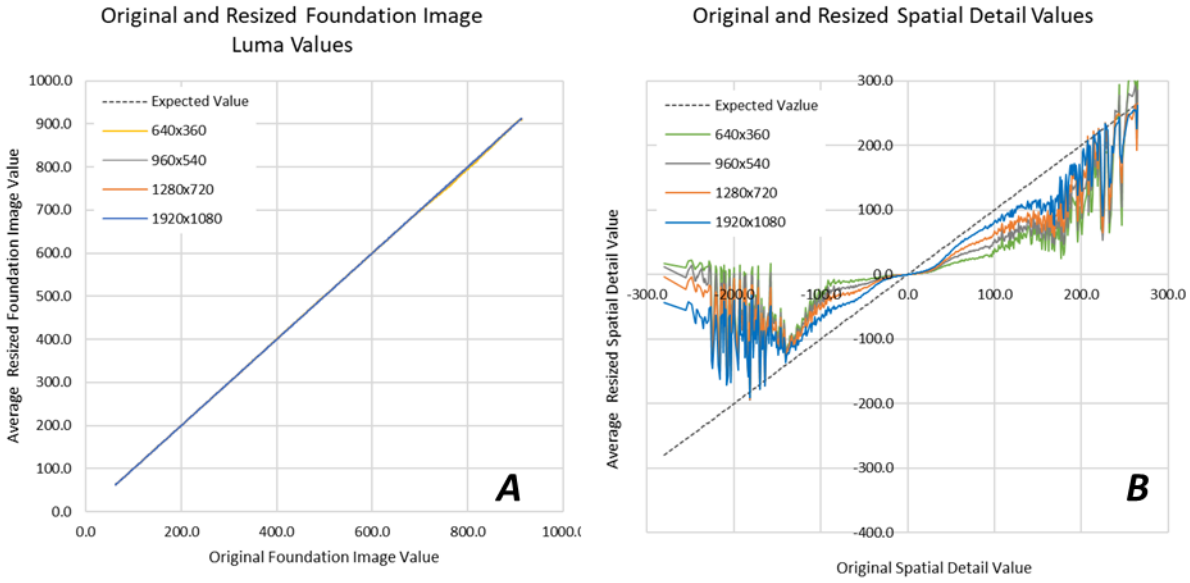


Figure 7 - Correlations for Rescaled Foundation Image and Spatial Detail Values

The distortion of the Spatial Detail component has several interesting features that provide insight into the effects of rescaling.

First, Spatial Detail signals of the rescaled images tend to follow a progressively shallower slope with increasingly aggressive downscaling. This indicates a progress systematic loss of local contrast in the HDR image. That is, the range of Spatial Detail values becomes narrower with rescaling. (The histogram of the Spatial Detail signal of rescaled images becomes narrower and more peaked.)

Second, the slope of the Spatial Detail signal is near zero near the origin of Figure 7B. This is the range in which the original Spatial Detail values have small absolute values. In other words, this is the range that tends to identify the regions in the original luma image that can be thought of as texture & background (see Figure 4). The conclusion is that rescaling more severely smooths low-amplitude textures and film grain and has a relatively lesser impact on larger-amplitude features.

Third, the average values of Spatial Detail signal of the rescaled image become progressively noisier and decorrelated with the value of the Spatial Detail signal of the original as the absolute value of the original Spatial Detail increases. We find that the decorrelation is content- and compression-dependent. Thus, quantifying the decorrelation can be a useful metric, as we show below.

7. Choosing Best Combinations of Resolution & Bitrate

HEVC compression introduces its own distortions beyond those introduced by rescaling. Total distortion depends on the interaction of rescaling and bitrate. Reducing encoded resolution at an encoded bitrate can actually reduce total distortion and increase video quality; but only to a point beyond which further reductions in encoded resolution increases total distortion. Minimizing distortion is a matter of finding the best balance between encoded resolution and bitrate.

In this study, we quantify HDR distortion by measuring the correlations between processed and original Spatial Detail signals for many bitrate-resolution combinations. Recall that the Spatial Detail signal represents most of the total distortion.

We measure correlation with the coefficient of determination, R^2, the square of the Pearson correlation coefficient, R, (see ref. 30). R^2 quantifies the "goodness of fit" of data to a linear regression line. The expectation is that the average code value of the encoded frame would be the same as the corresponding code value of the original frame. Mismatches are a manifestation of a lack of correlation and result in a lower value of R^2, which has a range of 0 to 1.

