# PRACTICAL METHODS TO VALIDATE ULTRA HD 4K CONTENT

Sean McCarthy, Ph.D.
ARRIS

*Abstract*

*Ultra High Definition television is many things: more pixels, more color, more contrast, and higher frame rates. Of these parameters, "more pixels" is much more mature commercially and Ultra HD 4k TVs are taking their place in peoples' homes. Yet, Ultra HD content and service offerings are playing catch-up. We don't yet have enough experience to know what good Ultra HD 4k content is nor do we know how much bandwidth to allocate to deliver great Ultra HD experiences to consumers. In this paper, we describe techniques and tools that could be used to validate the quality of uncompressed and compressed Ultra HD 4k content so that we can plan bandwidth resources with confidence. We will describe the statistical methods we use to validate Ultra HD 4k content, and will present some of our results. We will also explore the impact of high-efficiency video coding (HEVC) compression on the statistics of Ultra HD 4k content. The data and analysis we present are intended to provide tools and data that could be used to optimize bandwidth allocation and design Ultra HD 4k service offerings.*

## INTRODUCTION

Only a decade ago, high definition HD was the big new thing. With it came new wider 16:9 aspect ratio flat screen TVs that made the living room stylish in a way that old CRTs couldn't match. Consumers delighted in the new better television experience. Studios, programmers, cable, telco, and satellite video providers delivered a new golden-age of television. HD is now table stakes most places, and where that is not yet the case, it will be soon enough.

Yet now, before we hardly got used to HD, we are talking about Ultra HD (UHD) with at least four times as many pixels as HD. In addition to and along with UHD, we are getting a brand new wave of television viewing options. The Internet has become a rival of legacy managed television distribution pipes. Over-the-top (OTT) bandwidth is now often large enough to support 4k UHD exploration. New compression technologies such as HEVC are now available to make better use of video distribution channels. And the television itself is no longer confined to the home. Every tablet, notebook, PC, and smartphone now has a part time job as a TV screen; and more and more of those evolved-from-computer TVs have pixel density and resolution to rival dedicated TV displays.

Is all that resolution going to make a difference to consumers? If yes, what bandwidth will 4k UHD programming need? Those are two big questions our industry is exploring with respect to planning UHD services; yet they are not independent questions.

4k UHD is still new enough in the studios and post-production houses that 4k-capable cameras, lenses, image sensors, and downstream processing are still being optimized. Can we be sure yet that the optics and post processing are preserving every bit of "4k" detail? On the distribution side, could video compression and multi-bitrate adaptive streaming protocols change the amount of visual detail to an extent that it could conceivably turn "4k" quality into something more like "HD" or even less?

If the 4k content we have available today for bandwidth and video quality testing does not truly have a "4k"-level of detail, then we could go astray and plan for less bandwidth than we might need for future 4k UHD services. If the 4k content we have available today is truly "4k", then we should also want to be sure that we do not over compress and turn 4k UHD into something less impressive.

Indeed, during our UHD 4k testing, we have found several candidate test sequences that appeared normal to the eye but turned out to have unusual properties when examined mathematically. Such content could lead to wrong conclusions when planning for UHD 4k bandwidth and services.

In this paper, we present mathematical techniques to help answer the question "How 4k is it?" Our method examines 4k UHD video to see if it has a statistically expectable distribution of spatial detail as a function of 2-dimensional spatial frequency. The benchmark for our statistical expectations is drawn from numerous studies of the statistics of natural scenes.

Our main objective in writing this paper is to describe methodology that might be useful in helping to decide which 4k UHD content should be included in the video test library intended to be used for bandwidth and video quality planning.

## SOURCES OF 4K UHD CONTENT

There are many online places from which to obtain 4k content that could be considered for testing purposes. Industry-focused sources include the European Broadcasting Union[1,2] (EBU), CableLabs[3], and blenderfoundation[4]. Stock footage typically intended for promotional projects, but which might also be considered for testing purposes, is available online sites such as Shutterstock[5], NYC B.Roll[6], NatureFootage[7], and others that can be found by searching keywords such as "4k stock." Video-sharing sites such as YouTube[8] and Vimeo[9] host compressed 4k UHD content that could be candidates for testing certain kinds of 4k UHD services.

## CAMERA CONSIDERATIONS

The quality of 4k content depends on the quality of the camera, the particulars of the post processing such as filtering and compression, and the skill of the camera operator and crew.

4k-capable cameras available today range from consumer camcorders to cream-of-the-crop professional 4k-cameras that are used to create premium cinema and television content. Even some smartphones boast 4k cameras.