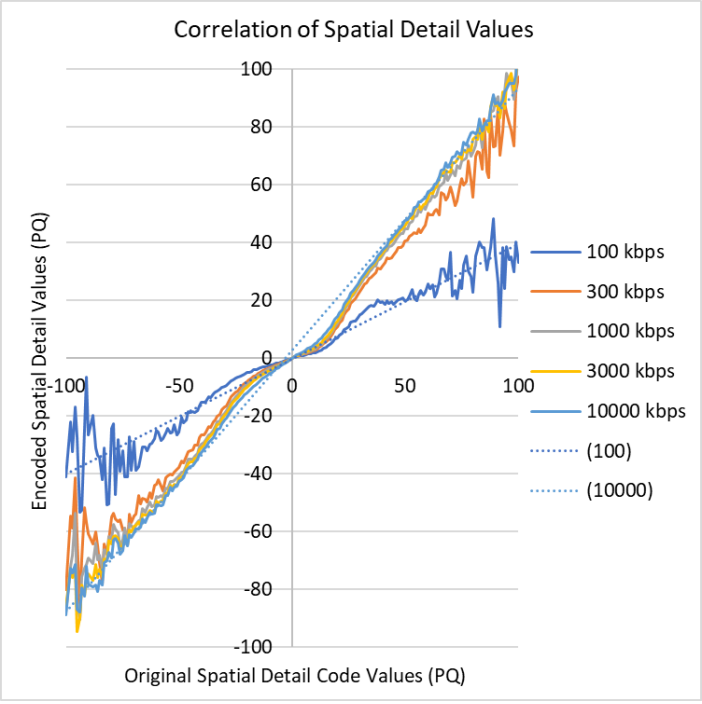


Figure 8 - Spatial DetailCorrelations for HEVC Compressed and Rescaled HDR Content

The data plotted in in Figure 8 illustrate Spatial Detail correlations for HEVC-compressed versions of the HdM-HDR-2014 test content. Each test sequence was encoded at 1920x1080 resolution and bitrates from

100 kbps to 10 Mbps. Though not shown, we also encoded the test clips at the same bitrates for each downscaled resolution 1440x1080, 1280x720, 960x540, 720x540, and 640x360. We used the HD-HDR test sequences because they comprise a richer more complete test set, and thus more reliable, than the single 4k-HDR example of the Meridian test sequence.

In Figure 8, there is a general overall linear relationship between encoded and original Spatial Detail, but there are three significant differences worth noting. First, the magnitude of the variation around the fitted straight-line (dashed) increases with decreasing bitrate. This indicates that Spatial Detail correlation decreases with decreasing bitrate. Second, the slope of the fitted straight-line decreases with decreasing bitrate. This indicates that localized contrasts of textures and HDR highlights are diminished with decreasing bitrate. Third, the slope of the correlation is flatter near the origin (small Spatial Detail values) than for large Spatial Detail values. This indicates that low contrast textures and details (such as those typically associated with faces and background textures) are systematically impacted more severely by HEVC compression than are high contrast HDR textures.

(Although not the main purpose of this paper, it is worth pointing out that Spatial Detail plots such as Figure 8 could be a useful engineering tool. Such plots provide a quantitative view of how compression and processing algorithms differentially affect textures, features, noise, and overall Spatial Detail contrast. Video encoder developers and video compressionists could use such information to optimize algorithms and parameter selection to improve visual quality and compression efficiency.)

We calculated an R^2 value for each of the 30 permutations of HEVC bitrate and encoded resolution for each of the 5 test sequences. We analysed the resulting R^2 values in the spirit of ATIS-0800061[31], "Methodology for Subjective or Objective Video Quality Assessment in Multiple Bit Rate Adaptive Streaming."

Table 1 summarizes the goodness-of-fit, R^2 values, for the bitrate-resolution combinations used in this study. The cells highlighted in green indicate which encoded resolution maximizes the goodness-of-fit for each bitrate. In other words, the green-highlighted cells correspond to the bitrate and resolution combinations that maximize the correlation between encoded Spatial Detail and original Spatial Detail. These are the bitrate-resolution combinations that best preserve the textures, highlights, and local contrast variations in HDR video. Higher resolutions at lower bitrates introduce compression distortion. Lower resolutions at higher bitrates introduce rescaling distortion. Our quantification of Spatial Detail correlation highlights the bitrate-resolution combinations that provide the best balance.