Lens quality and image sensor size are key issues in 4k capture. Obviously, consumer and prosumer grade cameras should not be expected to have the top-quality lenses and image sensors found in high-end professional cameras. Yet, even in high-end cameras one needs to consider the interaction between the lens and image sensor. At this point in time, the image sensors found in many 4k cameras are larger than HD image sensors. As a result, depth-of-field tends to be shallower. Background and foreground details that are out of the plane of focus can be softer than they are in HD. Depth-of-field can be increased by decreasing the aperture, but at the expense on less light which can result in noisier video because of sensor noise. Longer exposure times could improve the amount of light captured, but then motion blur could become an issue. All of these opto-electrical items are capable of producing 4k content that has less spatial detail in the subject matter and more noise than would otherwise be expected. More important, such content could lead to wrong conclusions about the amount of bandwidth that will be needed to deliver great 4k experiences to consumers.

## COMPRESSION CONSIDERATIONS

Video compression works mainly by strategically reducing the amount of spatial detail in video. Each compressed video frame is predicted from previously stored frames as much as possible. Whatever is unpredictable is packaged as a residual signal and sent to decoders, but not before the residual is further refined by being converted into a signal having less numerical precision through a technique called quantization. In the MPEG family of compression standards (MPEG-2, AVC/H.264, and HEVC/H.265), quantization has the effect of preferentially reducing the spatial details that are associated with higher spatial frequencies. In addition, AVC and HEVC employ spatial blurring filters that reduce the noticeability of spatial discontinuities between coding blocks. Both frequency-sensitive quantization and spatial blurring can reduce the kind of fine detail that 4k aims to show off. Without that "4k" detail in our test content our bandwidth predictions could be off when the next even-better generation of 4k camera arrives.

## ADAPTIVE STREAMING CONSIDERATIONS

Over-the-top streaming services have taken the lead in delivering 4k content to consumers. Such services typically employ one of several adaptive streaming protocols that enable video to play smoothly even when the consumer's bandwidth fluctuates significantly. Each adaptive-streaming video player senses its available bandwidth and requests a segment of programming that fits within its capabilities. If enough bandwidth is available, a 4k-capable video player would select lightly-compressed high bitrate segments having full 4k resolution (3840x2160). More restricted bandwidth could force selection of more aggressively compressed versions of the content though still at full 4k resolution. Even more restricted bandwidth can force selection of aggressively compressed sub-4k resolution (for example, 1920x1080 or 960x540 etc.). The modulation of the resolution and the compression level mean that 4k adaptive-streaming services might not always be fully 4k. Thus, any test content derived from such a source could impact how test results should be interpreted.

## SPATIAL FREQUENCY

An image is normally thought of as a 2-dimensional array of pixels with each pixel being represented by red, green, and blue values (RGB) or luma and 2 chrominance channels (for example, YUV or YCbCr). An image can also be represented as a 2-dimensional array of spatial-frequency components as illustrated in Figure 1. The visual pixel-based image and the spatial-frequency representation of the visual image are interchangeable mathematically. They have identical information, just organized differently.

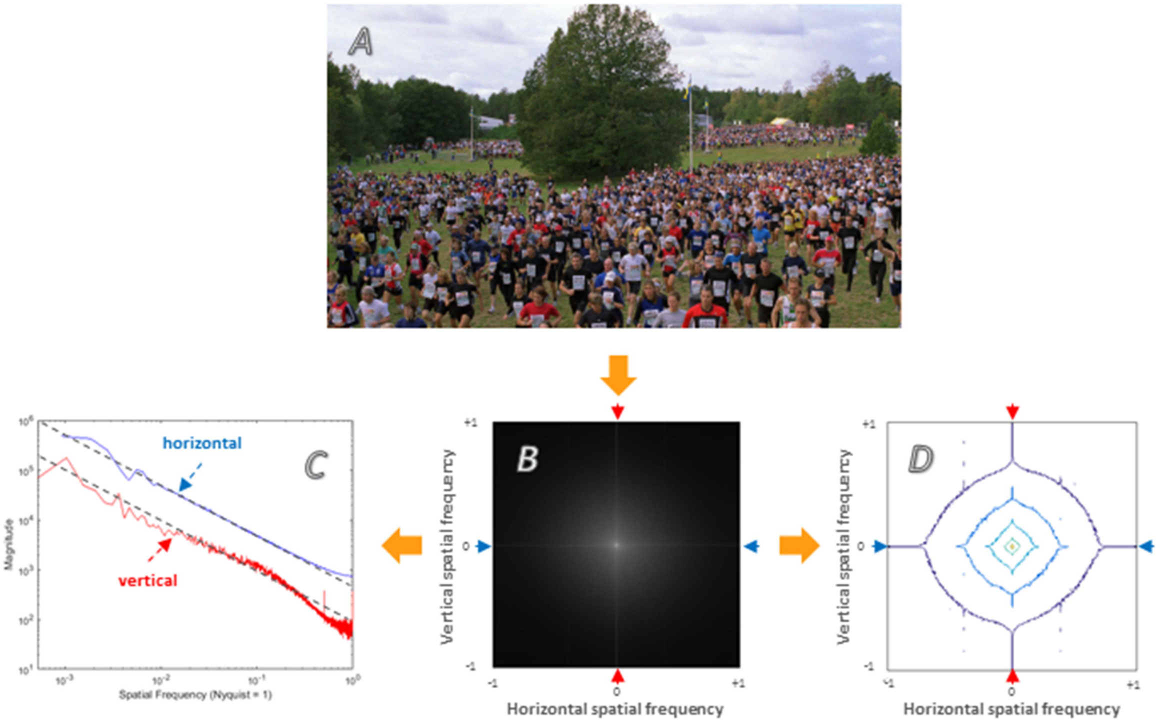

**Figure 1.** Illustration the representation of an image in terms of spatial frequencies. The visual pixel-based image (***A***) can be represented as a 2-dimensional array of complex numbers using Fourier transform techniques. The absolute-value of the complex numbers is shown as a 2-dimensional magnitude spectrum (***B***) in which brighter areas correspond to larger magnitude values. (Note that the log of the magnitude spectrum is shown in ***B*** to aid visualization. The horizontal and vertical frequency axes are shown relative to the corresponding Nyquist frequency (±1).) The magnitudes of the main horizontal and vertical spatial frequency axes are shown in ***C*.** The main horizontal spatial frequency axis corresponds to zero vertical frequency (blue arrows in ***B***), and the main vertical spatial frequency axis corresponds to zero horizontal spatial frequency (red arrows in ***B***). (Note that the magnitude spectrum is mirror symmetric around the 0,0 point (center of ***B*** & ***D***) along the main horizontal and vertical axes. Thus only the values from 0 to 1 (Nyquist frequency) are shown in ***C***.) The dashed lines in **C** indicate the 1/spatial frequency statistical expectation for natural scenes (the 1/spatial frequency appears as a line in the log plot). A contour map of the log of the magnitude spectrum is shown in ***D***. Contour maps provide useful *gestalts* of the overall 2D magnitude spectrum. The data shown in ***B***, ***C***, and ***D*** were obtained by averaging the magnitude spectrum of individual frames over 250 frames (5 seconds) of the SVT CrowdRun 2160p50 test sequence. (All video processing and analysis discussed in this were performed using MATLAB[10] and ffmpeg[11].)

Spatial-frequency representations of images can further be represented by a magnitude component and a phase component. The magnitude component, called the magnitude spectrum, provides information on how much of the overall variation within the visual (pixel-based) image can be attributed to a particular spatial frequency. (Spatial frequency is 2-dimensional having horizontal and vertical parts.) The phase component, called the phase spectrum (not shown), provides information on how the various spatial frequencies interact to create the features and details we recognize in images.

For the purposes of this study, we find it useful to use contour maps of the log of the magnitude spectra (as shown in Figure 1.) to create a *gestalt,* the 2-dimensional spatial frequency composition of images.

## STATISTICS OF NATURAL SCENES

Images of natural 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[12]. 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 scenes fall off as one-over-spatial-frequency." This statement applies to both horizontal and vertical spatial frequencies. Examples of images adhering to this statistical expectation are shown in Figure 2.