Table 1 – Choosing Bitrate & Resolution Combinations based on Spatial Detail Correlation

Resolution	Bitrate (kbps)				
	100	300	1000	3000	10000
1920x1080	0.900	0.933	0.946	0.966	0.983
1440x1080	0.893	0.939	0.949	0.968	0.982
1280x720	0.919	0.942	0.955	0.965	0.976
960x540	0.903	0.934	0.946	0.955	0.962
720x540	0.892	0.924	0.935	0.943	0.950
640x360	0.869	0.897	0.902	0.909	0.914

The data in Table 1 are the average over all HdM-HDR-2014 test sequences and thus represent a compromise trade-off for any particular test sequence. The compromise is arguably unfairest at low bit

rates. As illustrated in Table 2, some test sequences from benefit from lower-resolution encoding at low bit rates whereas others benefit from higher-resolution encoding.

Table 2 illustrates the use of Spatial Detail correlation to choose the best variants to include in HDR adaptation in a content-aware manner, also called per-title encoding. Note that the *carousel_fireworks* test sequence benefits from lower resolution variants than does the *bistro* test sequence. The *carousel_fireworks* test sequence is more challenging because it has a wider range of luminance and more motion. From a practical standpoint, Table 2 is a guide to setting encoder parameters to produce the best content-specific per-title encoding ladders.

Table 2 – Choosing Content-Dependent Bitrate & Resolution Combinations

Resolution	bistro					carousel_fireworks				
	100	300	1000	3000	10000	100	300	1000	3000	10000
1920x1080	0.958	0.987	0.995	0.997	0.999	0.841	0.848	0.878	0.942	0.985
1440x1080	0.956	0.990	0.995	0.997	0.998	0.848	0.879	0.898	0.956	0.989
1280x720	0.979	0.990	0.993	0.995	0.996	0.841	0.901	0.932	0.966	0.986
960x540	0.963	0.985	0.987	0.988	0.988	0.860	0.906	0.927	0.948	0.968
720x540	0.966	0.983	0.988	0.988	0.989	0.866	0.916	0.931	0.946	0.960
640x360	0.950	0.966	0.967	0.967	0.968	0.869	0.912	0.920	0.927	0.929

Conclusion

In this paper, we provided methods to quantify HDR distortions and take steps to mitigate those distortions by selecting the best combinations of bitrates and resolutions to include in HDR adaptation sets used in adaptive streaming services.

A key part of our approach is to decompose HDR video into two components for the purpose of mathematical analysis. A Foundation Image contains the overall luminance and contrast variations in HDR video. A Spatial Detail signal contains the localized luminance and contrast variations.

We showed that most of the distortion in encoded HDR content is a result of distortions of the Spatial Detail signal.

We proposed that the correlation between the encoded and original Spatial Detail is a useful metric by which to design encoding ladders; i.e., the best combinations of bitrate and resolution for HDR television services. This approach maximizes the similarity of the local contrasts and highlights between encoded and original HDR videos. These local contrasts and highlights add significant impact to HDR video and represent a large part of the content creator's original intent. Significantly, the use of Spatial Detail correlation can be used to design content-aware per-title encoding ladders as well as overall generic encoding ladders.