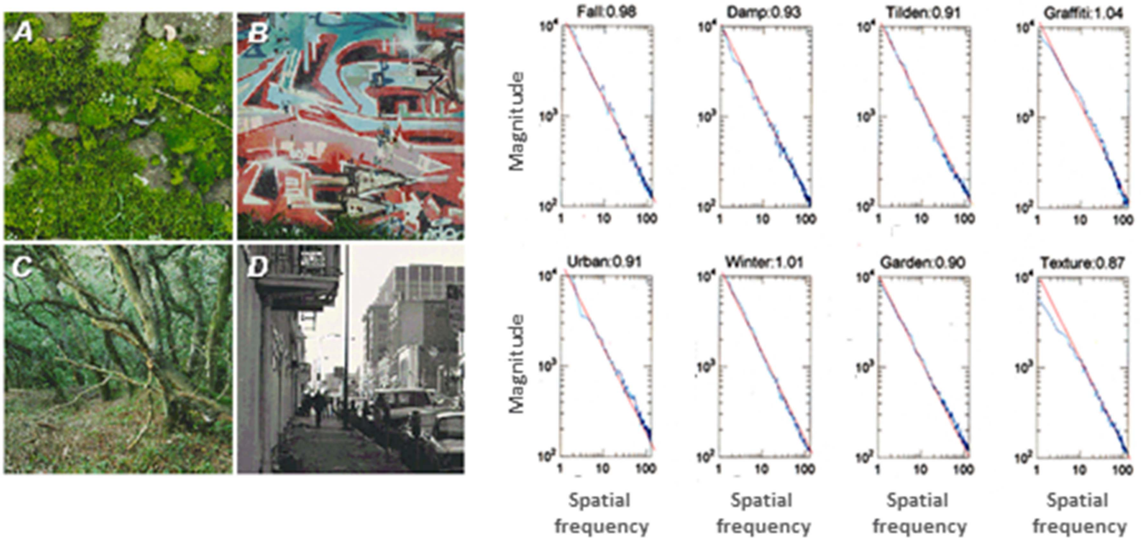

**Figure 2.** Examples of adherence to the 1/f statistical expectation. These 8 image series created and analyzed by McCarthy & Owen[13] illustrate the well-established statistical expectation that the magnitude spectra of images of natural scenes tend to be inversely proportional to spatial frequency. (Note that inverse proportionality appears as a straight line in the log plots shown.) The images shown in ***A***, ***B***, ***C***, and ***D*** are representatives of the "Texture" (close-ups of grass, sand, etc.), "Graffiti" (urban art), "Tilden Park" (woodland scenes), and "Urban" (San Francisco street scenes) image series, respectively. Other image series are "Fall" (colorful Fall foliage), "Damp" (puddles and environments where one might expect to find newts), "Winter" (snow scenes in New England), and "Garden" (UC Berkeley botanical garden).

Note that "natural-scene" images are not limited to pictures of grass and trees and the like. Any visually complex image of a 3-dimensional environment tends to have the one-over-frequency characteristic, though man-made environments tend to have stronger vertical and horizontal bias than unaltered landscape. The one-over-frequency characteristic can also be thought of as a signature of scale-invariance, which refers to the way in which small image details and large image details are distributed. Images of text and simple graphics do not tend to have one-over-frequency magnitude spectra.

## A BENCHMARK FOR SPATIAL DETAIL

In this paper, we leverage the one-over-frequency statistical expectation to see if it holds for 4k UHD content we have considered for use in our lab tests. Examples of 4k (2160p50) video sequences that largely adhere to the one-over-frequency statistical expectation are shown in Figure 3. Examples of candidate UHD 4k (3840x2160) videos that violate the one-over-frequency statistical expectation are shown in Figure 4.

According to the results shown in Figures 3 & 4, the test sequences shown in Figure 3 remain viable candidates to be used in experiments to explore UHD bandwidth planning; but we would exclude the test sequences represented in Figure 4 from our UHD 4k test library.

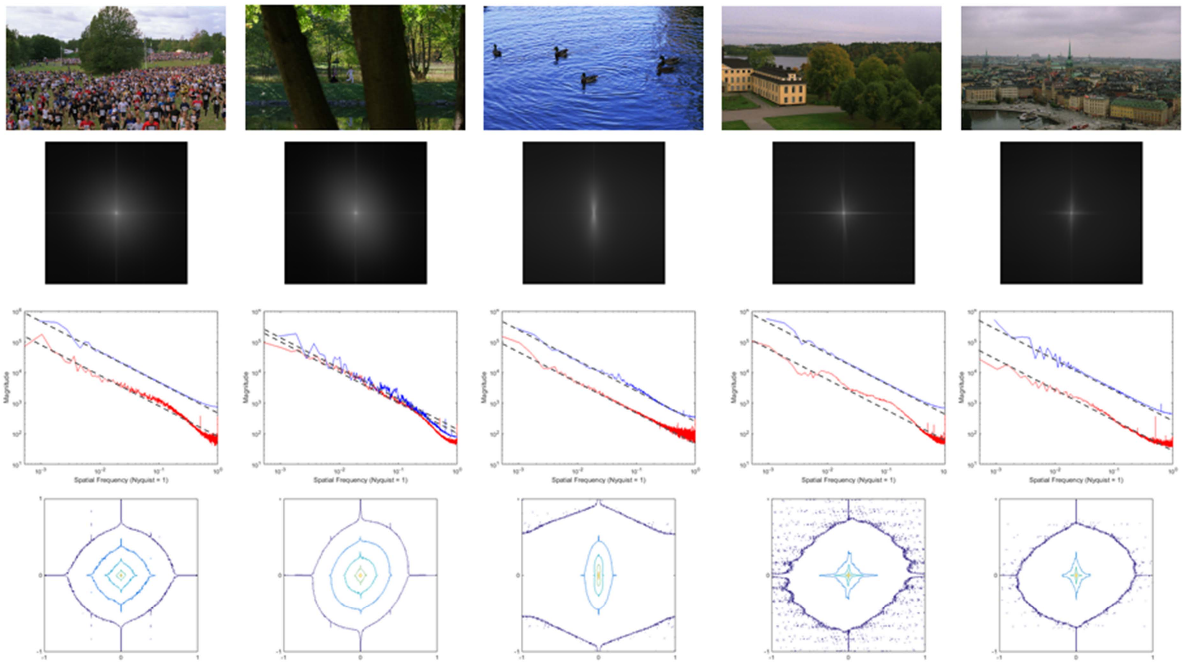

**Figure 3.** UHD 4k test sequences that have statically expectable magnitude spectra. The SVT UHD 4k test sequences (3840x2160 at 50 frames per second) “CrowdRun”, “ParkJoy”, “DucksTakeOff”, “InToTrees”, and “OldTownCross” are shown left to right in columns. The corresponding visual (pixel-based) image, log of the magnitude spectrum averaged over 250 frames (5 seconds), main horizontal and vertical axis components, and contour map of the log average magnitude spectrum are shown top to bottom in each column. Note that each of the sequences can be well-described by the one-over-frequency statistical expectation (dashed lines in the plot on the third row from the top); though there are some subtle deviations from statistical expectation (see Figure 5). Note also that the contour maps provide concise distinguishing information about each image sequence.

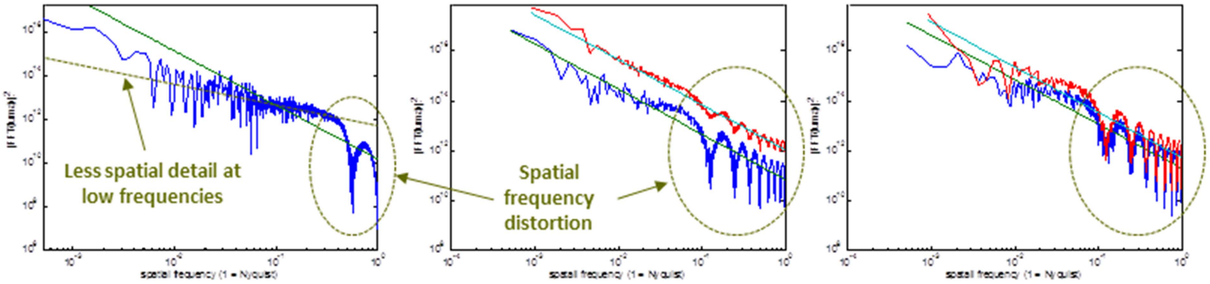

**Figure 4.** Example of candidate UHD test sequences that do not have statically expectable magnitude spectra. Some UHD 4k test sequences that we have obtained from various sources that appeared normal to the eye were found to have spatial magnitude spectra that were inconsistent with statistical expectations. Typical deviations from statistical expectations included: notch-like frequency distortions; excessive or diminished high or low frequency spatial detail (non-one-over-frequency behavior); and extraneous noise (see Figure 5).

The test sequences shown in Figure 3 are broadly in line with statistical expectations, however, they do show subtle deviations as illustrated in Figure 5. These deviations are mainly the presence of isolated narrow-band noise-like distortions and mild loss of high-frequency high spatial detail. The sole reason we present Figure 5 is to illustrate a method of scrutinizing candidate UHD 4k test content to an extent not possible with the eye alone. (It should be noted that the SVT UHD 4k test sequences shown in Figures 3 and 5 were produced 10 years ago, long before the emergence of UHD 4k as a consumer service, and thus were on the cutting edge of UHD 4k research and development.)