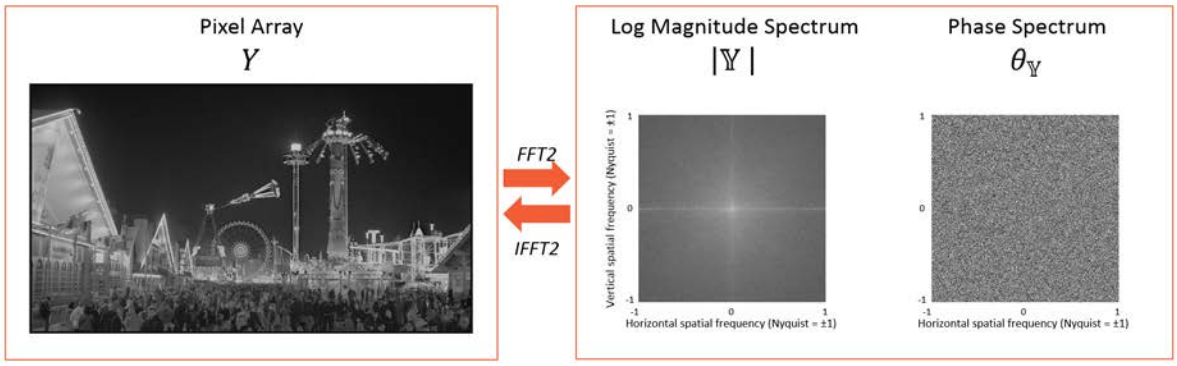
We also provided data that indicate that plots of Spatial Detail correlation could be a useful engineering tool. Such plots provide insight into the relative impact of compression and processing on different characteristics of HDR video, such as: noise, textures, and features.

In addition, we provided data that indicates that there is room to develop better HDR-sensitive resizing algorithms. The data show systematic distortions in very dark and very bright regions. New luminance-adaptive resizing algorithms might be worth investigating to mitigate such distortions.

Our next steps are to evaluate our proposed methods on a larger set of HDR data. Current publicly available HDR test material is limited and was produced before the latest HDR international standards were published. Our methodology is independent of any particular HDR transfer characteristics. Nonetheless, the application of our methodology to PQ and HLG HDR transfer characteristics might provide better or worse agreement with human opinions of video quality. This is the topic for a follow-on paper.

Appendix

The method of creating the Spatial Detail signal can perhaps best be understood by thinking of an image in terms of spatial frequency spectra as illustrated in Figure 9 (only the luma channel is shown). Any 2-dimensional array of pixel values can also be represented without loss of information as the product of a magnitude spectrum and a phase spectrum in 2-dimensional spatial frequency space. Spatial-frequency spectra can be obtained from an image pixel array by performing a 2-dimensional Fast Fourier Transform (FFT2). The pixel array can be recovered by performing a 2-dimensional Inverse Fast Fourier Transform (IFFT2). FFT2 and IFFT2 are well known signal processing operations that can be calculated quickly in modern processors.



$$FFT2\big(Y(x,y)\big) = \mathbb{Y}(k_x,k_y) = \big|\mathbb{Y}(k_x,k_y)\big| * exp\left(i\theta_{\mathbb{Y}}(k_x,k_y)\right)$$