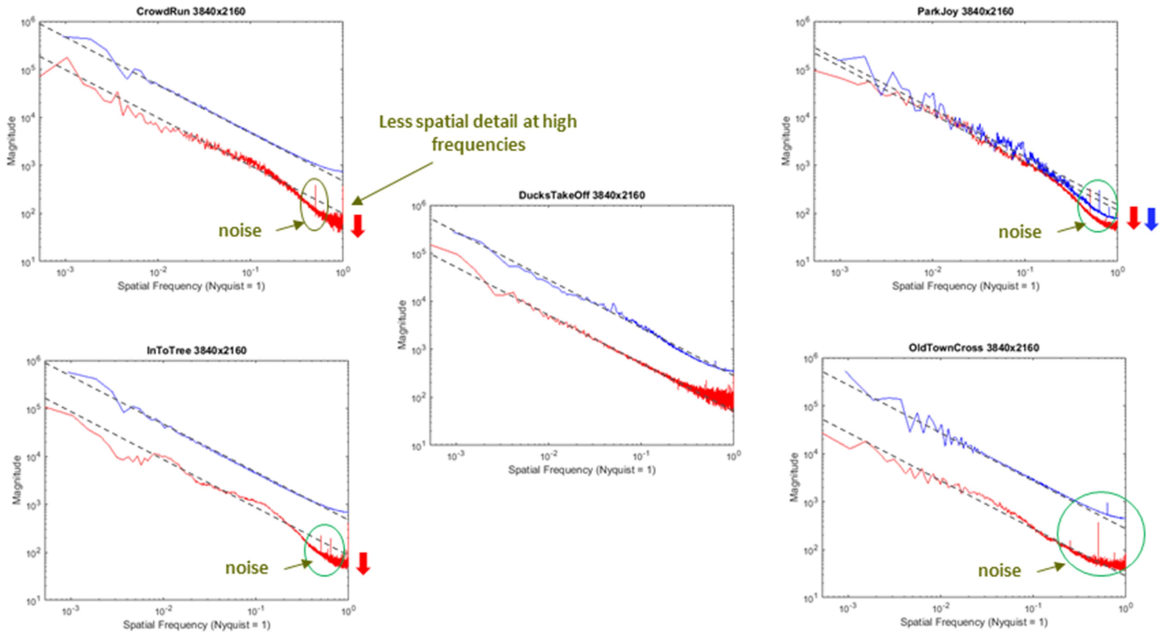

**Figure 5**. A closer scrutiny of subtle deviations from the statically expectable magnitude spectrum. Although most UHD 4k candidate test content matches the one-over-frequency statistical expectation in general, some sequence do show subtle deviations. As illustrates for the

SVT UHD 4k test sequences, these subtle deviations typically take to the form of extraneous noise that show up as isolated peaks and less-than-expected levels of high-frequency spatial detail (indicated by arrows pointing down). Note that the “DucksTakeOff” sequence meets statistical expectations particularly well.

## EFFECTIVE RESOLUTION

A key feature of adaptive streaming protocols is the inclusion of reduced-resolution versions of content in order to provide uninterrupted video service even when a consumer’s available bandwidth is significantly curtailed. Although compressed at resolution less than full 4k resolution, the content seen by a viewer is upconverted to 4k resolution by either a set top box or the television display itself. In this way, the effective resolution is less than the displayed resolution.

UHD 4k displays have such high resolution, and upconversion algorithms have become so good, that it is sometimes difficult to see by eye if a particular video is pristine full resolution or if some upconversion has occurred in the preparation of the content.

Figure 6 illustrates a method of analyzing the effective resolution of “4k” (3184x2160) resolution test content more quantitatively than can be done by eye. It is well-known that a reduced effective resolution correlates to a loss of high-frequency spatial detail. This loss could, in principle, be evident by inspecting the main horizontal and vertical axes of the magnitude spectrum. We find that is not always the case. Modern rescaling algorithms are very sophisticated and the difference between lowered effective resolution and full resolution can be subtle. Instead, we find that the contour maps of the log of the magnitude spectrum are a much more sensitive indicator of effective resolution. A test for accepting candidate UHD 4k test content into a master library could be along the lines of determining the average radius of the outermost contour, accepting test content only when a certain radius threshold is exceeded.

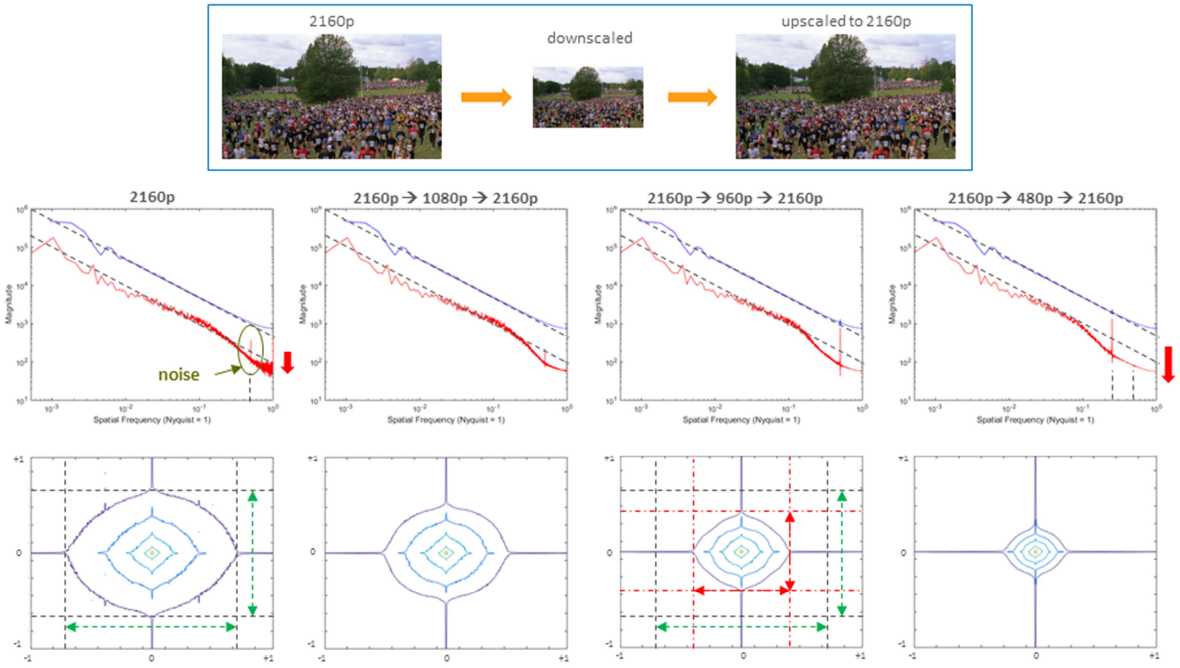

**Figure 6.** An example of using contour maps of magnitude spectra to examine effective spatial resolution. Some UHD 4k candidate test content could have lower effective resolution than full UHD 4k (3840x2160). We simulate such a situation by downscaling and then upscaling back to 3840x2160 resolution using ffmpeg. From left to right, the downscaled resolution is: unaltered 3840x2160; 1920x1080; 960x540; and 480x270 as an extremum. Note that examination of only the main horizontal and vertical axes of the magnitude spectrum (middle row) reveals some differences, most notably some reduction in high-frequency spatial detail and shift in the narrow-band noise; but these details are too subtle to make confident decisions, particularly when the original full resolution content is unavailable for comparison. The contour maps of the log of the average magnitude spectrum provide more clear-cut evidence. The contour levels are that same for all columns. Thus the constriction of the contours towards the center indicated that the magnitude spectrum narrows (loses high-frequency spatial detail) thus quantifying the reduced effective resolution. This is, of course, expected. The significance of this figure is that is illustrates that the contour map method can be a sensitive measure of effective resolution of candidate test video.

## EFFECT OF HEVC COMPRESSION

Video compression changes the amount of spatial detail in video, but the extent to which spatial detail is lost depends of the content itself and the aggressiveness of compression; i.e. the target bitrate.

In Figure 7 we demonstrate that our method of evaluating test content provides a way of testing the effective resolution of HEVC compressed content. Note that our method indicates that effective resolution is more sensitive to compressin for some kinds of content compared to other kinds of content. As such, our contour-map method could be used to optimize the selection of compressed UHD 4k content for testing purposes in terms of both the intrinsic image characteristics of content and the impact of bit rate. In this way, our contour-map method could serve as a content-independent method of measuring effective resolution and thus selection of useable UHD 4k test content.