Figure 9 - Representation of a Video Frame in Terms of Spatial Frequency

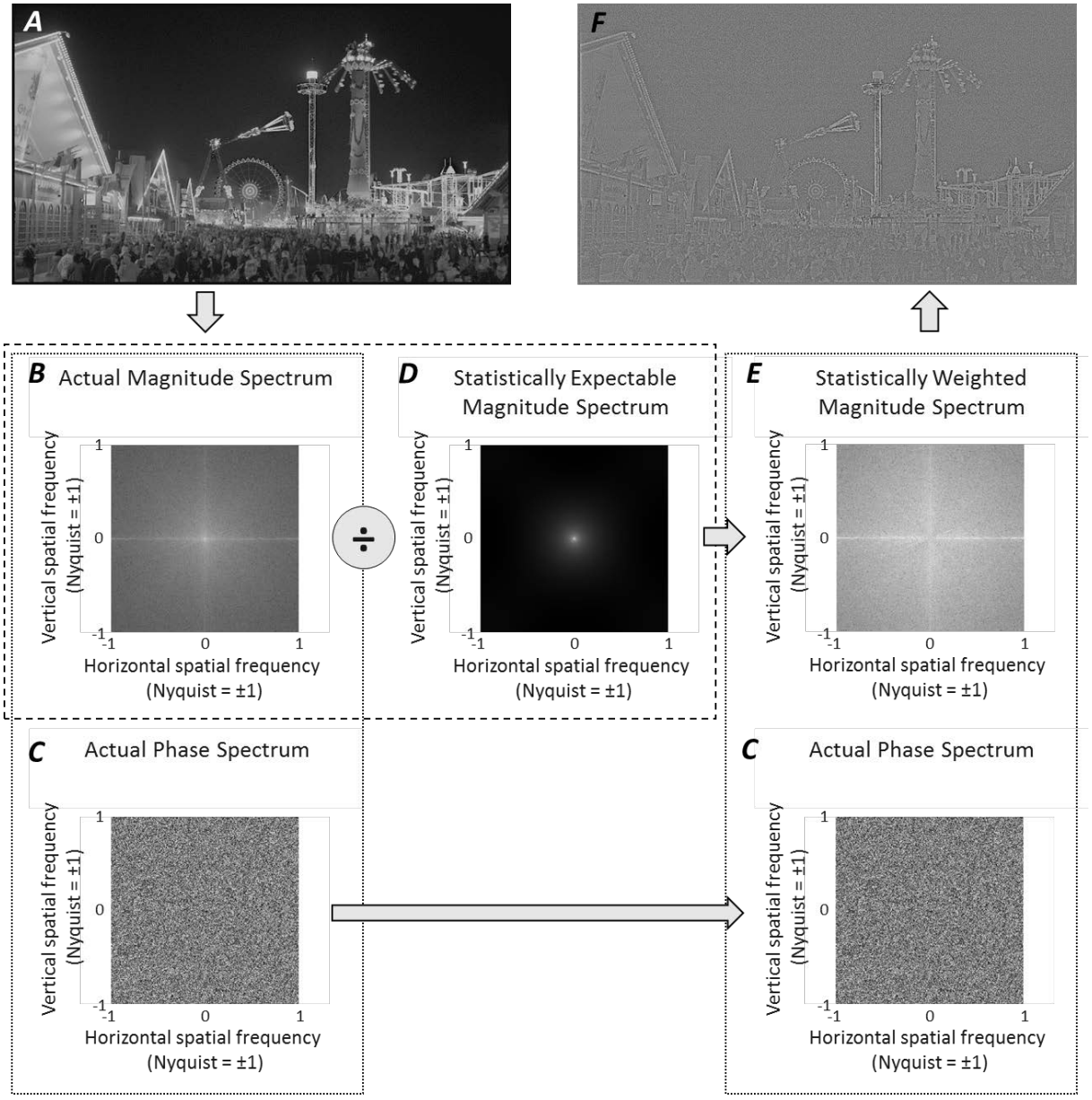


Figure 10 - Method of Calculating the Spatial Detail Signal

The Spatial Detail signal is calculated as illustrated in Figure 10. First, the magnitude (***B***) and phase spectra (***C***, shown twice) are calculated from the image pixel array (***A***). Next, a predetermined archetype of the statistically expectable one-over-frequency magnitude spectrum (***D***) is divided into the actual magnitude spectrum to produce a statistically weighted magnitude spectrum (***E***). Third, the statistically weighted magnitude spectrum is combined with the actual phase spectrum (***C***). Finally, a 2-dimensional Inverse Fast Fourier Transform is applied to produce a pixel array that we call the Spatial Detail signal (***F***).

Abbreviations

4k	3840x2160 resolution
ABR	Adaptive Bitrate
ATIS	Alliance for Telecommunications Industry Solutions
ATSC	Advanced Television Systems Committee
DOCSIS	Data Over Cable Interface Specifications
EOTF	Electro-Optical Transfer Function
HD	High Definition
HDTV	High-Definition Television
HD-HDR	High Definition-High Dynamic Range
HdM	Hochschule der Medien (Stuttgart Media University)
HDR	High Dynamic Range
HEVC	High Efficiency Video Coding
HLG	Hybrid-Log Gamma
IMF	Interoperable Master Format
IP	Internet Protocol
ITU-R	International Telecommunication Union - Radiocommunication Sector
JPEG2000	Joint Photographic Experts Group-2000
MXF	Material Exchange Format
MSE	Mean-Squared Error
MSO	Multiple-System Operator
MVPD	Multichannel Video Programming Distributor
OETF	Opto-Electronic Transfer Function
PQ	Perceptual Quantizer
PSNR	Peak Signal-to-Noise Ratio
RGB	Red Green Blue
SHVC	Scalable High Efficiency Video Encoding
SMPTE	Society of Motion Picture and Television Engineers
TIFF	Tagged Image File Format
Ultra HD	Ultra High Definition
v210	Quicktime Raw Component Video Picture Format
WCG	Wide Color Gamut
YCbCr	Luma and Chroma Video Picture Format