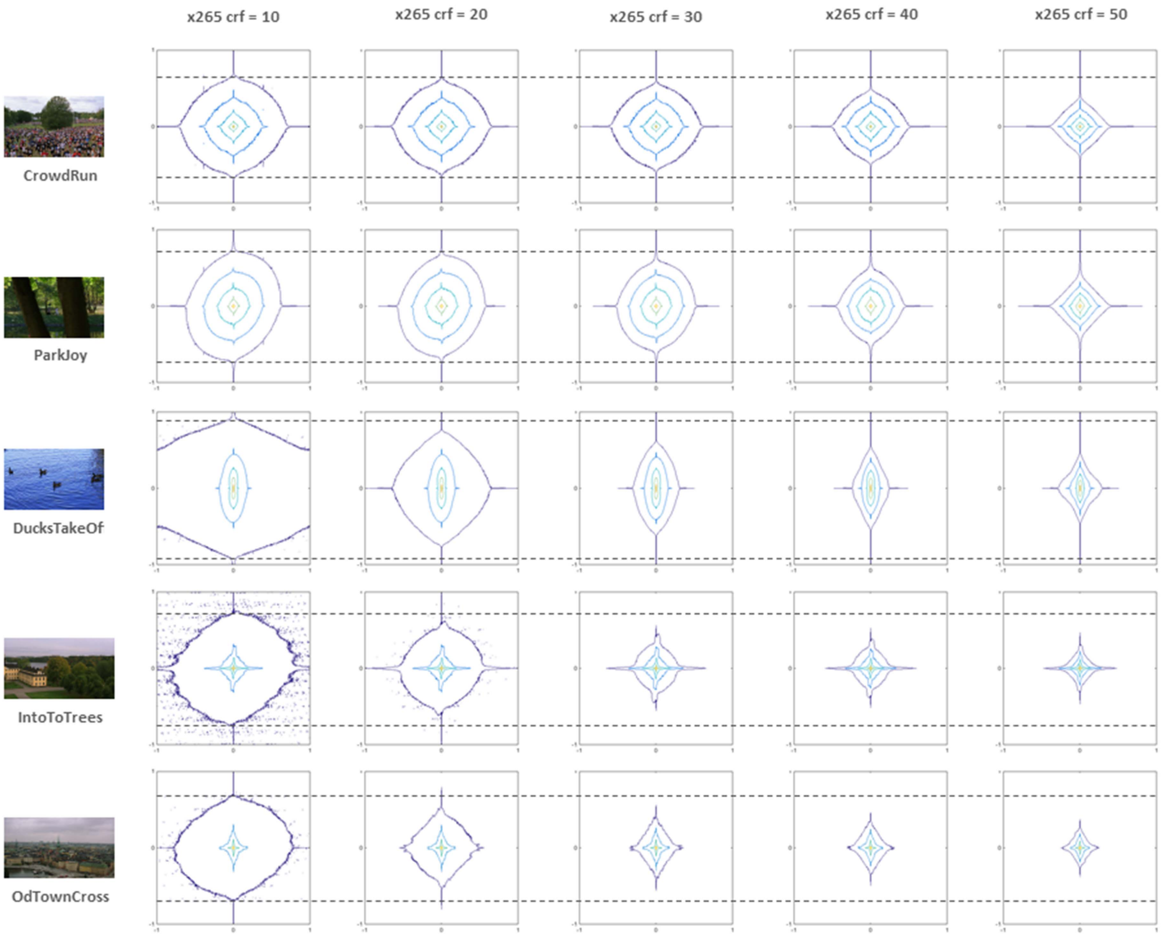

**Figure 7.** An example of using contour maps of magnitude spectra to examine the effect of video compression. Shown here are the contour maps of the average (250 frames, 5 seconds) of the log magnitude spectrum of each of the SVT UHD 4k test sequences compressed with HEVC to various extents. We used the libx265[14] library with ffmpeg to perform the HEVC compression. The ***crf*** value noted at the top of each column indicates the value of the constant rate factor (***crf***) parameter used in the ffmpeg libx265 command line. Smaller values of ***crf*** created more lightly compressed video. Video compressed with a ***crf*** value of 50 is typically very heavily artifacted. Video compressed with a ***crf*** value of 10 produces contour maps that are very similar to those for uncompressed video (see Figure 3). Note that the impact of the ***crf*** value is content dependent. For "CrowdRun" and "ParkJoy", the ***crf*** values below ~30 do not a have a major impact on effective resolution. On the other hand, a noticeable change in effective resolution is evident for a ***crf*** value of 20 for "IntoToTrees", "DucksTakeOff", and "OldTownCross". (The dashed lines provide a reference for the radial extent of outer contour of the lightly compressed and uncompressed versions of the video.)

## DISCUSSION & CONCLUSIONS

The objective of this paper was to present techniques that might be useful in evaluating UHD 4k video sequences that could be candidates for testing related to planning UHD 4k products and services. Selection of test content that is not representative of anticipated UHD 4k programming – including future UHD 4k programming that will be available when the end-to-end UHD 4k ecosystem has been optimized – could lead to wrong conclusions about what bandwidth and level of video quality would be needed.

We show in this paper that ensemble-average statistical expectations related to spatial frequency magnitude spectra of images of natural scenes can be used as a benchmark of comparison to address the question: "How 4k is it?" We find that major deviations from statistical expectations can be considered grounds for excluding content from a test library. (Though perhaps some synthetic compression busters should be retained to stress compression equipment and distribution services.)

We also find that examination of contour maps of the log of the magnitude spectra are sensitive indicators of effective resolution. Contour maps that indicate a lack of 4k-level effective resolution can be considered grounds for excluding content from a test library. The main horizontal and vertical components of the magnitude spectrum seem to be good probes for detecting added noise and gross distortions; but they are not highly sensitive probes of effective resolution.

Significantly, our contour-map method is also a sensitive content-independent probe that can be used to evaluate compressed content for inclusion in UHD 4k video test libraries to be used for planning UHD 4k video quality and bandwidth.

## REFERENCES

1) CableLabs 4k Resources. www.cablelabs.com/resources/4k/

2) EBU UHD-1 Test Sequences. https://tech.ebu.ch/testsequences/uhd-1

3) SVT High Definition Multi Format Test Set. tech.ebu.ch/webdav/site/tech/shared/hdtv/svt-multiformat-conditions-v10.pdf

4) blenderfoundation "Tears of Steel". mango.blender.org

5) ShutterStock. www.shutterstock.com

6) NYC B.Roll. www.nycbroll.com

7) NatureFootage. www.naturefootage.com

8) YouTube. www.youtube.com

9) Vimeo. www.vimeo.com

10) The Mathworks. www.mathworks.com

11) ffmpeg. www.ffmpeg.org

12) D.J. Field. "Relationship between the statistics of natural images and the response properties of cortical cells." J. Opt. Soc. Am. A. Vol. 4, No. 12 1987

13) S. McCarthy and W. G. Owen. Personal communication. University of California, Berkeley.

14) x265. www.x265.org