Bibliography & References

1. ITU-R Report BT.2246-2 (2017) “The present state of ultra-high definition television.”
2. ITU-R Report BT.2390-2 (2017) “High dynamic range television for production and international programme exchange.”
3. Roettgers, J. “The Story Behind ‘Meridian’: Why Netflix Is Helping Competitors With Content and Code.” Variety, Sep 15, 2016.
4. ITU-R BT.709-6 (2015) “Parameter values for the HDTV standards for production and international programme exchange.”
5. Xiph.org https://media.xiph.org/video/derf/
6. SMPTE OV 2067-0:2017 “SMPTE Overview Document – Interoperable Master Format – Overview for the SMPTE 2067 Document Suite.”
7. SMPTE ST 377-1:2011 “SMPTE Standard – Material Exchange Format (MXF) – File Format Specification.”
8. ISO/IEC 15444-1:2016 “Information Technology – JPEG 2000 image coding system: Core coding system”
9. ffmpeg.org https://www.ffmpeg.org
10. ITU-R Rec. BT.2100-0 (2016) “Image parameter values for high dynamic range television for use in production and international programme exchange.”
11. SMPTE ST 2084:2014 “SMPTE Standard – High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays.”
12. SMPTE ST 2067-21:2016 “SMPTE Standard – Interoperable Master Format – Application #2E”
13. Froehlich, J., et al., “HdM-HDR-2014 Project,” http://www.hdm-stuttgart.de/~froehlichj/hdm-hdr-2014
14. Froehlich, J., Grandinetti, S., Eberhardt, B., Walter, S., Schillin, A., and Brendel, H. 2014. “Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays,” Proc. SPIE 9023, Digital Photography X
15. ITU-R Rec. BT.2020-2 (2015) “Parameter values for ultra-high definition television systems for production and international programme exchange.”
16. “V210 Video Picture Encoding” at Library of Congress (Sustainability of Digital Formats). https://www.loc.gov/preservation/digital/formats/fdd/fdd000353.shtml
17. ITU-R Rec. BT.1886 (2011) “Reference electro-optical transfer function for flat panel displays used in HDTV studio production.”
18. MathWorks, MATLAB. https://www.mathworks.com/
19. x265 (x265-64bit-10bit-2017-05-01.exe) https://builds.x265.eu/
20. McCarthy, S.T. 2017. “A Biologically-Inspired Approach to Making HDR Video Quality Assessment Easier” SMPTE Motion Imaging Journal, vol. 124, no. 4, pp 47-58
21. McCarthy, S.T. 2014. “Theory and practice of perceptual video processing in broadcast encoders for cable, IPTV, satellite, and internet distribution,” Proc. SPIE 9014, Human Vision and Electronic Imaging XIX
22. McCarthy, S. 2012. “A Biological Framework for Perceptual Video Processing and Compression,” SMPTE Mot. Imag. J., 119(8):24-32, Nov/Dec.
23. McCarthy, S.T. and Owen, W.G. 2006. “Apparatus and Methods for Image and Signal Processing”. US Pat. 6014468 (2000). US Pat. 6360021 (2002), US Pat. 7046852 (2006)
24. Field, D.J. 1987. “Relationship between the statistics of natural images and the response properties of cortical cells,” J. Opt. Soc. Am. A. Vol. 4, No. 12

25. Boyce, J.M., Ye, Y., Chen, J., and Ramasubramonian, A.K. “Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard.” IEEE Tans Circuits and Systems for Video Technology, Vol. 26, No. 1, 2015.
26. ATSC A/341:2017 “Video – HEVC”
27. Winkler, S. 2005 Digital Video Quality: Vision Models and Metrics, John Wiley & Sons
28. Wang, Z. and Bovik, A.C. 2009. “Mean squared error: love it or leave it? - A new look at signal fidelity measures,” IEEE Signal Processing Magazine, vol. 26, no. 1, pp. 98-117
29. “Lanczos resampling” https://en.wikipedia.org/wiki/Lanczos_resampling
30. Bansal, G. 2017. “What is the difference between coefficient of determinations and coefficient of correlation?” http://blog.uwgb.edu/bansalg/statistics-data-analytics/linear-regression/what-is-the-difference-between-coefficient-of-determination-and-coefficient-of-correlation/
31. ATIS 0800061 (2013) “Methodology for subjective or objective video quality assessment in multiple bit rate adaptive streaming”

